

**MADDE DÜZEYİNDE BOYUTLULUK MODELLERİNİN  
BİLGİSAYAR ORTAMINDA BİREYSELLEŞTİRİLMİŞ TEST  
YÖNTEMLERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ**

**EXAMINING THE EFFECTS OF ITEM LEVEL  
DIMENSIONALITY MODELS ON MULTIDIMENSIONAL  
COMPUTERIZED ADAPTIVE TESTING METHODS**

**Burhanettin ÖZDEMİR**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

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

Öngördüğü

Doktora Tezi

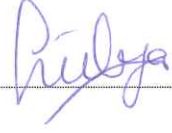
olarak hazırlanmıştır.

2015

Eđitim Bilimleri Enstitüsü M¼d¼rl¼ę¼'ne,

Burhanettin ÖZDEMİR'in hazırladığı "Madde D¼zeyinde Boyutluluk Modellerinin Bilgisayar Ortamında Bireyselleřtirilmiř Test Y¼ntemleri Üzerindeki Etkisinin İncelenmesi" bařlıklı bu alıřma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde Ölme ve Deęerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiřtir.

Bařkan Prof. Dr. H¼lyya KELECİOđLU



¼ye (Danıřman) Prof. Dr. Selahattin GELBAL



¼ye Do. Dr. řeref TAN



¼ye Yrd. Do. Dr. Murat AKYILDIZ



¼ye Yrd. Do. Dr. Sevda ETİN



ONAY

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim-Öęretim ve Sınav Y¼netmelięi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından 22 /07 / 2015 tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca ..... / ..... / ..... tarihinde kabul edilmiřtir.

Prof. Dr. Berrin AKMAN  
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

# MADDE DÜZEYİNDE BOYUTLULUK MODELLERİNİN BİLGİSAYAR ORTAMINDA BİREYSELLEŞTİRİLMİŞ TEST YÖNTEMLERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

**Burhanettin ÖZDEMİR**

## ÖZ

Bu çalışmanın amacı, farklı yetenek kestirimi yöntemleri, madde seçim yöntemleri ve test sonlandırma kurallarını dikkate alarak bireylerin yabancı dil yeteneklerinin telafi-edici modellere dayalı Çok Boyutlu Bilgisayar Ortamında Bireyselleştirilmiş (BOB) Testi yöntemleri ile ölçülmesi ve madde-içi ve maddeler-arası boyutluluğun çok-boyutlu BOB testi yöntemlerinin performansları üzerindeki etkisinin incelenmesidir. Bu amaç doğrultusunda, Hacettepe Üniversitesi tarafından uygulanan dinleme, okuduğunu anlama ve dilbilgisi olmak üzere üç boyuttan oluşan İngilizce Yeterlik Sınavlarına (İYS) ilişkin gerçek veri seti kullanılarak gerçek verilere dayalı simülasyon (post-hoc simulation) yapılmıştır.

Bu çalışmada, 2009-2013 eğitim-öğretim yıllarında uygulanan 10 İngilizce yeterli sınavına ait veri seti kullanılmış ve her bir testte yer alan maddelere ait madde parametreleri telafi-edici (compensatory) çok boyutlu 2 parametrelili lojistik model (CM-2PLM) kullanılarak kestirilmiştir. Madde-içi boyutluluk modeline ait madde havuzu 565 maddeden oluşurken, maddeler-arası boyutluluk modeline ait madde havuzu ise 559 maddeden oluşmaktadır. Bu çalışmada en uygun çok-boyutlu BOB testine karar vermek için iki farklı yetenek kestirim yöntemi (Fisher'in puanlama ve Bayesyen MAP yöntemi), üç farklı madde seçim yöntemi (A-optimality, D-optimality, Seçkisiz madde seçim yöntemi) ve iki farklı test sonlandırma kuralı (sabit madde sayısı ve hata varyansı durdurma kuralı) kullanılmıştır. Toplamda 72 koşul analiz edilmiş ve her bir koşula ilişkin analiz sonuçları güvenilirlik katsayıları, ölçmenin standart hatası, ortalama madde sayısı, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon ve RMSD değerleri açısından karşılaştırılmıştır.

Madde düzeyinde boyutluluk modellerine dayalı çok boyutlu BOB testi analiz sonuçlarına bakıldığında, farklı madde seçim ve yetenek kestirim yöntemlerinin kullanımının standart hata, testin uzunluğu, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon ve RMSD değerlerini etkilediği bulgusuna

ulařılmıştır. D-optimality madde seim yntemi yerine A-optimality madde seim yntemi kullanıldığında her bir boyutluluk modeli iin hem test uzunluęunun ve RMSD deęerlerinin azaldığı hem de her bir boyuta iliřkin testin gvenirlięinin arttığı bulgusuna ulařılmıştır. Dięer taraftan, madde seim yntemlerinden D-optimality ve yetenek kestirim yntemlerinden MLE'ye dayalı Fisher'in puanlama ynteminin madde dzeyinde boyutluluk modellerinden etkilendięi grlmektedir. Gerek verilere dayalı (post-hoc) simlasyon analizi bulgularına gre kaęıt-kalem testleri ile karřılařtırıldığında ok boyutlu BOB testlerinin daha az madde ile daha yksek gvenirlikte lmler yaptığı grlmektedir. Sonu olarak, A-optimality madde seim ve Bayesyen MAP yetenek kestirim yntemlerinin kullanıldığı madde-ii boyutluluk modeline dayalı ok boyutlu BOB testlerinin dięer ok boyutlu BOB testlerine gre daha gvenilir ve tutarlı sonu verdięi sylenebilir. Bu alıřmanın sonuları İYS sınavının gerek ok-boyutlu BOB testi yntemleri ile uygulanmasında nemli bir katkı saęlayabilir.

.

**Anahtar szckler:** Bilgisayar ortamında bireyselleřtirilmiş testler, madde dzeyinde boyutluluk modelleri, ok boyutlu madde tepki kuramı, yabancı dil sınavları

**Danıřman:** Prof. Dr. Selahattin GELBAL Hacettepe niversitesi, Eęitim Bilimleri Anabilim Dalı, Eęitimde lme ve Deęerlendirme Bilim Dalı

# **EXAMINING THE EFFECTS OF ITEM LEVEL DIMENSIONALITY MODELS ON MULTIDIMENSIONAL COMPUTERIZED ADAPTIVE TESTING METHODS**

**Burhanettin ÖZDEMİR**

## **ABSTRACT**

The purpose of this study is to measure students' language abilities with Compensatory Multidimensional Computerized Adaptive Testing (MCAT) designs using different ability estimation, item selection methods and stopping rules; and to examine the effect of item-level dimensionality models on MCAT. For this purpose, real data set from English Proficiency Test (EPT) administered by Hacettepe University was used to conduct post-hoc simulation, in which each test consist of three dimensions listening, reading and grammar, respectively.

In this study, 10 EPT data sets administered between 2009 and 2013, were used to conduct analysis. Item parameters were estimated with compensatory multidimensional 2 parameter logistic model (CM-2PLM) and item pool for with-in item dimensionality model consisted of 565 items, while item pool for between item dimensionality consisted of 559 items. In order to determine the best MCAT algorithm for EPT, two different theta estimation (Fisher scoring and Bayesian MAP) methods, three different fisher information based item selection methods (A-optimality, D-optimality and Random) and two different termination methods (fixed number of item, precision based) were used. In total, 72 different conditions were taken into consideration, and results of these conditions were compared with respect to, reliability index, SEM, averaged number of items administered and RMSD values between full bank theta and estimated MCAT theta.

MCAT Results indicated that using different theta estimation and item selection methods affected SEM, averaged number of administered items, correlation between true and estimated theta and RMSD values. Using A-optimality rather than D-optimality to select items both decreased average number of items administered, RMSD values and increased test reliability for both dimensionality models. On the other hand, both D-optimality item selection and MLE-based Fisher's scoring methods were affected from item-level dimensionality methods. Results also indicated that post-hoc MCAT simulation for EPT provided ability estimations with

higher reliability and fewer items compared to paper and pencil format. Overall, MCAT designs based on within-item models with A-optimality and Bayesian theta estimation method outperformed other MCAT designs. Results of this study would also provide an important guideline for live MCAT application of EPT.

**Keywords:** Computerized adaptive testing, item-level dimensionality models, multidimensional item response theory, language testing

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

## ETİK BEYANNAMESİ

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

- 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.

Burhanettin ÖZDEMİR

## TEŞEKKÜR

Çalışmalarım boyunca fikirleri ve önerileriyle beni yönlendiren, desteğini her zaman hissettiğim, her sıkıntıda ilgisini ve yardımlarını esirgemeyen, çalışmamı tamamlamada büyük katkıları olan danışmanım ve değerli hocam Prof. Dr. Selahaddin GELBAL'a sonsuz teşekkürlerimi sunarım.

Doktora eğitimim süresince kendisinden çok faydalandığım, emek ve zamanını harcayarak tez çalışmamı destekleyen, değerli hocam Prof. Dr. Hülya Kelecioğlu'na ve Doç. Dr. Nuri DOĞAN'a görüş ve önerileriyle tez çalışmama katkı sağlayan jüri üyeleri, Doç. Dr. Şeref Tan'a ve Yrd. Doç. Dr. Sevda ÇETİN ve Yrd. Doç. Dr. Murat AKYILDIZ'a sonsuz teşekkürlerimi sunarım.

Bu alanda yetişmemde emeği olan, mesleğimi sevmeme katkıları olan ve tez konumu belirlemeye çalıştığım ilk aşamadan itibaren görüş ve önerileriyle çalışmama değerli katkılarda bulunan değerli hocalarıma teşekkürü bir borç bilirim.

Hayatım boyunca yanımda olan, sevgisini, şefkatini ve desteğini esirgemeyen, hayatımın en zor dönemlerinde hep yanımda olan canım Annem, Babam ve Kardeşlerime sonsuz teşekkürlerimi bir borç biliyorum.

Ayrıca, evlendiğimiz ilk günden doktora sürecini tamamladığım güne kadar hep yanımda olan, beni hep en iyisini yapmak için motive eden ve bu tez sürecinde en az benim kadar yorulan canım eşim Esengül ÖZDEMİR'e sonsuz teşekkürlerimi sunuyorum. İyi ki varsınız....



## İÇİNDEKİLER

ÖZ .....	iii
ABSTRACT .....	v
TEŞEKKÜR .....	viii
İÇİNDEKİLER .....	ix
TABLolar DİZİNİ .....	xi
ŞEKİLLER DİZİNİ .....	xii
SİMGELER VE KISALTMALAR DİZİNİ .....	xiii
1. GİRİŞ .....	ix
1.1. Bilgisayar Ortamında Bireyselleştirilmiş Testlerin Tarihsel Gelişimi .....	2
1.2. Tek Boyutlu ve Çok Boyutlu BOB Testi Yöntemleri .....	5
1.3. Problem Durumu .....	6
1.4. Araştırmanın Amacı ve Önemi: .....	7
1.5. Problem Cümlesi: .....	9
1.5.1. Alt Problemler: .....	9
1.6. Sayıtlar: .....	11
1.7. Sınırlılıklar: .....	11
1.8. Araştırmanın Kuramsal Temeli .....	11
1.8.1. Madde tepki kuramı ( Item Response Theory) .....	11
1.8.2. Madde ve test karakteristik eğrisi .....	13
1.8.3. Madde Tepki Kuramının Varsayımları .....	15
1.8.4. Çok boyutlu madde tepki kuramı .....	16
1.8.5. Çok boyutlu MTK modelleri .....	17
1.8.6. MTK'ya dayalı madde ve test bilgi fonksiyonları .....	21
1.8.7. BOB Testlerinin Uygulama Basamakları .....	26
1.8.7.1. BOB Testlerinde Başlama Koşulunun Belirlenmesi .....	28
1.8.7.2. Madde Havuzunun Oluşturulması .....	29
1.8.7.3. Madde Seçim Yöntemleri .....	29
1.8.7.3.1. D-Optimality Madde Seçim Yöntemi .....	31
1.8.7.3.2. A-Optimality Madde Seçim Yöntemi .....	32
1.8.7.3.2. C-Optimality Madde Seçim Yöntemi .....	32
1.8.7.3.2. E-Optimality Madde Seçim Yöntemi .....	33
1.8.7.4. Yetenek Kestirim Yöntemleri .....	34
1.8.7.4.1. Maksimum Olabilirlik Kestirim Yöntemi (MLE) .....	34
1.8.7.4.2. Fisher'in Puanlama Metodu (Fisher's Scoring Method) .....	35
1.8.7.4.3. Bayesyen Olabilirlik Kestirim Yöntemleri .....	36
1.8.7.5. BOB Testlerinde Kullanılan Durdurma Kuralları .....	37
1.8.7.5. İçerik Ağırlıklandırılması .....	38
1.8.8. BOB Testlerinin Avantaj ve Dezavantajları .....	39
1.8.9. Bireyselleştirilmiş Testlerin Eğitimdeki Uygulamaları .....	40
2. İLGİLİ ARAŞTIRMALAR .....	44
2.1. BOB Testi İle İlgili Yurtdışında Yapılan Çalışmalar .....	44
2.2. BOB Testi İle İlgili Yurtiçinde Yapılan Çalışmalar .....	47
3. YÖNTEM .....	51
3.1. Araştırmanın Yöntemi .....	51

3.2. Çalışma Grubu.....	51
3.2.1. Çalışma Grubunun Özellikleri.....	51
3.3. Araştırma Verileri .....	51
3.4. Verilerin Analizi .....	52
3.4.1. Veri Analizinin Birinci Aşaması .....	52
3.4.1.1. Madde Parametrelerinin Kestirilmesi .....	52
3.4.1.2. Model-Veri Uyumlu İstatistiklerinin İncelenmesi .....	53
3.4.1. Veri Analizinin İkinci Aşaması .....	54
4. BULGULAR VE TARTIŞMA .....	57
4.1. Alt Problem 1: A-optimality madde seçim yönteminin kullanıldığı madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB test analizlerine ilişkin her bir koşula ait güvenilirlik katsayısı, standart hata, korelasyon, testin uzunluğu ve RMSD değerleri nasıldır? .....	57
4.2. Alt Problem 2: D-optimality madde seçim yönteminin kullanıldığı madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB test analizlerine ilişkin her bir koşula ait güvenilirlik katsayısı, standart hata, korelasyon, testin uzunluğu ve RMSD değerleri nasıldır? .....	74
4.3. Alt Problem 3: Seçkisiz madde seçim yönteminin kullanıldığı madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB test analizlerine ilişkin her bir koşula ait güvenilirlik katsayısı, standart hata, korelasyon, testin uzunluğu ve RMSD değerleri nasıldır? .....	90
4.4. Alt Problem 4: Çok boyutlu BOB testi yöntemi için belirlenen madde seçme yöntemlerinden hangisi daha iyi sonuç vermektedir? .....	106
4.5. Alt Problem 5: Çok boyutlu BOB testi yöntemi için belirlenen yetenek kestirim yöntemlerinden hangisi daha iyi sonuç vermektedir? .....	109
4.6. Alt Problem 6: Çok boyutlu BOB testi yöntemi için belirlenen test sonlandırma kurallarından hangisi daha iyi sonuç vermektedir?.....	111
4.7. Alt Problem 7: Analiz bulguları doğrultusunda en uygun çok boyutlu BOB testi algoritması nasıldır? .....	114
5. SONUÇ ve ÖNERİLER .....	116
5.1. Sonuçlar.....	116
5.2. Öneriler.....	121
5.2.1. Araştırmaya Dönük Öneriler.....	121
5.2.2. Uygulamaya Dönük Öneriler .....	122
KAYNAKÇA.....	124
EKLER DİZİNİ .....	132
EK-1: ETİK KURUL ONAY BİLDİRİMİ.....	133
EK-1: Çok-Boyutlu Madde-İçerik Boyutluluk ve Tek Boyutlu Modele İlişkin Uyum İstatistikleri .....	134
EK-2: Tek Boyutlu ve Çok-Boyutlu Maddeler-Arası Boyutluluk Modeline İlişkin Uyum İstatistikleri .....	135
EK-3: Maddeler-Arası ve Madde-İçerik Boyutluluk Modeline İlişkin Uyum İstatistikleri .....	136
EK-1: ORJİNALLİK RAPORU.....	137
ÖZGEÇMİŞ .....	138

## TABLULAR DİZİNİ

Tablo 4.1. A-optimality Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin BOB Testi Bulguları.....	58
Tablo 4.2. A-optimality Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralına İlişkin BOB Testi Bulguları.....	66
Tablo 4.3. D-optimality Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin BOB Testi Bulguları.....	75
Tablo 4.4. D-optimality Madde Seçim Yöntemi ve Hata varyansı Durdurma Kuralına İlişkin BOB Testi Bulguları.....	81
Tablo 4.5. Seçkisiz (Random) Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin BOB Testi Bulguları.....	91
Tablo 4.6. Seçkisiz (Random) Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralına İlişkin BOB Testi Bulguları.....	98

## ŞEKİLLER DİZİNİ

Şekil 1.1. Madde karakteristik eğrisi .....	14
Şekil 1.2. Çok boyutlu madde karakteristik yüzeyi.....	19
Şekil 1.3. Maddeler-arası ve madde-içi boyutluluk modeli.....	20
Şekil 1.4. Çok boyutlu teste ait madde bilgi ve test bilgi yüzeyi grafiği .....	25
Şekil 1.5. Tek boyutlu ve çok boyutlu BOB testleri ile bireyin yetenek kestirimi grafiği .....	27
Şekil 4.1. A-optimality Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin Grafikler.....	63
Şekil 4.2. A-optimality Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralına İlişkin Grafikler .....	71
Şekil 4.3. D-optimality Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin Grafikler.....	79
Şekil 4.4. D-optimality Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralına İlişkin Grafikler .....	87
Şekil 4.5. Seçkisiz (Random) Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin Grafikler.....	95
Şekil 4.6. Seçkisiz (Random) Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralına İlişkin Grafikler.....	103

## SİMGELER VE KISALTMALAR DİZİNİ

**CAT:** Computerized Adaptive Testing

**MCAT:** Multidimensional Computerized Adaptive Testing

**MIRT:** Multidimensional Item Response Theory

**GMAT:** Graduate Management Admission Test

**GRE:** Graduate Record Examination

**TOEFL:** Test of English as a Foreign Language

**MTK:** Madde Tepki Kuramı

**KTK:** Klasik Test Kuramı

**IRT:** Item Response Theory

**EPT:** English Proficiency Test

**İYS:** İngilizce Yeterlik Sınavı

**MLE:** Maximum Likelihood Estimation

**MAP:** Maximum A Posteriori

**RMSD:** Root Mean Squared Difference

**ÖSH:** Ölçmenin Standart Hatası

## 1. GİRİŞ

Günümüzde bireyin belli alandaki özelliğini veya yeteneğini kestirmek için genellikle kâğıt-kalem testleri uygulanmaktadır. Ülkemizde yaygın olarak kullanılan bu geleneksel yöntemde bireylere ölçülecek özelliğe ilişkin aynı soruların yer aldığı ve sabit sayıda soru içeren farklı kitapçıklar verilerek belli bir süre zarfında bireyin bütün sorulara yanıt vermesi beklenir. Daha sonra bireyin hangi soruları cevapladığına bakılmaksızın doğru cevaplanan soru sayısına bağlı olarak bireyin testten aldığı toplam puan hesaplanır. Yani bireyin cevapladığı maddelerin nitelikleri göz önünde bulundurulmaz. Ayrıca, testi alan bireyin yetenek düzeyinin düşük ya da yüksek olmasına bakılmaksızın bireylerin bütün maddelere cevap vermesi beklenir. Bireylerin yetenek düzeyleri farklı olduğundan testin zorluk derecesi bireylere göre değişkenlik gösterir. Dolayısıyla, testi alan öğrenci kendisi için çok kolay ve çok zor olan soruları da yanıtlamak zorunda kalır. Bireyin yetenek düzeyine göre çok zor ve çok kolay sorular ölçülmek istenen özellik hakkında yeterli bilgi sağlamayabilir.

Sabit madde sayısına sahip geleneksel ölçme araçları genellikle ranji sınırlandırılmış özellikleri veya yetenekleri ölçmek için kullanılır; ayrıca testi oluşturan maddelerin çoğu orta düzeyde yeteneğe sahip bireyler için daha uygundur. Klasik test teorisinde teste yer alan maddelerin bir kısmının çok zor ya da çok kolay olması nedeniyle yetenek düzeyi yüksek ve düşük olan bireyler için ölçümlerin güvenilirliğinin düşük çıkmasına neden olabilir. Geleneksel yöntemlerin bir diğer sınırlılığı, test için belirlenen sabit bir süre olduğundan, bireyin maddeleri cevaplama hızının düşük olması bireyin yetenek düzeyinin olduğundan daha düşük kestirilmesine neden olur (Weiss, 2004).

Geleneksel kâğıt-kalem testleri ekonomik olmasına rağmen, bilişim alanındaki ve öğrenme psikolojisindeki gelişmeler, “ test uygulama süresinin” kullanılabilirlik özelliğini yeniden gözden geçirilmesini zorunlu hale getirmektedir. Özellikle de, bilgisayar programlarındaki yenilikler, bilgisayarlar aracılığıyla soruların geliştirilmesine imkân vermektedir (Çıkrıkçı-Demirtaşlı, , 1999).

Bilgisayar teknolojilerindeki gelişmeler sosyal hayatımızı, çevremizi ve yaşam tarzımızı etkilediği gibi eğitimde kullanılan ölçme ve değerlendirme araçlarını, yöntem ve tekniklerini de aşamalı olarak etkilemektedir. Bu gelişmelere bağlı olarak günümüzde daha nitelikli ve verimli eğitim vermek için bilgisayar teknolojilerinden önemli ölçüde faydalanılmaya çalışılmaktadır. Bu amaç doğrultusunda bireylerin özelliklerini veya

yeteneklerini bilgisayar teknolojilerini kullanarak ölçmeyi amaçlayan alternatif ölçme yöntemleri geliştirilmiştir. Bunlardan en temel olanı testlerin kâğıt-kalem yerine bilgisayar ortamında sorulmasıdır. Bu yöntem genellikle bilgisayar destekli testler (Computer-Based Tests-CBT) olarak adlandırılmaktadır. Diğer bir alternatif ölçme yöntemi, bireyin yetenek düzeyi ile maddelerin özelliklerinin bilgisayar ortamında eşleştirildiği bilgisayar ortamında bireyselleştirilmiş test (Computerized Adaptive Testing-CAT) yöntemleridir (McBride Sz Martin, 1983; Weiss ve Kingsbury, 1984).

Kağıt-kalem testleri ile karşılaştırıldığında, bilgisayar ortamında bireyselleştirilmiş testleri (BOB testleri) kullanmanın birçok avantajı vardır. En önemli avantajı ise daha az madde ile daha güvenilir ölçümler yapmasıdır (Wainer, 1993). Bunun temel sebebi, BOB testi sürecinde kestirilen yetenek düzeyine ilişkin en yüksek bilgiyi sağlayan maddelerin seçilmesi ve bireyin yetenek düzeyi ranjının dışındaki çok zor ve çok kolay maddelerin teste dâhil edilmemesidir. Ayrıca, BOB testleri uygulama zamanı açısından esneklik sağladığı gibi, test biter bitmez bireye test sonucu hakkında bilgi ve dönüt verebilme özelliğine sahiptir (Lin, 2012).

Bireyin yeteneğinin bilgisayar ortamında ölçülmesi yeni bir fikir olmamasına karşın, teknolojik altyapı yetersizliği ve bilgisayarların karmaşık ölçümleri yapabilecek nitelikte olmaması önceki yıllarda bu alanda yapılan çalışmaları sınırlandırmıştır. Ancak, yirmi birinci yüzyılın başında bilgisayar teknolojilerindeki hızlı gelişim, yeni BOB testi yöntemlerinin gelişimini tetiklediği gibi uygulanabilirliğini de arttırmıştır.

### **1.1. Bilgisayar Ortamında Bireyselleştirilmiş Testlerin Tarihsel Gelişimi**

Günümüzde de artık hemen hemen her alanda kullanılan bilişim teknolojileri sistemleri ile bunlara bağlı olarak geliştirilen yazılımların, sınav ve test sistemlerinde de kullanılması kaçınılmaz hale gelmiştir. Bu gelişmelere bağlı olarak ortaya çıkan “Bilgisayar Ortamında Bireyselleştirilmiş Test” (Computerized Adaptive Testing -CAT) uygulaması, bireyin ölçülen yeteneği ile test maddelerinin güçlük düzeyinin, bilgisayar ortamında eşleştirilmesidir (Weiss, 2004).

Alfred Binet tarafından geliştirilen IQ testi (Binet ve Simon,1905) ilk bireyselleştirilmiş test olarak kabul edilmektedir. Binet zekâ testinin ilk bireyselleştirilmiş test olarak kabul edilmesinin temel sebebi, günümüzde kullanılan BOB testlerine ait uygulama aşamalarına sahip olmasıdır. Bir diğer önemli özelliği ise ölçme kuramlarından klasik test kuramı (KTK) yerine madde tepki kuramına (MTK) dayanmasıdır. Bu testin temel

amacı her bir bireyin yeteneğinin kestirilmesi için kullanılacak en uygun maddelerin seçilerek bireyin yeteneğinin ölçülmesidir. Bu amaç doğrultusunda, Binet tarafından testin bireyselleştirilmiş bir şekilde uygulanabilmesi için önceden hazırlanmış bir madde havuzu, zorluk düzeylerine göre sınıflandırılmış sorular, başlangıç kuralı, puanlama cetveli, madde seçim kuralı ve testi sonlandırma kuralı belirlenmiştir. Dolayısıyla, Binet'in zekâ testi ile BOB testleri arasındaki en önemli fark; Binet'in bireyselleştirilmiş zekâ testinde uygulayıcı bir psikolog iken, BOB testlerinde ise uygulayıcı bilgisayar programıdır (Weiss, 2004).

Ancak, 1900'lu yıllarında başındaki koşullar ve bunu takip eden yarım asırlık süreçte meydana gelen dünya savaşları Binet'in geliştirmiş olduğu zekâ testinin sadece kâğıt-kalem testi formatında bireyselleştirilmiş test olarak kalmasına neden olmuş ve bireyselleştirilmiş testlerin gelişimini engellemiştir. Daha sonra, 1950'li yıllarda Amerika Ordusundaki araştırmacılar, Binet'in zekâ testinin hem kâğıt kalem formatında hemde test makinesi ile bireyselleştirilmiş olarak uygulanmasına ilişkin çalışmalar yapmış, ancak başarısız olmuşlardır (Bayroff, 1964; Bayroff, Thomas ve Anderson, 1960). Bu tarihe kadar bireyselleştirilmiş testler, Binet'in geliştirmiş olduğu geleneksel kâğıt-kalem testi ile sınırlı kalmıştır.

Özellikle 1960'lı yılların sonunda Amerikan Deniz Kuvvetlerinin personel eğitimi araştırmaları ofisi ile Frederic Lord (1970, 1971a) tarafından madde tepki kuramı ve bireyselleştirilmiş testlere ilişkin yürütülen çalışmalar, bireyselleştirilmiş testlerin gelişimine önemli katkı sağlamıştır. Ayrıca Ferguson'un (1969) bilgisayar destekli eşleştirilmiş testlere ilişkin çalışmaları, eğitimde başarının ölçülmesine ilişkin çalışmalar (örn, Cleary, Linn, ve Rock, 1968, 1969) ve Weiss ve Betz'in (1973) BOB testleri üzerindeki çalışmaları bu alandaki gelişmelerin temelini oluşturmaktadır. Günümüzde interaktif bilgisayarların kullanılmasına olanak sağlayan BOB testlerinin gelişimine ise Wainer ve arkadaşları (2000) ile Van der Linden ve Glas'ın (2000) yapmış olduğu çalışmalar önemli katkı sağlamıştır.

Özellikle, yirmi birinci yüzyılın başından itibaren, bilgisayar teknolojisi alandaki hızlı gelişim, bireyselleştirilmiş testlerin temelini oluşturan karmaşık MTK modellerine uyumlu programların gelişimi gerçek BOB testi uygulamalarının artmasına olanak sağlamıştır. Bu süreçte gerçek BOB testi uygulamalarına olanak sağlayan birçok yeni Test Şirketleri (Testing Companies) kurulmuştur. Son on yılda ise çok boyutlu MTK



modellerine ve bilişsel tanı modellerine dayalı çok-boyutlu BOB testi çalışmalarına ağırlık verilmektedir.

Son yıllarda, BOB testlerine ilişkin yapılan çalışmalar bilişsel tanı (cognitive diagnostic-CD) ve çok-boyutlu madde tepki kuramı yaklaşımları (multidimensional item response theory -MIRT) yönünde gelişmektedir. Bilişsel tanı modelleri hakkında birçok çalışma yapılmasına karşın, bilişsel tanı modellerinin çoğu bireyleri ölçülen özellikler bakımından başarılı ve başarısız olmak üzere iki grupta sınıflamayı amaçlamaktadır (Wang, Chang ve Boughton, 2011). Diğer taraftan, çok boyutlu MTK'ya dayalı BOB testlerinde ise ölçülen her bir boyuta ilişkin bireyin yeteneği sürekli bir ölçek puanı üzerinden hesaplanırken, asıl amaç birey hakkında daha fazla bilgi toplamaktır (Wang, Chang ve Boughton, 2011; Lin, 2012).

Bilişsel tanı modellerine ve çok-boyutlu BOB testlerine dayalı araştırmaların ve uygulamaların yapılmasını engelleyen en önemli faktör, bilgisayar teknolojilerinin yeterli düzeyde gelişmemiş olmasıydı (Mulder ve van der Linden, 2010). Fakat özellikle son on yıldaki gelişen teknoloji kapasitesi ve hesaplama gücündeki artış çok-boyutlu MTK'nın sağladığı avantajların uygulanmasını kolaylaştırdı (Mulder ve van der Linden, 2010). Çok-boyutlu madde tepki kuramına dayalı BOB testleri, çok-boyutlu BOB (Multidimensional Computerized Adaptive Testing-MCAT) olarak adlandırılmaktadır.

Tek boyutlu BOB testleri ile karşılaştırıldığında, çok boyutlu BOB testlerinin farklı boyutlara ilişkin bilginin kullanılmasına olanak sağlamasından dolayı daha az madde ile daha yüksek güvenilirlikte ölçümler yaptığı belirtilmektedir (Segall, 1996; W. C. Wang ve Chen, 2004). Alan yazınında çok-boyutlu BOB testlerinin bireylerin yetenek parametrelerini kestirme (Veldkamp ve van der Linden, 2002; Yao, Pommerich ve Segall, 2013) veya bireyleri sınıflama (Luecht, 1996) amacına dönük çeşitli uygulama çalışmaları yapılmıştır. Bu uygulama çalışmalarını desteklemek ve geliştirmek amacıyla; madde seçim yöntemleri (Mulder ve van der Linden, 2009; W. C. Wang ve Chang, 2011; W. C. Wang, Chang ve Boughton, 2011; Yao, 2012), en uygun (optimal) madde havuzunun oluşturulması (Yao, 2014), yetenek kestirim yöntemleri ve test sonlandırma kuralları (Yao, 2013) gibi çok-boyutlu BOB testlerinin uygulama aşamaları hakkında çeşitli çalışmalar yapılmıştır. Ayrıca çoktan seçmeli maddeler, yapılandırılmış cevap (constructed response) ve çok kategorili maddelerden oluşan testler için geliştirilmiş çok-boyutlu BOB testleri vardır (Yao ve Schwarz, 2006; Lin, 2012; Sie 2015).

## 1.2. Tek Boyutlu ve Çok Boyutlu BOB Testi Yöntemleri

Bilgisayar ortamında bireyselleştirilmiş testler (CAT), psikolojik ve eğitimsel testlerin daha etkili ve verimli bir şekilde uygulanması için yeniden düzenlenerek interaktif bilgisayarlar ile uygulanmasıdır (Van der Linden ve Glas, 2000; Wainer et al., 2000). Bu yöntemin temel amacı, her bireyin ölçülen özelliğini en etkili ve verimli bir şekilde ölçecek bireyin yetenek düzeyine uygun maddeleri seçmektir.

Madde tepki tepki kuramının geliştirilmesine paralel olarak, bilgisayar teknolojisindeki ilerlemeler, bilgisayar ortamında bireyselleştirilmiş (BOB) testlerin uygulanabilirliğini arttırmıştır.. Madde seçim ve yetenek kestirim yöntemleri için tek boyutlu madde tepki kuramının kullanıldığı bireyselleştirilmiş testler tek boyutlu bilgisayar ortamında bireyselleştirilmiş test (Tek boyutlu BOB testleri-Unidimensional CAT) olarak adlandırılmaktadır ( Wang ve Chen, 2004).

Bilgisayar ortamında bireyselleştirilmiş testler, Binet'in geliştirmiş olduğu uyarlanmış test yönteminin uygulayıcı birey yerinde bilgisayar programı kullanılarak uygulanmasıdır. BOB testi yönteminin uygulanma sürecinde test maddeleri kullanılacak bilgisayar programına yüklenir ve bilgisayar ekranında görünmesi sağlanır. Daha sonra birey klavyeyi veya fareyi kullanarak maddelere cevap verir.

Binet testini uygulayan araştırmacı gibi, bilgisayar programı bireyin teste nasıl başlayacağına karar verir; bireyin önceki maddelere vermiş olduğu cevaplara bağlı olarak maddeleri seçer ve bir ya da birkaç kural belirleyerek testi sonlandırır. İlk BOB uygulamaları Binet'in kullanmış olduğu yöntemin farklı versiyonları iken (Weiss, 1973), sonraki uygulamalar ise madde havuzunun oluşturulma biçimine göre farklılık göstermektedir (Lord, 1971b, 1971c).

Burada bilgisayar bir çoklu ortam aracı gelişmiş grafik, ses vs. gibi özellikler sunmaktan çok daha fazlasını yapmaktadır. Bilgisayar önceden belirlenmiş bir soru grubunu sırayla uygulamak yerine, soruları testin gidişine göre seçerek sormaktadır. Çünkü bir bireye kendi yetenek düzeyinin çok üstünde (çok zor) sorular sormak ya da yetenek düzeyinin çok altında sorular sormak (çok kolay) testi alanın yetenek düzeyi hakkında fazla bilgi sağlamamaktadır. Ayrıca yetenek düzeyine uygun olmayan sorular sormak kişinin sıkılmasına da neden olabilmektedir (Sands, Waters, McBride, 1997).

BOB testlerinde tek boyutlu MTK yaygın olarak kullanılmasına rağmen, gerçek test uygulamaları için uygun olmayabilir. Özellikle bilişsel özelliklerin ölçülmesi ve

değerlendirilmesinin gerektiği portfolyo değerlendirmeleri, performans görevleri, klinik yeteneklerinin ölçülmesi ve değerlendirilmesi, yazma ve konuşma becerilerinin ölçülmesi ve projelerin değerlendirilmesi söz konusu olduğunda bireylerin yeteneklerinin doğru bir şekilde ölçülmesi için çok boyutlu madde tepki kuramına ihtiyaç duyulmaktadır (van der Linden ve Hambleton, 1997, s. 221). Çok boyutlu madde tepki kuramının uygulama alanlarının yaygınlaşması ve bilgisayar ortamında bireyselleştirilmiş testlerin kâğıt-kalem testlerine alternatif olarak görülmesi, her iki yöntemin birleşimi olan *çok boyutlu bireyselleştirilmiş testlerin* (Segall, 1996, 2001) geliştirilmesine olanak sağlamaktadır.

### 1.3. Problem Durumu

Birçok araştırmada birey hakkında en çok bilgi veren testlerin bireyin yetenek düzeyine uygun olan testler olduğu vurgulanmaktadır. Böylece bireyin yetenek düzeyine uygun olmayan çok kolay ve çok zor sorulara cevap verme gereksinimi duyulmadan, daha az soru ile yetenek kestirimi yapılabilir. Bu ise genellikle madde tepki kuramına dayanan bireyin ölçülen yeteneği ile test maddelerinin güçlük düzeyinin, bilgisayar ortamında eşleştirildiği bilgisayar ortamında bireyselleştirilmiş (BOB) testler ile gerçekleştirilebilir (Diao ve Reckase, 2009; Yoo, 2011). Nitekim yapılan çalışmalarda, BOB testi yönteminin, geleneksel kâğıt kalem testleri ile karşılaştırıldığında, soru sayısını %50 oranında azalttığı ve bireyin yetenek düzeyini daha yüksek güvenilirlikte hesapladığı vurgulanmaktadır (Lord,1980; Weiss,1983; Weiss 2004).

Bireyselleştirilmiş testlerin geleneksel kâğıt-kalem testlerine göre avantajlı yönleri olduğunu belirtenlere karşın, bu durum her bireyselleştirilmiş testin geleneksel kâğıt-kalem testlerinden üstün olduğu anlamına gelmeyebilir. Ayrıca en uygun BOB testine karar vermek için testin içeriğine ve ölçtüğü özelliğin yapısına uygun bir BOB testi algoritmasının geliştirilmesi gerekir. Dolayısıyla, her hangi bir BOB testi geliştirme aşamasında cevaplanması gereken bazı temel sorular vardır. Bu temel sorular; hangi modelin kullanılacağı, ilk maddenin nasıl seçileceği, test sürecinde bireyin yetenek parametresinin nasıl hesaplanacağı, bir sonraki maddenin nasıl seçileceği ve testin nasıl sonlandırılacağıdır (Diao ve Reckase, 2009). Bu sorular uygun bir şekilde cevaplandırıldığında geliştirilmesi amaçlanan en uygun BOB testi belirlenmiş olur. Ayrıca, bu temel sorular, BOB testi geliştirme sürecinin aşamalarını oluşturmaktadır.

Bu araştırma ile birçok sınavda olduğu gibi, yükseköğretimde İngilizce Yeterlik Sınavı (İYS)'na giren öğrencilerin yabancı dil yeteneklerinin, bireyin yetenek düzeyine

bakılmaksızın bütün bireylere aynı soruların sorulduğu geleneksel yöntem yerine, bireyin yetenek düzeyine uygun soruların sorulmasına olanak sağlayan BOB testi yöntemleri ile ölçülmesi amaçlanmaktadır. İYS testi dinleme, dilbilgisi ve okuduğunu anlama olmak üzere üç temel alt boyuttan oluşmaktadır. Dolayısıyla, bu çalışmada bireylerin her bir boyuta ilişkin yeteneklerini kestirmek amacıyla çok boyutlu madde tepki kuramına dayalı çok boyutlu BOB testi yöntemleri kullanılmıştır. Çok-boyutlu BOB testi için en uygun modele karar vermek amacıyla, çok boyutlu telafi-edici modele dayalı madde düzeyindeki boyutluluk modellerinden *madde-içi boyutluluk* ve *maddeler-arası boyutluluk modelleri* karşılaştırılmıştır. Her bir modele dayalı çok boyutlu BOB testi için en uygun madde seçim yöntemini belirlemek amacıyla Fisher'in bilgi matrisine dayalı *A-optimality*, *D-optimality* ve her hangi bir kuralın kullanılmadığı *seçkisiz (Random)* madde seçim yöntemleri kullanılmıştır. Çok-boyutlu BOB testi için en uygun yetenek kestirim yöntemini belirlemek için MLE yetenek kestirim yöntemlerinden Fisher'in puanlama yetenek kestirim yöntemi ve Bayesyen MAP yetenek kestirim yöntemi kullanılmıştır. Ayrıca, çok-boyutlu BOB testinde en uygun test sonlandırma kuralını belirlemek için ise sabit madde sayısı durdurma kuralı ve hata varyansı durdurma kuralı kullanılmıştır.

Bu çalışmada, en uygun çok-boyutlu BOB testine karar vermek için farklı madde seçim yöntemleri, yetenek kestirim yöntemleri ve durdurma kurallarının kullanıldığı çok-boyutlu BOB testi algoritmalarına ait her bir boyuta ilişkin *güvenirlilik katsayıları*, yetenek parametrelerine ilişkin *RMSD katsayıları*, gerçek ve kestirilen yetenek parametreleri arasındaki *korelasyon katsayıları* ve her bir boyuta ilişkin *ölçmenin standart hatası* (standard error of measurement-SEM) değerleri hesaplanarak karşılaştırılmıştır.

Her bir koşul için maddenin kullanım sıklığını (item exposure) kontrol etmeye olanak sağlayan randomesque yöntemi kullanılmıştır. Bu yöntem ile analizlerde madde havuzundan madde seçilirken test bilgisini maksimum yapan ilk on madde arasından birinin seçkisiz olarak atanması kuralı uygulanmıştır.

#### **1.4. Araştırmanın Amacı ve Önemi:**

Bu çalışmanın amacı, farklı yetenek kestirimi yöntemleri, madde seçim yöntemleri ve test sonlandırma kurallarını dikkate alarak bireylerin yabancı dil yeteneklerinin telafi-edici modellere dayalı Çok Boyutlu Bilgisayar Ortamında Bireyselleştirilmiş Test Yöntemleri ile ölçülmesi (Multidimensional Computerized Adaptive Testing-MCAT) ve

madde düzeyinde boyutluluğun çok boyutlu BOB testi yöntemlerinin performansları üzerindeki etkisinin incelenmesidir.

Bilgisayar ortamında bireyselleştirilmiş testler günümüzde GMAT (Graduate Management Admission Test), GRE (Graduate Record Examination), TOEFL (Test of English as a Foreign Language) gibi uluslararası testlerde kullanılmaktadır. Bu araştırma ile yukarıda belirtilen sınavlarda olduğu gibi, Hacettepe Üniversitesinde İngilizce Yeterlik Sınavına (İYS) giren öğrencilerin yabancı dil yeteneklerinin, bireyin yetenek düzeyine bakılmaksızın bütün bireylere aynı soruların sorulduğu geleneksel yöntem yerine, bireyin yetenek düzeyine uygun soruların sorulmasını olanak sağlayan, Çok Boyutlu Bilgisayar Ortamında Bireyselleştirilmiş Test yöntemleriyle ölçülmesi amaçlanmaktadır. Ayrıca bu çalışma ile bireylerin yabancı dil yeteneklerinin kestirilmesi için kullanılacak farklı telafi-edici çok boyutlu BOB testi yöntemlerinin performanslarının karşılaştırılması ve en uygun BOB testi yöntemine karar verilmesi amaçlanmaktadır.

Alan yazınına bakıldığında, özellikle ülkemizde çok boyutlu BOB testi yöntemleri ile ilgili yapılan çalışmaların henüz olmadığı görülmektedir. Bu araştırma çok boyutlu BOB testi yöntemlerinin performanslarının karşılaştırılmasına olanak sağladığından önemli görülmektedir.

Çok-boyutlu BOB testlerine ilişkin yapılan çalışmalarda genellikle çok-boyutlu telafi-edici MTK modellerinden madde-içi boyutluluk modeli (karmaşık model) kullanılmıştır. Bu çalışmada ise hem madde-içi boyutluluk hem de maddeler-arası boyutluluk modelleri kullanılmış ve her iki modele dayalı çok-boyutlu BOB testlerine ilişkin analiz bulguları karşılaştırılmıştır. Bu çalışma, farklı çok boyutlu MTK modellerine dayalı çok-boyutlu BOB testlerinin performanslarını karşılaştırmaya olanak sağladığından önemli görülmektedir.

Bu çalışmayı önemli kılan bir diğer özelliği ise İYS sınavına ait gerçek verilerin kullanılması ve gerçek verilere dayalı simulasyon (post-hoc simulation) yönteminin kullanılmasıdır. Bu çalışma, birçok ülkede yapılan uluslararası sınavlarda olduğu gibi, Türkiye’de de bireylerin yabancı dil yeteneklerinin bireyselleştirilmiş testlerle (adaptive tests) bilgisayar ortamında ölçülmesine bir ön hazırlık olması açısından büyük önem taşımaktadır. Araştırmanın problem cümlesi aşağıdaki gibi düzenlenmiştir.

### **1.5. Problem Cümlesi:**

İngilizce Yeterlik Sınavı (İYS)'na giren öğrencilerin yabancı dil yeteneklerinin bilgisayar ortamında ölçülmesi için geliştirilen farklı telafi-edici modellere dayalı farklı yetenek kestirimi yöntemleri, madde seçim yöntemleri ve test sonlandırma kurallarının kullanıldığı çok boyutlu BOB testi yöntemlerinin performansları nasıldır?

#### **1.5.1. Alt Problemler:**

**1.** A-optimality madde seçim yönteminin kullanıldığı madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB test analizlerine ilişkin farklı yetenek kestirim yöntemleri ve durdurma kuralları altında her bir koşula ait güvenilirlik katsayısı, standart hata, yetenek parametrelerine ait korelasyon, testin uzunluğu ve RMSD değerleri nasıldır?

**a)** Maddeler-arası boyutluluk modeline ait sabit madde sayısı durdurma kuralları ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**b)** Madde-içi boyutluluk modeline ait sabit madde sayısı durdurma kuralları ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**c)** Maddeler-arası boyutluluk modeline ait farklı standart hata durdurma kuralları ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**d)** Madde-içi boyutluluk modeline ait farklı standart hata durdurma kuralları ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**2.** D-optimality madde seçim yönteminin kullanıldığı madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB test analizlerine ilişkin farklı yetenek kestirim yöntemleri ve durdurma kuralı altında her bir koşula ait güvenilirlik katsayısı, standart hata, yetenek parametrelerine ait korelasyon, testin uzunluğu ve RMSD değerleri nasıldır?

**a)** Maddeler-arası boyutluluk modeline ait sabit madde sayısı durdurma kuralları ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**b)** Madde-içi boyutluluk modeline ait sabit madde sayısı durdurma kuralları ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**c)** Maddeler-arası boyutluluk modeline ait farklı standart hata durdurma kuralları ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**d)** Madde-içi boyutluluk modeline ait farklı standart hata durdurma kuralları ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**3.** Seçkisiz (Random) madde seçim yönteminin kullanıldığı madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB test analizlerine ilişkin farklı yetenek kestirim yöntemleri ve durdurma kuralı altında her bir koşula ait güvenilirlik katsayısı, standart hata, yetenek parametrelerine ait korelasyon, testin uzunluğu ve RMSD değerleri nasıldır?

**a)** Maddeler-arası boyutluluk modeline ait sabit madde sayısı durdurma kuralları ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**b)** Madde-içi boyutluluk modeline ait sabit madde sayısı durdurma kuralları ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**c)** Maddeler-arası boyutluluk modeline ait farklı standart hata durdurma kuralları ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**d)** Madde-içi boyutluluk modeline ait farklı standart hata durdurma kuralları ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?

**4.** Çok boyutlu BOB testi yöntemi için belirlenen test sonlandırma kurallarından telafi-edici hata varyansı ile sabit soru sayısından hangisi güvenilirlik, hata istatistikleri ve test uzunluğu açısından daha iyi sonuç vermektedir?

**5.** Çok boyutlu BOB testi yöntemi için belirlenen madde seçme yöntemlerinden hangisi daha iyi sonuç vermektedir?

**6.** Çok boyutlu BOB testi yöntemi için belirlenen yetenek kestirim yöntemlerinden hangisi daha iyi sonuç vermektedir?

## 7. Analiz bulguları doğrultusunda en uygun çok boyutlu BOB testi algoritması nasıldır?

### 1.6. Sayıtlar:

Genellikle çok-boyutlu BOB testlerinde her bir boyuta ilişkin yetenek parametreleri arasındaki korelasyonun veya varyans-kovaryans matrisi önsel dağılımının bilindiği varsayılmaktadır (Yoo, 2011). Bu çalışmada bireylerin kestirilen yetenek parametreleri arasındaki korelasyona bakılarak varyans-kovaryans matrisine ait önsel dağılımının bilindiği varsayılmaktadır ( $v_{cv}=c[0.9, 0.8, 0.8]$ )

Ayrıca, 2009-2013 yıllarında uygulanan İYS testlerinin bireyin ölçülen özelliklerini doğru ve güvenilir bir şekilde ölçtüğü ve bireylerin test maddelerine verdiği cevapların gerçek yetenek düzeylerini yansıttığı varsayılmaktadır.

### 1.7. Sınırlılıklar:

Bu çalışmada, gerçek verilere dayalı simülasyon yapılmasına olanak sağlayan ve çok-boyutlu BOB testleri için geliştirilmiş "MAT " (multidimensional adaptive testing, Choi, 2011) R paket programı kullanılmıştır. Dolayısıyla, bu çalışmada kullanılan madde seçim yöntemleri, yetenek kestirim yöntemleri, test sonlandırma kuralı ve madde kullanım sıklığı kontrol yöntemleri paket programda tanımlı yöntemlerle sınırlıdır.

Bu çalışmada 2009- 2013 akademik yılları arasında İngilizce yeterlik sınavında (İYS) çıkan gerçek veri seti kullanılmıştır. Madde havuzu oluşturulurken İYS sınavlarında çıkmış gerçek maddelere ait madde parametreleri kullanılmıştır. Analiz sürecinde ise sadece gerçek madde parametrelerine bağlı olarak bireylere ilişkin yetenek parametreleri türetilmiş ve her bir analiz için aynı yetenek parametreleri kullanılmıştır. Her bir koşul için yapılan analizler uzun sürdüğünden çok-boyutlu BOB testi sürecinde testi alan birey sayısı 500 ile sınırlandırılmıştır.

### 1.8. Araştırmanın Kuramsal Temeli

Tek boyutlu ve çok boyutlu BOB testlerinin temelini oluşturan ölçme modeli tek boyutlu ve çok boyutlu madde tepki kuramı olduğundan bundan sonraki bölümde madde tepki kuramı ve farklı tek boyutlu ve çok boyutlu madde tepki kuramı modelleri hakkında bilgi verilmiştir.

#### 1.8.1. Madde tepki kuramı ( Item Response Theory)

Bilgisayar ortamında bireye uyarlanmış testler için kullanılan en yaygın matematiksel kuram Madde Tepki Kuramıdır (MTK). MTK, test geliştirme, madde parametrelerinin



kestirilmesi ve bireye ait ölçülen özelliklere ilişkin yetenek parametrelerinin kestirilmesi açısından güçlü psikometrik temellerinden dolayı önemli avantajlara sahiptir. Bilgisayar ortamında bireyselleştirilmiş (BOB) testler açısından en önemli avantajı ise, bireyi ve maddeleri aynı ölçeğe yerleştirip, birey ile bireyin kestirilen yetenek düzeyini en iyi ölçen maddelerin eşleştirilmesini kolaylaştırmasıdır (Weiss, 2011).

Araştırmacıların BOB testlerinin geliştirilme ve uygulama basamaklarını anlayabilmesi için tek boyutlu ve çok boyutlu MTK modelleri hakkında genel bir bilgiye sahip olması gerekir. Bundan dolayı, bu bölümde MTK kuramı ve MTK kuramına dayalı tek boyutlu MTK modelleri ile çok boyutlu MTK modelleri hakkında bilgi verilmiştir.

Madde Tepki Kuramı, kişinin ölçülen özellikteki düzeyi (yetenek düzeyi) ile verdiği yanıtlar arasındaki ilişki olduğunu kabul eder ve bu ilişkiyi matematiksel bir fonksiyon ile açıklayan olasılıklı bir model önerir (Embretson ve Reise, 2000). Madde Tepki Kuramı modelleri maddeye verilen yanıt ile ölçülen yetenek veya özellik arasındaki ilişkiyi göstermektedir. Madde tepkisi doğru-yanlış, evet-hayır, katılıyorum-katılmıyorum gibi iki kategorili (dichotomous) veya Likert tipi bir ölçekte olduğu gibi iki kategoriden fazla (polytomous) olabilir (DeMars, 2010). İkili puanlanan veriler için uygun olan madde tepki kuramı modelleri; bir parametrelili, iki parametrelili ve üç parametrelili lojistik modeller (1PLM, 2PLM, 3PLM) olarak adlandırılmaktadır. Bir parametrelili model madde güçlüğü parametresi ( $b_i$ ) ile bireyin yetenek düzeyi arasında ilişkiye dayanır. İki parametrelili modelde madde zorluk parametresinin yanında belirli bir yetenek düzeyinde maddenin düşük ve yüksek yetenek grubundaki bireyleri ayırma gücü olarak tanımlanan ayırt edicilik parametresi ( $a_i$ ) eklenmektedir. Üçüncü modelde ise iki parametrelili modelde düşük yetenek düzeyindeki bireyleri maddeyi doğru yanıtlama olasılığını veren şans parametresi ( $c_i$ ) eklenmektedir (Hambleton ve Swaminathan, 1984).

İkili puanlanmış cevaplar için geliştirilen tek boyutlu modellerden biri Birnbaum (1968) tarafından geliştirilen üç parametrelili lojistik (3PL) modeldir. Bu modelde,  $\theta$  yetenek düzeyine sahip  $j$  bireyinin  $i$  maddesine doğru cevap verme olasılığı:

$$P(U_{ij} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{[a_i(\theta_j - b_i)]}}{1 + e^{[a_i(\theta_j - b_i)]}} \quad (1.1)$$

formülü ile hesaplanmaktadır. Burada  $a_i$ ,  $i$  maddesine ait ayırt edicilik parametresi;  $b_i$ ,  $i$  maddesine ait madde güçlüğü ve  $c_i$  ise en düşük yetenek düzeyindeki bireyin

maddelere doğru cevap verme olasılığı ya da  $i$  maddesine ait şans başarısı olarak tanımlanmaktadır.

İki parametrelili lojistik (2PL) model, maddelere ait şans başarısı sıfıra eşit olduğu ( $c_i=0$ ) 3PL modelin özel halidir. Dolayısıyla 3PL modelde  $c_i$  yerine sıfır yazıldığında, 2PL modele ait olasılık fonksiyonu elde edilmiş olur. Bu modelde,  $\theta$  yetenek düzeyine sahip  $j$  bireyinin  $i$  maddesine doğru cevap verme olasılığı:

$$P(U_{ij} = 1 | \theta_j, a_i, b_i) = \frac{e^{[a_i(\theta_j - b_i)]}}{1 + e^{[a_i(\theta_j - b_i)]}} \quad (1.2)$$

formülü ile hesaplanmaktadır. Burada  $a_i$ ,  $i$  maddesine ait ayırt edicilik parametresi;  $b_i$ ,  $i$  maddesine ait madde güçlüğü olarak tanımlanmaktadır.

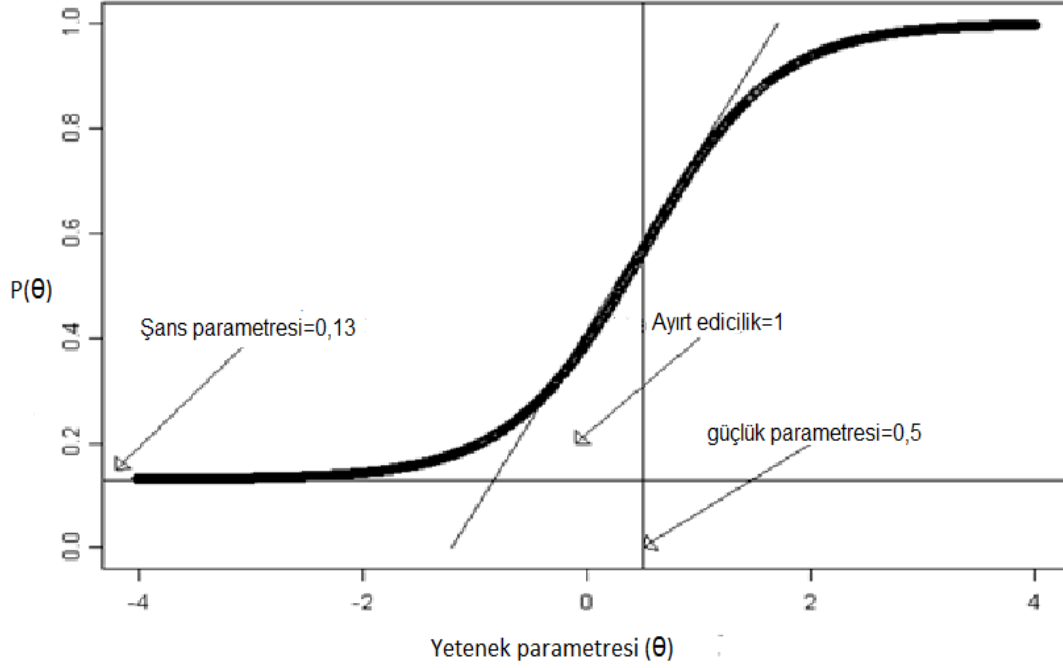
Bir parametrelili lojistik (1PL) model ise, maddelere ait ayırt edicilik parametresinin bütün maddeler için eşit olduğu 2PL modelin özel halidir. MTK modellerinden 1PL modelde,  $\theta$  yetenek düzeyine sahip  $j$  bireyinin  $i$  maddesine doğru cevap verme olasılığı sadece  $b_i$  madde güçlük parametresi kullanılarak aşağıdaki formülle hesaplanmaktadır:

$$P(U_{ij} = 1 | \theta_j, b_i) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \quad (1.3)$$

Madde tepki kuramı modellerine ait olasılık fonksiyonları, maddelere verilen cevaplar (cevap örüntüleri) ve madde parametreleri kullanılarak maddelere ait madde karakteristik eğrisi (Item Characteristic Curve-ICC) ve test karakteristik eğrisi (Test Characteristic Curve-TCC) çizilir.

### 1.8.2. Madde ve test karakteristik eğrisi

Madde karakteristik eğrisi (MKE), bir madde için bireylerin yetenek düzeyleri ile modele ait olasılık yoğunluk fonksiyonları aracılığıyla hesaplanan maddelere cevap verme olasılığı arasındaki ilişkiyi verir. MKE'nin hesaplanabilmesi için öncelikle uygun MTK modelinin belirlenmesi ve madde parametrelerinin kestirilmesi gerekir (Ree, 1979). MKE'nin şeklini MTK modelleri ile hesaplanan madde parametreleri belirler. Diğer bir ifade ile madde parametreleri; MKE'nin konumunu, eğimini ve başlangıç noktasını belirler. Aşağıda testin tek boyutlu olduğu durumda madde karakteristik eğrisine bir örnek verilmiştir.



**Şekil 1.1: Madde karakteristik eğrisi**

Şekil 1.1’de verilen grafikte yatay eksen yetenek düzeylerini ( $\theta$ ) dikey eksen ise sorunun doğru cevaplanma olasılığını temsil etmektedir. Grafikteki  $a_i$  grafiğin büküm noktasının ( $\text{prob}=0.5$ ) yatay eksen kestiği noktanın eğimidir. Grafikte yer alan  $a_i$ ’nin değeri, yani maddenin ayırt edicilik gücü arttıkça madde karakteristik eğrisi dikleşir. Madde güçlüğününün ( $b_i$ ) değeri madde karakteristik eğrisinin sola ya da sağa taşıyacaktır. Diğer bir ifade ile sorunun doğru yanıtlanma olasılığı yetenek düzeyleri için farklılaşacaktır (sola gittikçe soru kolay, sağa gittikçe soru zor demektir). Diğer parametre olan  $c_i$  eğrinin dikey eksen kestiği noktadır ve en düşük yetenek düzeyindeki bireylerin maddeye doğru cevap verme olasılığını ya da rastgele bir cevap işaretlendiğinde, bu cevabın doğru çıkma olasılığını verir; değeri yükseldikçe sorunun düşük yetenek düzeyindekiler tarafında çözülme olasılığı artar (Rudner, 1998).

Testi alan bireylerin testteki her bir maddeyi doğru cevaplama olasılıklarının toplamı ise bireyin testten alacağı toplam puanı verir. Bireylere ait toplam puan ise teorik olarak  $(-\infty, \infty)$  arasında değiştiği kabul edilmesine karşın, genellikle  $[-3,+3]$  aralığında değerler alır. Test karakteristik eğrisi ise bireylerin yetenek düzeyleri ile modele ait olasılık fonksiyonları aracılığıyla hesaplanan bireyin toplam puanı arasındaki ilişkiyi verir.

Bu kuramın en önemli noktası yetenek kestirimlerini uygulanan sorulardan bağımsız olarak vermesidir. Bireyler farklı sorularla test edilseler bile madde tepki kuramı farklı

kişilerin yeteneklerinin kestirebilmek için standart bir çerçeve sunmaktadır (Hambleton; Swaminathan ve Rogers, 1991). Aynı bireye farklı sorular içeren iki farklı test uygulansa bile her bir birey için kestirilen yetenek düzeyi (puanları) farklılık göstermez.

Çok kategorili puanlanan veriler için ise sınıflamalı cevap modeli ( Nominal Response Model, Bock,1972), kısmi puan modeli (Partial Credit Model, Masters, 1982), aşamalı tepki modeli (Graded Response Model, Smajima, 1969) örnek olarak gösterilebilir (Hambelton ve Swaminathan, 1985).

### 1.8.3. Madde Tepki Kuramının Varsayımları

Madde tepki kuramının klasik test kuramına göre üstünlüklerinden biri madde ve test parametrelerinin gruptan ve grupta yer alan bireylerin özelliklerinin madde örnekleminde bağımsız olarak kestirilmesidir. MTK'nın avantajları ancak model-veri uyumu tatmin edici derecede sağlandığında geçerlidir. Model-veri uyumu düşük olduğunda madde ve yetenek parametrelerinin değişmezliği sağlanamaz. Model-veri uyumu ise, birincil varsayımların karşılanması ile sağlanır.

Tüm tek boyutlu MTK modelleri için birincil varsayımlar:

1. Tek boyutluluk
2. Yerel bağımsızlık

Varsayımlarıdır.

**Tek boyutluluk** ile testin sadece tek bir yeteneği ya da özelliği ölçtüğü varsayılmaktadır ve bu varsayımın tam olarak karşılanması çok zordur. Bu varsayımın sağlanması için test performansını etkileyen baskın bir faktörün olması yeterlidir. Bu baskın faktör, test ile ölçülmesi amaçlanan yeteneği temsil etmektedir (Hambleton ve Swaminathan, 1985). Genellikle tek boyutluluk çalışması faktör analizi ile yapılmaktadır (Lord 1980; Hambleton ve Ravinelli 1986; Crocker ve Algina 1986; Harvey ve Hammer 1999; Doğan, 2002).

**Yerel bağımsızlık**, belirli bir yetenek düzeyinde olan alt grupların, bir maddeye verdikleri cevapların birbirinden istatistiksel olarak bağımsızlığına işaret etmektedir. Bu varsayımın doğru olması için bir bireyin, bir maddedeki performansının diğer maddelerdeki performansını etkilememesi gerekir. Başka bir ifadeyle, bir maddeye verilen doğru ya da yanlış cevabın, başka bir maddeye verilecek doğru ya da yanlış cevaptan bağımsız olmasıdır. Örneğin, bir madde diğer maddelerin cevaplanmasına

ilişkin ipucu vermemelidir ya da bir maddeye verilen cevap diğer maddenin cevaplanması için ön koşul olmamalıdır.

Tek boyutluluk ve yerel bağımsızlık varsayımları sağlandıktan sonra madde tepki kuramında veri yapısına en uygun model belirlenir (1PLM, 2PLM veya 3PLM). MTK'da beklenen model özellikleri ise;

1. Madde parametrelerinin değişmezliği
2. Yetenek parametrelerinin kestiriminin değişmezliğidir.

Bütün madde tepki modellerinin sahip olduğu en önemli özelliklerden biri madde parametrelerinin değişmezliği bir başka ifadeyle gruptan bağımsız olmasıdır. Hambleton ve Swaminathan (1985), madde parametrelerinin değişmezliğinin evrenden çekilen iki veya daha fazla alt gruptan (örneğin erkekler ve kızlar, üst ve alt grup gibi) elde edilen madde parametrelerinin karşılaştırılarak kontrol edilebileceğini belirtmişlerdir.

Benzer şekilde Hambleton ve Swaminathan (1985), yetenek parametresi kestirimlerinin değişmezliğinin ilgili madde havuzundan iki veya daha fazla madde örnekleme seçilerek elde edilen kestirimler karşılaştırılarak incelenebileceğini belirtmişlerdir.

#### **1.8.4. Çok Boyutlu Madde Tepki Kuramı**

Madde tepki kuramına dayanan ilk çalışmalarda, bireyi tanımlayan parametrelerin tek bir boyutta değişkenlik gösterdiği varsayımı kabul ediliyordu (Lord ve Novick, 1968). Buna paralel olarak, MTK'ya dayalı geliştirilen BOB testi yöntemleri tek boyutluluk varsayımına dayanıyordu. Fakat bu varsayımın sağlanmadığı ve tek boyutluluğun sağlanmamasının sonuçlarını belirlemek için çeşitli çalışmalar yapılmıştır (Ackerman, 1989; Folk ve Green, 1989; Reckase, Ackerman ve Carlson, 1988). Bu çalışmalar ışığında çok boyutlu madde tepki kuramına dayanan çok boyutlu BOB testi yöntemleri geliştirilmiştir.

Testte yer alan maddeler ile bireyler arasındaki ilişki genellikle tek boyutlu MTK modellerinde tanımlandığından çok daha karmaşıktır. Bu yüzden, çok boyutlu MTK bireyin bir maddeye cevap verebilmesi için birden fazla yetenek ve beceriye sahip olması gerektiği durumlarda kullanılması amacıyla geliştirilmiş bir ölçme modelidir (Chen, 2012).

Eğer test birden fazla boyutu ölçüyorsa, tek boyutlu modele uyum sağlayamayabilir ve yerel bağımsızlık varsayımının sağlanabilmesi için çok boyutlu modellere ihtiyaç duyulur. Çok boyutlu MTK'da bireyin yeteneği çok boyutlu uzayda her bir yapı (yetenek) bir düzlem ile temsil edilecek şekilde tanımlanır. Dolayısıyla, çok boyutlu MTK, bireyin çok boyutlu uzaydaki yeri ile bireyin bir maddeye doğru cevap verebilme olasılığını ilişkilendiren bir matematiksel modele dayanır (Reckase,2009). Ayrıca çok boyut MTK modelleri, ölçülen özellikleri ve testin karmaşık yapısını temsil eden ve daha güvenilir sonuçlar veren bir modeldir. Dolayısıyla klasik ölçme modellerinden elde edilen bilgilerden daha detaylı ve kapsamlı bilgi sağlar. Örneğin, çok boyutlu MTK modelleri ile model uyumuna bakılarak, bir testin hangi özellikleri ölçtüğü ve bireyin bir maddeye cevap verebilmesi için hangi özelliklere sahip olması gerektiği ya da maddeye verilen cevapların performansının ölçülen özellikleri nasıl etkilediği incelenebilir.

Günümüzde eğitimde kullanılan ölçme araçları farklı formattaki maddelerden oluşabilmektedir. Örneğin bazı ölçme araçları hem açık uçlu hem de çoktan seçmeli sorular içermektedir. Dolayısıyla, farklı özellikleri ölçen ve farklı formatta maddeleri içeren testlerin, testin karmaşık yapısına uygun olan çok boyutlu modellerle ölçülmesi daha kapsamlı ve güvenilir sonuçlar elde etmemizi sağlar.

#### **1.8.5. Çok Boyutlu MTK Modelleri**

Çok boyutlu MTK modelleri telafi-edici (Compensatory) ve telafi-edici olmayan (noncompensatory) modeller olmak üzere ikiye ayrılır (Sijtsma ve Junker, 2006). Telafi-edici modellerde bir bireyin ölçülen bir alt boyuttan elde ettiği düşük puan diğer boyutlardan elde ettiği yüksek puanlarla telafi edilir. Telafi-edici olmayan modellerde bu durum söz konusu değildir. Hangi çok boyutlu modelin kullanılacağı ise testte ölçülen özelliklerin yapısına bağlıdır. Eğer boyutlar arasındaki ilişki yüksek ve bir boyuttan elde edilen düşük yetenek parametresi diğer boyutlardan elde edilen yüksek yetenek parametreleri tarafından telafi edilebiliyor ise, telafi-edici modellerin kullanılması gerekir.

Telafi-edici çok boyutlu MTK modeli, 3PL MTK modelinin testin bir den fazla özelliği ölçtüğü çok boyutlu duruma genelleştirilmiş şeklidir (Babcock, 2009). Aşağıda telafi-edici çok boyutlu 3PL MTK modeline (3PL –MIRT) ait olasılık yoğunluk fonksiyonun formülü verilmiştir:

$$P(U_{ij} = 1 | \theta_j, a_i, c_i, d_i) = c_i + (1 - c_i) \frac{e^{(a_i \theta_j' - b_i)}}{1 + e^{(a_i \theta_j' - b_i)}} \quad (1.4)$$

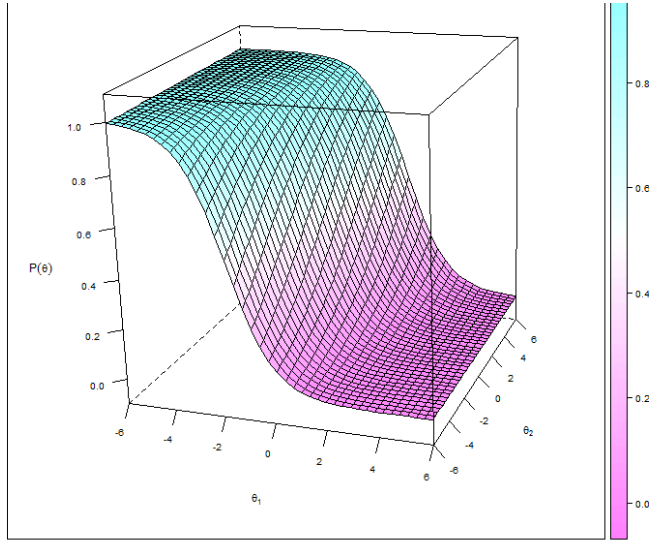
Burada  $\theta_j$ , m yetenek düzeyinin ölçüldüğü  $1 \times m$  şeklinde bir vektördür. Benzer şekilde  $a_i$   $1 \times m$  şeklinde ayırt edicilik vektörü olup  $b_i$  ve  $c_i$  skalar değerleri sırasıyla madde güçlük ve şans parametresini vermektedir (Weiss ve Yoes, 1991). Bu modele alternatif olarak geliştirilen bir diğer modelde ise  $b_i$  parametresi yerine  $d_i$  kesişim parametresi kullanılır. Bu modele ait formülü ise aşağıda verilmiştir:

$$P(U_{ij} = 1 | \theta_j, a_i, c_i, d_i) = c_i + (1 - c_i) \frac{e^{(a_i \theta_j' + d_i)}}{1 + e^{(a_i \theta_j' + d_i)}} \quad (1.5)$$

Burada  $d_i$  parametresi  $-a_i b_i$  değerine eşit olup kesişim ya da madde kolaylığı ( item easiness ) parametresi olarak adlandırılmaktadır (Reckase, 2009). Çok boyutlu MTK modelleri genellikle şans parametresinin olmadığını ya da şans parametresinin sıfıra eşit olduğunu varsaymaktadır ( $c_i=0$ ). Bu modellerde maddelere ait her bir boyuta ilişkin madde ayırt edicilik parametresi sıfırdan büyük olduğunda ( $a_{mi} > 0$ ), diğer boyutlara bakılmaksızın her hangi bir boyuta ilişkin bireyin yeteneğinin yüksek düzeyde olması bireyin maddeye doğru cevap verme olasılığını 1'e yaklaştırır. Bazı testler için bu durum anlamlı sonuçlar verirken, bazı durumlarda telafi-edici modellerin kullanılması anlamlı olmaz. Örneğin, bireyin hem okuduğunu anlama hem de işlem yapma becerisini ölçen bir madde için, ölçülen özelliklerin sadece birine sahip olunması maddeye doğru cevap verebilmek için yeterli olmaz. Okuma becerisindeki eksiklik, işlem yapma becerisi ile telafi edilemez. Böyle bir durumda, telafi-edici olmayan (noncompensatory) modellerin kullanılması daha gerçekçi sonuçlar verir.

Tek boyutlu MTK'da her bir madde için bireylerin yetenek düzeyleri ile modele ait olasılık fonksiyonları aracılığıyla hesaplanan maddelere cevap verme olasılığı arasındaki ilişki madde karakteristik eğrisi ile gösterilir. Çok boyutlu MTK modellerinde ise yetenek parametreleri ile maddeye doğru cevap verme olasılığı arasındaki ilişki çok boyutlu uzayda bir yüzey ile gösterilmektedir. Bu yüzeyin şeklini ise çok boyutlu MTK modelleri ile hesaplanan madde parametreleri ve yetenek parametreleri vektörü belirler.

Şekil 1.2'de testin iki boyutlu olduğu durumda testteki maddeye ait madde karakteristik yüzeyine bir örnek verilmiştir.



**Şekil 1.2: Çok boyutlu madde karakteristik yüzeyi**

Telafi-edici olmayan modellerde, bireyin bir maddeye doğru cevap verebilme olasılığının yüksek olabilmesi için testle ölçülen özelliklerin her birine belli bir düzeyde sahip olması gerekir. Bir yetenek düzeyindeki artışın diğer yetenek düzeylerindeki eksikliği ne kadar telafi edeceği noktasında bir sınırlılık vardır.

Telafi-edici çok boyutlu modelde toplamaya dayalı olasılık fonksiyonu kullanılırken, telafi-edici olmayan modelde ise çarpmaya dayalı olasılık fonksiyonu kullanılır. Telafi-edici olmayan modelde her boyut için tanımlanmış ayrı bir olasılık terimi vardır ve madde parametreleri kullanılarak elde edilen olasılık değerleri çarpılıp kestirilen doğru cevaplama olasılığı hesaplanır (Reckase ve McKinley, 1991). Telafi-edici olmayan çok boyutlu modele ilişkin olasılık yoğunluk fonksiyonu aşağıda verilmiştir:

$$P(U_{ij} = 1 | \theta_j, a_i, c_i, d_i) = c_i + (1 - c_i) \prod_{j=1}^J \frac{e^{[a_i(\theta'_j - b_i)]}}{1 + e^{[a_i(\theta'_j - b_i)]}} \quad (1.6)$$

Burada tanımlanan  $b_i$  parametresi m boyutlu bir test için  $1 \times m$  şeklinde her bir boyuta ilişkin madde güçlüğü veren bir vektördür. Bu modelde birçok olasılık değeri hesaplanır ve bu olasılık değerleri çarpılarak genel cevaplama olasılığı elde edilir. Telafi-edici modelin aksine, telafi-edici olmayan çok boyutlu MTK modellerinde madde ayırt edicilik parametresi gibi madde güçlük parametreleri de testin her bir boyutu için tanımlıdır (Chen, 2012).

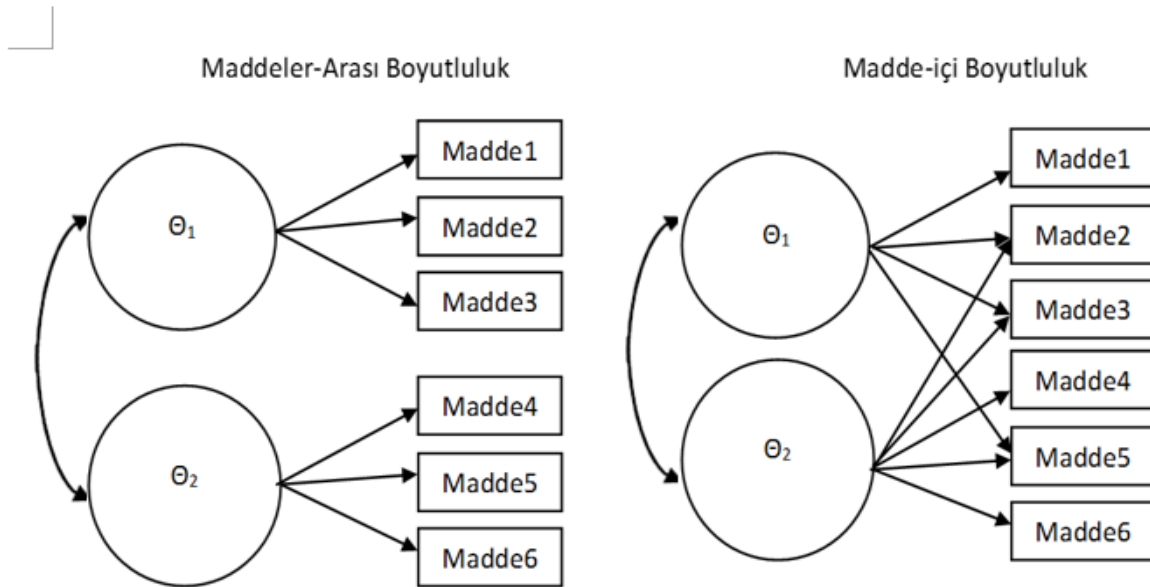
Araştırmacıların bu modeli telafi-edici olmayan model olarak tanımlamasının sebebi bir boyuttaki yüksek yetenek parametresi ( $\theta$ ), diğer boyutlardaki düşük yetenek



parametrelerini telafi edemez. Herhangi bir boyuta ilişkin düşük  $\theta$  değeri, düşük çarpma terimi anlamına gelip, genel olasılık değerini sıfıra yaklaştırır (Spray, Davey, Reckase, Ackerman ve Carlson, 1990).

Telafi-edici ve telafi-edici olmayan çok boyutlu modellerin dışında dikkat edilmesi gereken bir diğer nokta ise madde düzeyinde boyutluluktur. Madde düzeyindeki boyutluluk ise maddeler-arası boyutluluk (between-item dimensionality) ve madde-içi boyutluluk (within-item dimensionality) olmak üzere ikiye ayrılır (Wang, Chen, ve Cheng, 2004; Wang, Wilson ve Adams, 1997). Maddeler-arası boyutlulukta her bir madde sadece bir boyutta yük verir. Bu modelde maddelere ait ayırt edicilik parametreleri bir boyuttan sıfırdan farklı değer alırken diğer boyutlara ait ayırt edicilik parametresi sıfıra eşittir. Buna karşın, Madde-içi boyutluluk modelinde maddeler, birden fazla boyutta yük verir. Bu modelde maddelere ait madde ayırt edicilik parametresi ve madde yükleri diğer boyutlar içinde sıfırdan farklı değerler alabilir. Hartig and Höhler (2009) ise faktör analizine ait yük matrisindeki farklılıktan dolayı, maddeler-arası boyutluluk modelini basit yapıli model (simple structure model ve madde-içi boyutluluk modelini ise karmaşık yapıli model (complex structure model) olarak adlandırmıştır.

Şekil 1.3'te iki farklı boyutu ölçen bir test için telafi-edici çok boyutlu modellere ait madde düzeyinde boyutluluk modellerinden maddeler-arası ve madde-içi boyutluluk modelleri gösterimi şekillerle gösterilmiştir.



Şekil 1.3: Maddeler-arası ve madde-içi boyutluluk modeli (Wang ve Chen, 2004)

Bu çalışmada kullanılan telafi-edici çok boyutlu MTK modeli, maddelere ait şans başarısının sıfıra eşit olduğu kabul edilen iki-parametrelili çok boyutlu MTK modelidir (2PL-ÇBMTK/2PL MIRT). Bu modele ait formül aşağıda verilmiştir:

$$P(U_{ij} = 1 | \theta_j, a_i, d_i) = \frac{e^{(a_i \theta_j' + d_i)}}{1 + e^{(a_i \theta_j' + d_i)}} \quad (1.7)$$

Burada  $\theta_j$ , m yetenek düzeyinin ölçüldüğü  $1 \times m$  şeklinde bir vektördür. Benzer şekilde  $a_i$   $1 \times m$  şeklinde ayırıcı edicilik vektörü ve  $d$  kesişim parametresi ya da madde kolaylığı parametresi olarak adlandırılan skaler değerdir.

### 1.8.6. MTK'ya dayalı madde ve test bilgi fonksiyonları

Madde tepki kuramında bilgi fonksiyonları önemli bir yere sahiptir. Bireyin ölçülen özelliğinin (kestirilen yetenek parametresinin) doğruluk derecesi ve ölçme sürecine karışan hatanın düzeyi hakkında bilgi sağlar. Testte yer alan her bir madde bireyin ölçülen özelliği hakkında belli bir oranda bilgi sağladığından maddenin sağlanmış olduğu bilginin düzeyi ya da miktarı bireyin yetenek düzeyi ile madde parametrelerinin birbirleri ile olan uyumluluğuna bağlıdır. Madde tepki kuramına ait 1PL model kullanıldığında madde bilgisini etkileyen tek faktör madde güçlük parametresidir. Buna karşın, 2PL ve 3PL model kullanıldığında diğer parametrelerde ( $a_i$ ,  $c_i$ ) madde bilgisini yada madde bilgi fonksiyonunu etkilemektedir. Her bir maddeye ait bilgiyi ölçmek için kullanılan fonksiyona madde bilgi fonksiyonu (Item Information Function) denir.

Testte yer alan maddelere ilişkin bilgi fonksiyonuna ait matematiksel formül istatistiksel kestirim teorisinde kullanılmak üzere ilk olarak R.A. Fisher (1925) tarafından ortaya atılmıştır (Kullback, 1959). Herhangi bir  $\theta$  parametresi ile temsil edilen ve  $f(\cdot; \theta)$  olasılık yoğunluk fonksiyonuna sahip bir  $Y$  gözlenen değişkenine ait Fisher'in bilgi fonksiyonu;

$$I(\theta) = E_{\theta} \left[ \frac{d \log f(Y; \theta)}{d \theta} \right]^2 \quad (1.8)$$

Formülü ile hesaplanır. Burada  $E_{\theta}$ ,  $f(\cdot; \theta)$  olasılık yoğunluk fonksiyonu altında kestirilen yetenek parametresinin türevinin beklenen değerini vermektedir. Fisher, maksimum olabilirlik kestirimi (MLE) ile elde edilen  $\theta$  parametresinin asimtotik varyansının,  $I(\theta)$  değerinin çarpmaya göre tersine eşit olduğunu göstermiştir. Ayrıca diğer yanlılık göstermeyen yetenek parametrelerini kestirme yöntemlerinden elde edilen değerler için varyansın alt sınırını vermektedir.

Fisher'in bilgi fonksiyonunun bir diğerk formülü:

$$I(\theta) = -E_{\theta} \left[ \frac{\partial^2 \log f(Y;\theta)}{\partial^2 \theta} \right] \quad (1.9)$$

şeklindedir. Bu formül ile olasılık yoğunluk fonksiyonun logaritmasının  $\theta$  yetenek parametresine göre ikinci dereceden türevinin beklenen değeri maddeye ait bilginin miktarını verir.

Fisher'in bilgi fonksiyonu test tek boyutlu olduğunda veya yetenek parametresi ( $\theta$ ) skaler bir değer olduğunda negatif değer almaz. Benzer şekilde test birden fazla boyuttan oluştuğunda yetenek parametresi birden fazla değer alan bir yetenek vektörüne dönüşür. Testin tek boyutlu olduğu durumda olduğu gibi çok boyutlu durumda da Fisher bilgi fonksiyonu negatif değerler almaz. Fisher bilgi fonksiyonu test tek boyutlu olduğunda varyansın alt sınırını belirlerken test birden fazla boyuttan oluştuğu durumda ise herhangi bir yansız (unbiased) yetenek kestirim yöntemi için kovaryans matrisinin alt sınırını belirler.

Fisher'in test bilgi fonksiyonu ise bir testte yer alan maddelerin her bir yetenek düzeyinde sahip olduğu madde bilgilerinin toplamına eşittir. Fisher'in test bilgi fonksiyonu aşağıda verilmiştir:

$$I_j(\theta_j) = \sum_j E_{\theta} \left[ \frac{d \log f(Y;\theta)}{d\theta} \right]^2 \quad (1.10)$$

Test bilgi fonksiyonunun en önemli özelliği yetenek parametresi hakkında yapılan kestirimlerin doğruluk derecesi (the degree of accuracy) ve ölçmenin standart hatasını hesaplamaya olanak sağlamasıdır. Test bilgi fonksiyonu ile ölçmenin standart hatası arasında ters bir orantı vardır. Diğerk bir ifade ile toplam test bilgisi arttıkça ölçmenin standart hatası azalır. Fisher'in test bilgi fonksiyonu aracılığı ile ölçmenin standart hatası:

$$SEM(\theta) = \sqrt{\frac{1}{I(\theta)}} = \sqrt{\frac{1}{E_{\theta} \left[ \frac{d \log f(Y;\theta)}{d\theta} \right]^2}} \quad (1.11)$$

formülü ile hesaplanır. Kestirilen yetenek parametrelerine ait varyans ise;

$$Var(\hat{\theta}) = \frac{1}{I(\hat{\theta})} \quad (1.12)$$

formülü ile hesaplanır.

Bilgi teorisinde bir diğ er önemli kavram, Solomon Kullback ve Richard Leibler (1951) tarafından geliştirilen Kullback-Leibler (KL) uzaklığı (sapması ya da bilgisi- divergence or information) fonksiyonudur. Bu bilgi modelinde asıl amaç iki ö rneklemi birbirinden ayıran en iyi testi oluşt urmaktır. Aslında KL uzaklığı ya da bilgisi, Neyman-Pearson teoremine göre en iyi güç değ erini veren olabilirlik oran testinin asimtotik kuvvetini verir. İki olasılık yoğunluk fonksiyonu ( $f_1$  ve  $f_2$ ) arasındaki KL uzaklığı:

$$KL(f_1, f_2) = \int f_1 \log \left[ \frac{f_1(y)}{f_2(x)} \right] d_{\mu}(y) \quad (1.13)$$

Formülü ile hesaplanır. Burada  $\mu$ , deđ işkenin sürekli oldü ğ u durumda Lebesgue değ erini verirken, deđ işkenin kategorik oldu ğ u durumda ise kategorik sayı değ erinin temsil etmektedir. KL uzaklık fonksiyonu negatif değ erler almaz ve her iki olasılık yoğunluk fonksiyonları birbirine eş it oldu ğ unda sıfıra eş it olur. Simetrik olmadığı ndan tam anlamıyla uzaklığı ifade etmez.

KL bilgi fonksiyonun en önemli kullanım alanlarından biri ise  $\theta$  yetenek parametresine ait iki farklı sonsal dağılım arasındaki uzaklığın hesaplanmasıdır. Bu durumda  $u_j$  bireyin maddelere verdiği cevapları temsil etmek üzere, birey yeni bir maddeye cevap verdiği yeni sonsal dağılım ile bir önceki dağılım arasındaki farkı hesaplayarak son maddeye verilen cevabın sağladığı bilgi miktarı hesaplanır. İki sonsal dağılım (posterior distribution) arasındaki uzaklığı veren KL fonksiyonu:

$$E_{p(\theta|u, u_j)} = \left[ \log \frac{p(\theta|u, u_j)}{p(\theta|u)} \right] \quad (1.14)$$

formülü ile hesaplanır (Mulder ve van der Linden, 2010; Wang ve Chang, 2011).

Fisher'in bilgi fonksiyonu ile KL bilgi fonksiyonunun en önemli ortak özelliđ i gözlenen deđ işkene (y deđ işkeni) yeni bağımsız gözlemler eklendiğ inde elde edilen bilgilerin toplanabilir olmasıdır.

Fisher'in bilgi fonksiyonundan yola çıkarak tek boyutlu testlere ait MTK'ya dayalı 1PL model için bir maddenin sağlayacağı en yüksek bilgi 0,25 değ erine eş it olup, bireyin maddeye doğru cevap verme olasılığı ile yanlış cevap verme olasılığ ının birbirine eş it oldu ğ u ( $P_i=Q_i=0,5$ ) noktadaki değ eridir. Diğ er bir ifade ile 1PL modelde herhangi bir maddenin en yüksek bilgiyi sağlayabilmesi için madde güçlük parametresi ile bireyin  $\theta$  yetenek parametresi bir birine eş it olması gerekir. Bireyin yetenek parametresi madde güçlük parametresinden büyük ya küçük oldu ğ u durumda maddeye ait bilgi azalır.

Dolayısıyla bireylerin ölçülen özellikleri hakkında daha güvenilir ölçümler yapabilmek için testlerin farklı güçlük düzeyine sahip maddelerden oluşması gerekir.

2PL modelde madde ve test bilgi fonksiyonlarının madde ayırt edicilik parametresi ( $a_i$ ) ile doğru orantılıdır. Dolayısıyla, maddenin ayırt edicilik parametresi arttıkça teste ait bilginin miktarı artmaktadır. 3PL modele ait madde ve test bilgi fonksiyonlarının hesaplanması diğer iki modelle karşılaştırıldığında daha karmaşıktır. 2PL modelde olduğu gibi 3PL modelde madde ve test bilgi fonksiyonlarının madde ayırt edicilik parametresi ( $a_i$ ) ile doğru orantılı olduğu görülmektedir. Ancak, madde ve test bilgi fonksiyonu şans parametresi ( $c_i$ ) ile ters orantılıdır. Dolayısıyla, 3PL modelde maddenin ayırt edicilik parametresi arttıkça teste ait bilginin miktarı artmakta iken, maddelere ait şans parametresi arttıkça teste ait bilginin miktarı azalmaktadır.

Madde bilgi fonksiyonu ile ölçmenin standart hatası ters orantılıdır. Ayrıca yetenek parametresi ortalamaya yaklaştıkça teste ait toplam bilginin en yüksek değere ulaştığı, uçlara doğru gittikçe teste ait bilginin azaldığı vurgulanmaktadır. Buna karşın, yetenek parametresi ortalamaya yaklaştıkça teste ait standart hatanın en düşük değere ulaştığı, uçlara doğru gittikçe standart hatanın azaldığı belirtilmektedir. Genel olarak, tek boyutlu MTK'da ölçmenin standart hatası yetenek parametresine ait dağılımın uç noktalarındaki değeri, dağılımın orta noktalarındaki değerlerinden daha yüksektir ( Yao, 2014).

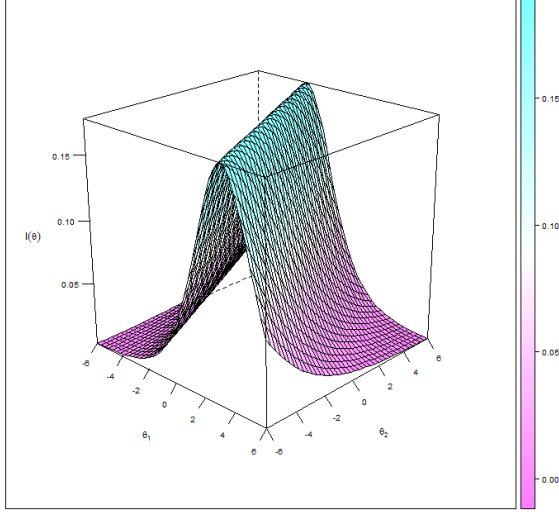
Madde tepki kuramında ölçmenin güvenilirliği (measurement precision) madde ve test bilgisine bakılarak değerlendirilir. Teste ait bilgi arttıkça yetenek parametrelerine ait varyans azalır, dolayısıyla, testin güvenilirliği artar. Bu durum test çok boyutlu olduğunda da geçerlidir ve çok boyutlu bilgi fonksiyonu ölçmenin güvenilirliğini verir. Ayrıca, tek boyutlu MTK modellerinde olduğu gibi, çok boyutlu modellerde de bilgi fonksiyonu ile ölçmenin standart hatası ters orantılıdır.

Çok boyutlu madde bilgi fonksiyonu, tek boyutlu MTK modellerine benzer şekilde hesaplanmaktadır. En önemli farklılık çok boyutlu MTK modellerinde bilgi fonksiyonun yönünde göz önünde bulundurulması gerekir (Ackerman, Gierl ve Walker, 2003)

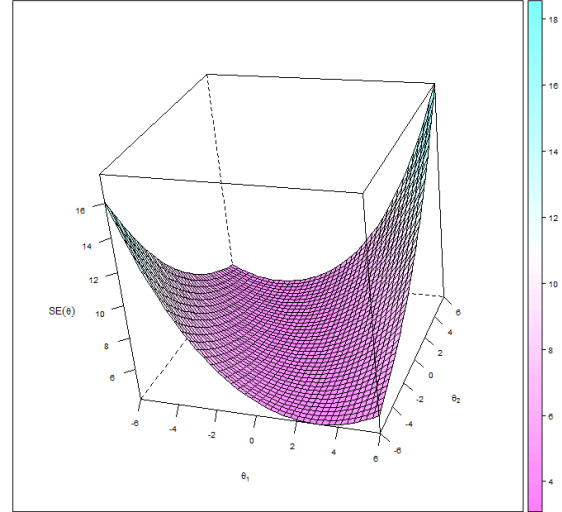
Tek boyutlu testlerde madde ve test bilgi fonksiyonları bir eğri ile gösterilirken çok boyutlu modellerde ise bir yüzey ile temsil edilmektedir. Çok boyutlu test bilgi fonksiyonu ise testte yer alan maddelere ait madde bilgi fonksiyonları ile elde edilen değerlerin toplamına eşittir.

Şekil 1.4'te çok boyutlu testlere ait madde bilgi ve test bilgi fonksiyonu yüzeyi ile madde ve teste ait standart hata yüzeyi verilmiştir. Madde ve test bilgi fonksiyonları BOB testlerdeki madde seçim yöntemlerinin temelini oluşturduğundan ayrı bir başlık altında incelenmiştir.

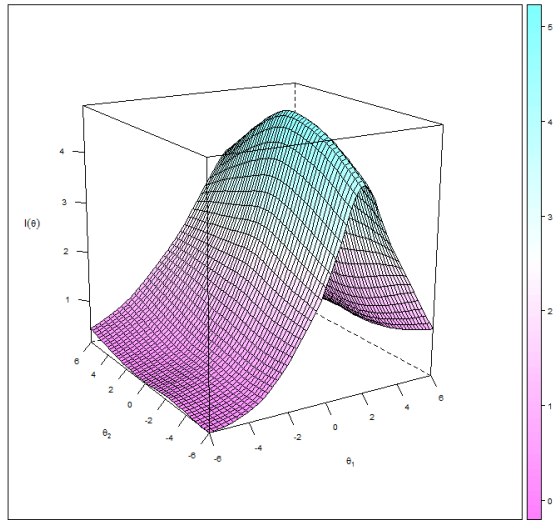
a) Madde bilgi yüzeyi



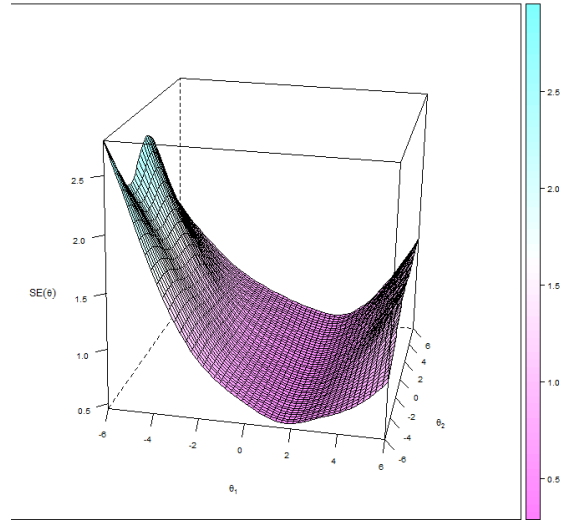
Madde standart hata yüzeyi



b) Test bilgi yüzeyi



Test standart hata yüzeyi



**Şekil 1.4: Çok boyutlu teste ait madde bilgi ve test bilgi yüzeyi grafiği**

Şekil 1.4 (a)'da sol taraftaki üç boyutlu grafik iki farklı özelliği ölçen maddeye ait bilgi yüzeyini gösterirken, sağ taraftaki üç boyutlu grafik ise ölçülen özelliklere ilişkin maddeye ait ölçmenin standart hatasını temsil etmektedir. Benzer şekilde, Şekil 1.4 (b)'de sol taraftaki üç boyutlu grafik iki boyutlu teste ait test bilgi yüzeyini gösterirken,

sağ taraftaki üç boyutlu grafik ise iki boyutlu teste ait ölçmenin standart hatasını temsil etmektedir. Tek boyutlu testlerde olduğu gibi teste ve maddeye ait bilgi fonksiyonu ile ölçmenin standart hatası ters orantılı olduğu görülmektedir. Yani maddeye ya da teste ait bilgi arttıkça standart hata azalmaktadır.

### **1.8.7. BOB Testlerinin Uygulama Basamakları**

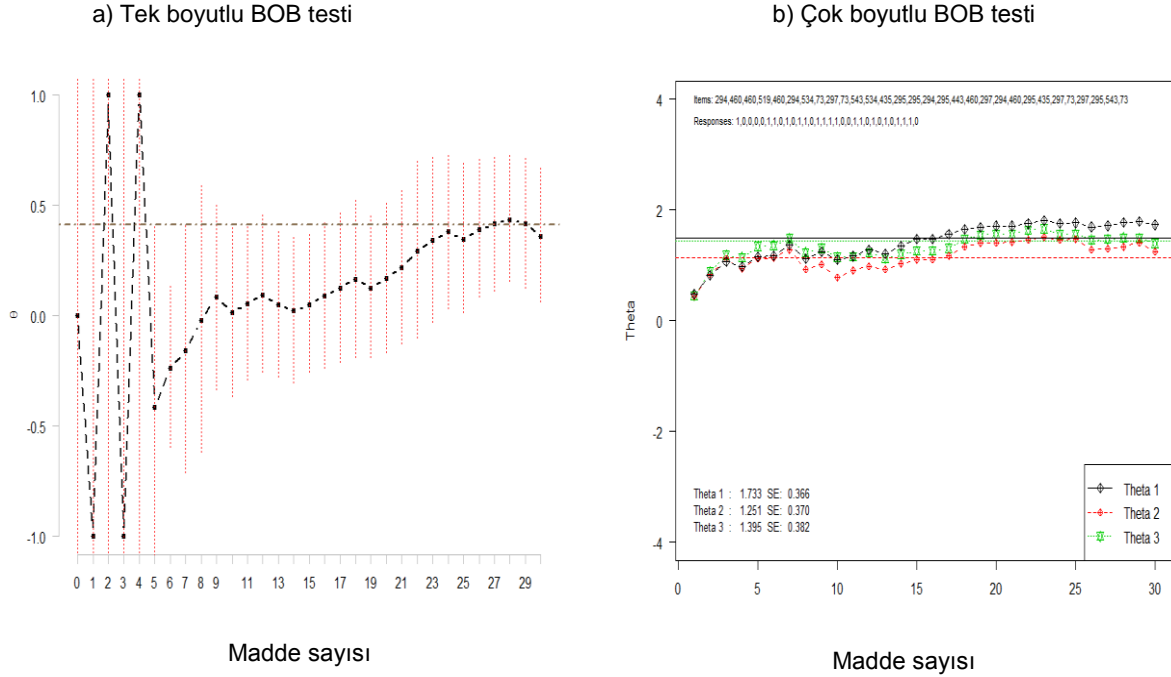
Binet testini uygulayan araştırmacı gibi, bilgisayar programı bireyin teste nasıl başlayacağına karar verir, bireyin önceki maddelere vermiş olduğu cevaplara bağlı olarak maddeleri seçer ve bir yâda birkaç kural belirleyerek testi sonlandırır. İlk BOB uygulamaları Binet'in kullanmış olduğu yöntemin farklı versiyonları iken (Weiss, 1973), sonraki uygulamalar ise madde havuzunun oluşturulma biçimine göre farklılık göstermektedir (Lord, 1971b, 1971c).

Burada bilgisayar bir çoklu ortam aracı gelişmiş grafik, ses vs. gibi özellikler sunmaktan çok daha fazlasını yapmaktadır. Bilgisayar önceden belirlenmiş bir soru grubunu sırayla uygulamak yerine, soruları testin gidişine göre seçerek sormaktadır. Çünkü bir bireye kendi yetenek düzeyinin çok üstünde (çok zor) sorular sormak ya da yetenek düzeyinin çok altında sorular sormak (çok kolay) testi alanın yetenek düzeyi fazla bilgi sağlamamaktadır. Ayrıca kişinin yetenek düzeyine uygun olmayan sorular sormak kişinin sıkılmasına da neden olabilmektedir (Sands, Waters ve McBride, 1997).

Bilgisayar ortamında bireye uyarlamış testlerde uygulama sürecinde öncelikle, bireyin düzeyine uygun zorluk derecesindeki soruları havuzdan çekip sorulur. Test sürecinde bireyin yetenek kestirimini en iyi biçimde yapabilmek için uygulama boyunca yetenek düzeyinde sürekli ayarlamalar yapılır. Bu ayarlamalar bireyin uygulamada o soruya gelene kadar gösterdiği performansa bağlıdır. Bir soru yanlış cevaplanırsa bir sonraki soru daha kolay; doğru cevaplanırsa sıradaki soru daha zor olmaktadır. Burada önemli nokta şudur: uygulanan sorular havuza eklenmeden önce yüzlerce hatta binlerce kişiye uygulandığından madde parametreleri bilinmektedir ve uygulanacak sorular bu parametrelere dayanarak seçilmektedir (Kalender, 2004).

Şekil 1.5 (a)'da tek boyutlu BOB testi analizi yöntemiyle bireyin yeteneğinin ölçülmesi süresince yanıtlamış olduğu sorular ve bireyin yetenek kestirim süreci görülmektedir. Öncelikle bireye yetenek parametresi ( $\theta$ ) sıfıra eşit olduğunda en yüksek bilgiyi veren maddeler madde havuzundan seçilerek sorulmaktadır. Daha sonra maddelere verilen cevaplara bağlı olarak daha kolay ya da daha zor sorular sorularak, bu süreç analiz

öncesi belirlenen durdurma kuralına ulaşıncaya kadar devam etmektedir. Maddelere verilen cevaplara bağlı olarak kestirilen yetenekler arasındaki hata düzeyi belirlenen ölçütün altına düştüğünde veya belli bir madde sayısına ulaşıldığında analiz program tarafından durdurulur. Şekil 1.5 (a)'da yer alan dikey kırmızı çizgiler ise kestirilen yetenek parametresine ilişkin güven aralığını vermektedir.



**Şekil 1.5: Tek boyutlu ve çok boyutlu BOB testleri ile bireyin yetenek kestirimi grafiği**

Şekil 1.5 (b)'de üç farklı yeteneğin ölçüldüğü çok boyutlu BOB testi yöntemiyle bireyin yeteneğinin ölçülmesi sürecinde yanıtlamış olduğu sorular ve bireyin her bir ölçülen boyuta ilişkin yetenek kestirimi süreci görülmektedir. Şekil 1.5 (b)'deki kırmızı, siyah ve yeşil eğriler ise her bir boyuta ilişkin kestirilen yetenek parametresini vermektedir. Öncelikle bireye ait yetenek parametreleri sıfıra eşit olduğunda en yüksek bilgiyi veren maddeler madde havuzundan seçilerek sorulmaktadır. Daha sonra maddelere verilen cevaplara bağlı olarak daha kolay ya da daha zor sorular sorularak, bu süreç analiz öncesi belirlenen durdurma kuralına ulaşıncaya kadar devam etmektedir. Maddelere verilen cevaplara bağlı olarak kestirilen yetenekler arasındaki hata düzeyi belirlenen değerin altına düştüğünde veya belli bir madde sayısına ulaşıldığında analiz program tarafından durdurulur.

BOB testlerin uygulama sürecine bakıldığında beş temel bileşenden oluştuğu söylenebilir (Weiss ve Kingsbury, 1984; Thompson, 2007). İlk bileşen testin uygulanacağı konu alanının belirlenmesi ve bu konu alanına ilişkin özelliği ölçtüğü



ifade edilen maddelerin madde tepki kuramına dayalı yöntemlerle kalibre edilip madde havuzunun oluşturulması. Diğer adımlar ise daha çok BOB test sisteminin işleyiş algoritmasına ait psikometrik özelliklerle ilgilidir. Bu aşamalar sırasıyla;

1. BOB testlerinin başlama koşulunun belirlenmesi
2. Maddelerin kalibre edilerek madde havuzunun oluşturulması
3. Test sürecinde madde seçim ölçütünün belirlenmesi
4. Puanlama algoritmasının belirlenmesi
5. Analizi durdurma kuralının belirlenmesi

Olarak tanımlanmaktadır (Weiss, 2011). Bundan sonraki bölümde BOB testlerinin uygulama basamaklarını oluşturan bileşenler hakkında bilgi verilecektir.

#### **1.8.7.1. BOB Testlerinde Başlama Koşulunun Belirlenmesi**

Bilgisayar ortamında bireyselleştirilmiş testlerde başlama koşulunun belirlenmesi, bireyin cevaplayacağı ilk maddenin belirlenmesi ve madde havuzundan seçilmesi sürecini belirler. Seçilecek ilk maddeyi belirleyebilmek için ise bireyin yetenek parametresini temsil eden bir başlangıç  $\theta$  değeri atanır. Başlangıç yetenek parametresini belirlemek için kullanılan farklı yöntemler vardır. En yaygın olarak kullanılan yöntem ise ortalama yetenek parametresine denk gelen bir sabit değer atamaktır. MTK' ya dayalı BOB testlerinde genellikle yetenek parametreleri birey merkezli ölçeklendiğinden bu değer 0 (sıfır)'a eşitlenir.

Bütün bireyleri aynı  $\theta$  değeri ile başlatmanın dezavantajlarından biri, bütün bireylere belirlenen yetenek parametresine ilişkin en yüksek bilgiye sahip aynı maddenin testin başlangıcında sorulmasıdır. Bu durum maddenin kullanım sıklığını (item exposure) artırır ve testin güvenliğini etkiler. Bu dezavantajları ortadan kaldırmak için, başlangıç yetenek parametresi herkes için belirlenmiş aynı sabit bir değer yerine, belirli bir aralıkta değişkenlik gösteren değerlerden seçkisiz olarak atanabilir. Bu dezavantajı ortadan kaldırmanın bir diğer yolu ise, madde kullanım sıklığı kontrol (item exposure control) yöntemlerini kullanarak bireylere farklı maddelerin sorulması sağlanabilir (Thompson ve Weiss, 2011).

Başlama koşulu olarak kullanılabilecek bir diğer yöntem ise, bireyler hakkında daha önce elde edilmiş ön bilgiler ya da dış faktörler kullanılarak madde seçimi belirlenebilir. Örneğin Castro, Suarez ve Chirinos (2010) motivasyon ve sosyo-ekonomik düzey gibi

dış faktörleri kullanarak her bir bireye ilişkin farklı başlangıç parametresi belirlemiştir (Thompson ve Weiss, 2011). Böylece testin başlangıcında farklı maddelerin seçilmesi ve maddelerin kullanım sıklığının azaltılması amaçlanmıştır.

Bu çalışmada, başlama koşulu olarak bireylerin her bir boyuta ilişkin yetenek parametrelerinin sıfıra eşit olduğu sabit yetenek parametresi yöntemi kullanılmıştır. Bireylerin testin başlangıcında farklı maddelere cevap vermelerini sağlamak ve maddelerin kullanım sıklığının azaltmak için madde kullanım sıklığı yöntemlerinden randomesque yöntemi kullanılmıştır. Böylece madde havuzundan maddeler seçilirken her bir birey için kestirilen yetenek parametrelerine ilişkin en yüksek bilgiyi veren ilk 10 madde içinden biri seçkisiz olarak seçilmesi koşulu ile maddeler seçilmiştir.

### **1.8.7.2. Madde Havuzunun Oluşturulması**

Bireyselleştirilmiş testlerin geliştirilme aşamalarından ilki belirlenen uygun ölçme modeli altında maddelerin kalibre edilerek madde havuzunun oluşturulmasıdır. Dolayısıyla madde havuzu bireyselleştirilmiş testleri oluşturan maddelerin seçildiği havuzdur. Ölçülmesi amaçlanan özelliklere göre maddeler yazılır ve MTK'ya dayalı modeller ile kalibre edilir ve ortak bir ölçeğe yerleştirilir (Song, 2010). Her bir bireyin ölçülen özelliklerine ilişkin güvenilir ölçümler yapabilmek için ranjı geniş güçlük parametresine sahip ve madde kalitesi yüksek olan maddelerin seçilmesi gerekir. Ayrıca her bir ölçülen özelliğe ilişkin yeterli sayıda maddelerin olması gerekir.

Bir madde havuzundaki maddelerin yeterli sayıda olabilmesi için ölçülen konu alanlarına ya da boyutlara ait maddelerin yeterli bir miktarda olması gerekir. Yeterli bir madde havuzunda bireyselleştirilmiş testteki madde sayısının 6 ile 12 katı kadar maddenin bulunması tavsiye edilmektedir (Luecht, 1998; Patsula ve Steffan, 1997; Stocking, 1998). Madde havuzundaki maddelerin sayısını etkileyen faktörlerin başında, madde kullanım sıklığı (item exposure) ve madde kullanım sıklığından kaynaklı madde kalitesinin düşmesi gelmektedir.

### **1.8.7.3. Madde Seçim Yöntemleri**

Tek boyutlu ve çok boyutlu bireye uyarlanmış testlerde kullanılan madde seçim yöntemleri birbirinden farklılık göstermektedir. Bu çalışmada çok boyutlu BOB testlerinde kullanılan madde seçim yöntemleri kullanıldığından, bu bölümde özellikle çok boyutlu BOB testlerinde kullanılan madde seçim yöntemleri hakkında bilgi verilmiştir

Tek boyutlu CAT analizinde birçok madde seçim yöntemleri olmasına rağmen en çok kullanılan yöntemler aşağıda sıralanmıştır.

Veerkamp ve Berger (1997) tarafından geliştirilen;

- Fisher'in maksimum bilgi (the maximum Fisher information-MFI) ölçütü ve
- Ağırlıklandırılmış maksimum olabilirlik fonksiyonu (the maximum likelihood weighted information-MLWI),

van der Linden (1998) tarafından geliştirilen;

- Maksimum sonsal ağırlıklandırılmış bilgi fonksiyonu (the maximum posterior weighted information -MPWI)
- Maksimum beklenen bilgi ölçütü (the maximum expected information –MEI criteiron)

Urry (1970) tarafından geliştirilen;

- Urry yöntemi (Urry's procedure)
- Minimum beklenen önsel varyans (the minimum expected posterior variance - MEPV) yöntemleri kullanılmaktadır.

Bu yöntemlerin yanında maddelerin tamamen seçkisiz (random) olarak belirlendiği yöntemler de vardır.

Çok boyutlu bilgisayar ortamında bireye uyarlanmış testlerde kullanılan madde seçim yöntemleri ile ilgili ilk çalışmalar Bloxom and Vale (1987) tarafından yürütülmüştür. Bloxom and Vale (1987) yapmış oldukları çalışmada Owen'ın (1969, 1975) tek boyutlu BOB testleri için geliştirmiş olduğu Bayesyen yakınsama madde seçim yöntemini (Bayesian approximation procedure) testin çok boyutlu olduğu durumlarda kullanımı için geliştirmişlerdir. Bloxom and Vale (1987)'in çok boyutlu BOB testleri ile ilgili yapmış olduğu çalışmayı Fan ve Hsu (1996), Luecht (1996), Segall (1996, 2000), van der Linden (1996, 1999), ve Veldkamp ve van der Linden'in (2002) yapmış olduğu çalışmalar takip etmiştir.

Çok boyutlu BOB testlerinde kullanılan madde seçim yöntemlerinden bazıları optimal desenlere bağlı madde seçim yöntemleridir. Optimal desenler bilgi matrisinin veya kovaryans matrisinin determinantı veya izinin istatistiksel çıkarımları optimize etmede kullanıldığı yöntemlerdir. Bu yöntemler sırasıyla D-optimality ve A-optimality olarak

adlandırılmaktadır (Silvey,1980, s. 10). Bu desenler istatistiksel çıkarımları optimize etmek için eğitim ve sağlık gibi birçok alanda kullanılmaktadır (Berger ve Wong, 2005). Fisher'in bilgi matrisi bu yöntemlerde önemli bir rol oynamaktadır. Çünkü Fisher'in bilgi matrisi gözlenen değişkenleri açıklayan gizil değişkenlere ait bilgiyi ölçmekte kullanılır.

Optimal desenler kullanılarak çok boyutlu BOB testleri için geliştirilmiş birçok madde seçim yöntemleri vardır. Bunlar; D-optimality (determinant of the posterior information matrix) (Luecht, 1996; Segall, 1996), A-optimality (minimizing the error variance of the composite measure) (van der Linden, 1999), C-optimality ve önsel dağılımın maksimize edildiği Kullback–Leibler bilgi fonksiyonun (Kullback–Leibler information-KLI) (Veldkamp ve van der Linden, 2002) yöntemleridir. Aşağıda çok boyutlu BOB testlerinde kullanılan madde seçim yöntemleri hakkında bilgi verilmiştir.

Çok boyutlu testler için,  $I_s(\theta)$  and  $I_i(\theta)$  sırasıyla test bilgi ve madde bilgi matrisi olsun. Çok boyutlu bireyselleştirilmiş testlerde, k.inci maddenin seçimi değerlendirilirken bu maddenin kestirilen yetenek parametresi ( $\hat{\theta}$ ) hakkında sağladığı bilgi daha önce cevaplanan k-1 maddeye ait bilgi matrisinin toplamı ile son maddeye ( $i_k$ ) ait bilgi matrisinin toplamına:

$$I_{Sk-1}(\hat{\theta}^{k-1}) + I_{ik}(\hat{\theta}^{k-1}) \quad (1.15)$$

eşittir (Mulder ve van der Linden,2009). Dolayısıyla, optimal desenlere dayalı madde seçim yöntemleri 1.15 formülüne uygulanır. Örneğin, optimal desenlerde bilgi matrisinin maksimum determinantı D-optimality ölçütü olarak adlandırılır (Segall, 1996). Optimal desenlere dayalı çok boyutlu BOB testlerinde kullanılan madde seçim yöntemleri sırasıyla verilmiştir.

#### 1.8.7.3.1. D-Optimality Madde Seçim Yöntemi

D-optimality madde seçim yönteminde formül 1.15'te verilen değerın determinantını maksimum yapan bir sonraki maddeyi seçer. Dolayısıyla, çok boyutlu testlerde bir sonraki madde:

$$arg \max_{ik \in R} \det(I_{Sk-1} + I_{ik}) \quad (1.16)$$

Ölçütüne göre seçilir. Bu madde seçim yöntemi göreceli olarak daha büyük asimtotik varyansa sahip yetenek parametresi için uygun olan yüksek ayırt edicilik parametresine sahip maddeyi seçer. Dolayısıyla, diğer boyutlara ilişkin yetenek parametresi ile karşılaştırıldığında daha yüksek varyansa sahip boyutun varyansını azaltacak maddeyi seçme eğilimindedir. Bu madde seçim yönteminde birden fazla boyutta yüksek ayırt edicilik parametresine sahip olan maddeler fazla bilgi sağlayamazlar. Sonuç olarak, D-optimality madde seçim yöntemi, birden fazla boyutu ölçen maddelerden ziyade sadece bir boyuta hassas olan maddeleri seçme eğilimindedir.

### 1.8.7.3.2. A-Optimality Madde Seçim Yöntemi

A-optimality madde seçim yönteminde yetenek parametrelerine ait asimtotik varyansın toplamının minimum değere indirgenmesi amaçlanır. Diğer bir ifadeyle, bilgi matrisinin tersinin izini (the trace of the inverse of the information matrix) minimuma indirgeyen maddeleri:

$$arg_{ik \in R}^{min} \text{trace}((I_{Sk-1} + I_{ik})^{-1}) = arg_{ik \in R}^{max} \frac{\det(I_{Sk-1} + I_{ik})}{\sum_{l=1}^3 \det([I_{Sk-1} + I_{ik}]_{[l,l]})} \quad (1.17)$$

formülünü kullanarak belirler. A-optimality madde seçim yönteminde D-optimality madde seçim ölçütünde olduğu gibi en önemli faktör bilgi matrisinin determinantıdır. Fakat bu madde seçim yönteminde bir sonraki madde seçilirken diğer yetenek parametrelerine ait varyans da dikkate alınır.

### 1.8.7.3.3. C-Optimality Madde Seçim Yöntemi

C-optimality madde seçim yöntemi, çok boyutlu testlerde her bir boyuta ilişkin yetenek parametrelerinin ağırlıklandırılmış bileşenin elde edildiği durumlarda kullanılır (Silvey, 1980, s. 11). Belirlenen ağırlık katsayılarına bağlı olarak tek bir yetenek parametresi hesaplanır. C-optimality madde seçim yönteminde  $\lambda$  her bir yetenek parametresine ilişkin ağırlık katsayısını veren vektör olmak üzere, çok boyutlu BOB testlerinde bir sonraki maddeyi seçmek için:

$$arg_{ik \in R}^{min} \lambda^T (I_{Sk-1}(\hat{\theta}^{k-1}) + I_{ik}(\hat{\theta}^{k-1})^{-1} \lambda) = arg_{ik \in R}^{max} (\lambda^T (I_{Sk-1}(\hat{\theta}^{k-1}) + I_{ik}(\hat{\theta}^{k-1})^{-1} \lambda)^{-1}) \quad (1.18)$$

Ölçütü kullanılır. Aslında bu madde seçim ölçütündeki  $\lambda$  ağırlık katsayıları, teste yer alan boyutların önem sırasını verir. Bu madde seçim yöntemi çok boyutlu testlerden her bir boyuta ilişkin yetenek parametresi kestirmek yerine, boyutların önemlilik

derecesinin farklılık gösterdiği ve test sonucunda tek bir yetenek parametresi elde edilmesinin amaçlandığı durumlarda kullanılması uygundur.

#### **1.8.7.3.4. E-Optimality Madde Seçim Yöntemi**

E-optimality madde seçim yöntemi bilgi matrisine ait en küçük Eigen değeri maksimum yapar. Diğer bir ifadeyle, testin boyutlarına ait yetenek kestirimlerinin genel varyansını maksimum yapar. Karmaşık yapısından dolayı bu madde seçim yöntemine ait formül burada verilmemiştir. Bu ölçütün dezavantajlarından biri ise kayıp verilerin olduğu ve veri miktarının yetersiz olduğu durumlarda güvenilir sonuçlar vermemesidir. Her bir boyut için eşit ayrıt edicilik parametresine sahip maddelerin katkısı, yetenek parametrelerine ait varyansın eşit olduğu durumda ortadan kalkar. Bu durum, yetenek kestirme yöntemlerine ait temel varsayım olan uygulanan madde sayısı arttıkça ortalama örneklem varyansının azalacağı kuralı ile çelişmektedir. Bundan dolayı çok boyutlu BOB testlerinde madde seçim algoritması olarak kullanılması tavsiye edilmemektedir (Mulden ve van der Linden, 2009).

Eğitim alanında çok boyutlu testlerde teste ait bilgi matrisi ile D-optimality ölçütü kullanılarak bireyin kestirilen yetenek puanına ait varyansı minimize eden madde kümesi seçilebilir. Bu seçilen maddeler kestirilen yetenek parametrelerine ait varyans için en küçük güven aralığını verir. Yani bireyin gerçek yetenek düzeyi ile kestirilen yetenek parametresi arasındaki fark en aza indirgenir. Madde seçim yöntemi olarak D-optimality yerine A-optimality yöntemi kullanıldığında seçilecek maddeler değişkenlik gösterir. Çünkü D- optimality sadece bireye ait kestirilen yetenek parametrelerinin varyansına bağlı olarak madde seçimi yapar.

Yukarıda belirtilen ve yaygın olarak kullanılan madde seçim yöntemlerinin dışında farklı madde seçim yöntemleri vardır. Bunlar; minimum açığı yöntemi (Reckase, 2009), hacim (volume, Segall, 1996), lineer kombinasyonların hata varyansının minimuma indirgeme yöntemi (van der Linden, 1999), komposit puanların hata varyansının minimuma indirgeme yöntemi (Yao, 2010b) ve Kullback–Leibler (KL) bilgi yöntemidir (Veldkamp ve van der Linden, 2002).

Bu çalışmada, farklı telafi-edici çok boyutlu BOB test yöntemlerinde madde seçim ölçütlerinden A-optimality, D-optimality ve seçkisiz (Random) madde seçim yöntemleri kullanılmış ve madde seçim yöntemlerinin performansları karşılaştırılmıştır.

#### 1.8.7.4. Yetenek Kestirim Yöntemleri

Bilgisayar ortamında bireyselleştirilmiş testlerde, bireye ait yetenek parametreleri bireyin maddelere verdiği her bir yanıtta sona güncellenir. Daha sonra bu güncellenen yetenek parametresine bağlı olarak daha önce belirlenen madde seçme ölçütü kullanılarak en uygun madde, madde havuzundan seçilir. Alan yazınına bakıldığında bir çok yetenek kestirme yöntemi olmasına rağmen en yaygın olan yöntemler maksimum olabilirlik kestirim yöntemi (MLE) ve Bayesyen yöntemlerdir (Bejar ve Weiss, 1979; Song, 2012).

##### 1.8.7.4.1. Maksimum Olabilirlik Kestirim Yöntemi (MLE)

MLE yetenek kestirim yönteminde amaç, bireyin maddelere verdiği cevap örüntüsüne ilişkin en yüksek olabilirliğe sahip yetenek parametresini kestirmektir. Madde parametrelerinin kestirildiği durumda bireyin cevap örüntülerine ilişkin olabilirlik fonksiyonu  $L(U_j|\theta_j)$  ile gösterilip;

$$L(U_j|\theta_j) = \prod_{i=1}^n P(u_{ij}|\theta_j) \quad (1.19)$$

formülü ile hesaplanır. Burada  $u_j$ ,  $j$  bireyinin maddelere verdiği cevapları içeren her bir bireye ilişkin cevap örüntüsünü verirken,  $\theta_j$  ise  $j$  bireyine ait kestirilen yetenek parametrelerini vermektedir. Ayrıca  $u_{ij}$  parametresi ise  $j$  bireyinin  $i$  maddesine vermiş olduğu cevabı göstermektedir.

Bireyin ölçülen yeteneğine ilişkin MLE yetenek parametresi,  $\hat{\theta}_j$ , olabilirlik fonksiyonunu maksimum yapan  $\theta$  değeridir. Olabilirlik fonksiyonuna ait en yüksek değeri veren  $\theta_j$  parametresi, olabilirlik fonksiyonun logaritmasının  $\theta$  -yetenek parametresine ait birinci türevi alınıp sıfıra eşitlenerek  $\hat{\theta}_j$  parametresi hesaplanır. Dolayısıyla MLE ile kestirilen yetenek parametreleri;

$$U(\theta; u_{ij}) = \frac{d \log(L(U_j|\theta_j))}{d\theta} \quad (1.20)$$

ile gösterilen eşitliği sıfır yapan ( $U(\theta; u_{ij}) = 0$ )  $\theta$  değerleridir. Bu eşitliği sağlayan  $\theta$  değerlerini hesaplamada kullanılan iteratif yöntemlerden biri Newton-Rapson yöntemidir. Burada  $U(\theta; u_{ij}) = 0$  eşitliğini çözmek için kullanılan Newton-Rapson iterasyonuna ait formül,  $k=1,2,3,\dots$ , olmak üzere:

$$\theta_{k+1} = \theta_k + \frac{U(\theta; u_{ij})}{I(\theta; u_{ij})} \quad (1.21)$$

formülü ile hesaplanır. Bu formülde paydadaki değer Fisher'in bilgi fonksiyonuna  $(I(\theta; u_{ij}))$  ait değerdir. Bu iteratif süreç belli koşullar altında  $\theta_0$  keyfi değeri atanarak başlar ve  $\hat{\theta}_j$  değerine yakınsayana kadar devam eder.

Alan yazınında MLE'nin genellikle daha az yanlılık göstermesine karşın Bayesyen yöntemlere göre daha büyük standart hata ve RMSE değerlerine sahip olduğu belirtilmektedir (Wang ve Vispoel, 1998; Warm, 1989; Weiss ve; McBride, 1984; Bock ve Mislevy, 1982). Bir diğer dezavantajı ise, BOB testlerinin başlangıç aşamasında bireyin bütün sorulara doğru cevap verdiği ya da yanlış cevap verdiği durumlarda yetenek kestirimi yapamamasıdır. Bu problemi çözmek için kestirilen yetenek parametreleri sınırlandırılabilir  $[-4,4]$  ya da Bayesyen yetenek kestirim yöntemleri gibi diğer yöntemler kullanılabilir (Song, 2010).

#### **1.8.7.4.2. Fisher'in Puanlama Metodu (Fisher's Scoring Method)**

MLE yetenek kestirim yöntemlerinde  $U(\theta; u_{ij}) = 0$  eşitliğini sağlayan  $\theta$  değerlerini hesaplamak için Newton-Rapson iterasyon yöntemine ait 1.21 formülü kullanılmaktadır. Newton-Rapson iterasyon yöntemine ait formülde paydadaki değer Fisher'in bilgi fonksiyonuna  $(I(\theta; u_{ij}))$  ait değerdir. Fisher'in Puanlama yönteminde ise  $U(\theta; u_{ij}) = 0$  yapan yetenek parametrelerini hesaplamak için Newton-Rapson iterasyon yöntemine ait formül 1.21'deki bilgi fonksiyonu yerine Fisher'in beklenen bilgi fonksiyonu kullanılmaktadır. Fisher'in beklenen bilgi fonksiyonu  $J(\theta)$  ile gösterilip:

$$J(\theta; u_{ij}) = E[I(\theta; u_{ij})] \quad (1.21)$$

Formülü ile hesaplanır. Dolayısıyla Fisher'in puanlama yöntemine ait formül:

$$\theta_{k+1} = \theta_k + \frac{U(\theta; u_{ij})}{J(\theta; u_{ij})} \quad (1.22)$$

Şeklindedir. Formül x bakıldığında Fisher'in puanlama yönteminde, Newton-Rapson iterasyon yöntemindeki Fisher'in bilgi fonksiyonu yerine Fisher'in beklenen bilgi fonksiyonun kullanıldığı görülmektedir.

Fisher'in puanlama yöntemi, Newton-Rapson yöntemi ile karşılaştırıldığında daha fazla iterasyon gerektirir. Fakat Fisher'in beklenen bilgisini hesaplamak Fisher'in bilgisini



hesaplamaktan daha kolay olduğundan yetenek parametresini hesaplama süresi daha kısadır. Ayrıca, Fisher'in puanlama metodunun yazılımlarda programlanması daha kolaydır. Testin çok boyutlu olduğu durumlarda ise benzer formül kullanılır. Tek farklılık  $\theta$  parametresi tek boyutlu testler için skaler bir değer iken, çok boyutlu testler için ölçülen boyut sayısı kadar skaler değer içeren bir vektördür.

#### 1.8.7.4.3. Bayesyen Olabilirlik Kestirim Yöntemleri

Bayesyen kestirim yöntemlerinde örnekleme ilişkin yetenek parametrelerine ait önsel bilgi (prior information) kullanılır. Bu varsayılan dağılım önsel dağılım olarak adlandırılır ve bireyselleştirilmiş test uygulamalarında önsel dağılımın ortalamasının 0, standart sapmasının 1'e eşit olan bir normal dağılım gösterdiği kabul edilmektedir. Yetenek parametresine ilişkin önsel dağılımın bilindiği durumda,  $\theta$  yetenek parametresine ilişkin sonsal dağılım;

$$h(\theta|U_j) = \frac{L(U_j|\theta) f(\theta)}{\int_{\theta} L(U_j|\theta) f(\theta) d\theta} \quad (1.24)$$

formülü ile hesaplanır. Burada  $L(U_j|\theta)$  olabilirlik fonksiyonunu ve  $f(\theta)$  MTK modellerine ilişkin olasılık yoğunluk fonksiyonunu vermektedir. Yetenek parametresine ilişkin sonsal dağılım,  $h(\theta|U_j)$ , bir sonraki maddeye verilecek cevabın önsel dağılımı olarak kabul edilir. Bireye ait enson yetenek kestirimine ulaşıncaya kadar bu döngüsel işlem devam eder (Song, 2010).

Bayesyen kestirim yöntemleri arasında ise en yaygın olanları MAP (maximum a posteriori), EAP (expected a posteriori) ve Owen'ın normal yaklaşım (Owen's normal approximation) metodudur. Bayesyen yetenek kestirim yöntemlerinden, *EAP yetenek kestirim yöntemi* sonsal dağılımın ortalamasını (beklenen değerini hesaplayarak) olarak yetenek parametresini hesaplar. EAP yöntemiyle kestirilen yetenek parametresi;

$$\hat{\theta} = E(\theta|U_j) = \int_{-\infty}^{\infty} \theta h(\theta|U_j) d\theta \quad (1.25)$$

formülü ile hesaplanır.

Samejima (1969) tarafından geliştirilen *MAP yetenek kestirim yönteminde*, bireye ait yetenek parametresi kestirilirken yetenek parametresine ait sonsal dağılımın,  $h(\theta|U_j)$ , modu hesaplanır. Diğer bir ifadeyle, sonsal olasılık yoğunluk fonksiyonunu

maksimum yapan deęer bireye ait MAP yetenek kestirimini verir. Bu deęer sonsal olasılık yoęunluk fonksiyonunun türevinin sıfır eřit olduęu noktadaki deęeridir.

Owen'ın (1975) normal yaklařım metodunda, sonsal daęılımın ortalama ve varyansının matematiksel formunu elde etmek için geręek yetenek parametresine ait sonsal daęılımın normal daęılım gösterdięi kabul edilir. Dolayısıyla, bu yöntemde sonsal daęılımın ortalaması bireyin kestirilen yetenek parametresi olarak kabul edilmiřtir (Song, 2010).

Bu alıřmada yetenek parametresi kestirme yöntemlerinden, Fisher'in puanlama yöntemi ve Bayesyen yöntemlerden MAP yetenek kestirim yöntemi kullanılmıřtır.

#### **1.8.7.5. BOB Testlerinde Kullanılan Durdurma Kuralları**

Bilgisayar ortamında bireyselleřtirilmiř testler, belirlenen durdurma kuralına ulařıldıęında son bulan döngüsel bir süreçtir(Reckase, 2009; Wainer, 2000). Bireyselleřtirilmiř testler daha önce belirlenen bir güvenilirlik düzeyine ulařıldıęında, sabit bir soru sayısına ulařıldıęında ya da kestirilen yetenek parametreleri belli bir güven aralıęında kestirildięinde durdurulur (Yao, 2012).

Durdurma kuralı olarak belli bir güvenilirlik düzeyine ulařma ölçütü kullanıldıęında, bireylerin test sürecinde cevapladıęı madde sayısı beklenenin çok üstünde olabilir ya da test süresi kâğıt-kalem testleri ile karřılařtırıldıęında beklenenden çok uzun olabilir. Segall, Moreno ve Better (1997), test uzunluęunun sabit olmadıęı bireyselleřtirilmiř test uygulamalarında, yetenek parametre daęılımının uç noktalarındaki bireylerin cevapladıęı madde sayısının çok daha fazla olduęunu vurgulamıřtır. Bu durum ise yorgunluęu ve dikkatsizlikten kaynaklı hata yapma oranını arttırmaktadır. Ayrıca bireyin kestirilen yeteneęi hakkında çok az bilgi saęlayan maddelerin sorulmasına neden olmaktadır.

Durdurma kuralı olarak sabit soru sayısı belirlendięinde ise bireyselleřtirilmiř testlerin uygulanması daha kolay olur ve test süresi kabul edilebilir bir aralıкта deęiřkenlik gösterir. Fakat bu durumda her bir birey için aynı güvenilirlik düzeyine ulařmak mümkün olamayabilir. Bu dezavantajların ortadan kaldırılması için sabit soru sayısı ve güvenilirlik ya da ölçmenin standart hatasına dayalı durdurma kurallarının birlikte kullanılması daha kararlı sonuçlar verebilir.

Çok boyutlu bireyselleřtirilmiř testlerde kullanılan bir dięer durdurma kuralı ise kestirilen yetenek parametrelerine ait hata varyansının belli bir deęerin altına

düştüğünde testin sonlandırıldığı standart hataya dayalı durdurma kuralıdır. Standart hata durdurma kuralı ise telafi-edici (compensatory) ve telafi-edici olmayan (conjunctive) olmak üzere ikiye ayrılır. Telafi-edici durdurma kuralında boyutlara ilişkin kestirilen yetenek parametrelerine ait hata varyansının birleşimi belli bir değerin altına düştüğünde test sonlandırılır. Telafi-edici olmayan durdurma kuralında ise her bir boyuta ilişkin kestirilen yetenek parametrelerine ait hata varyansının belirlenen ölçütün altına düştüğünde test sonlandırılır.

Bu çalışmada, çok boyutlu BOB testleri için geliştirilen sabit soru sayısı ve her bir boyuta ilişkin kestirilen yetenek parametrelerine ait hata varyansının belli bir değerin altına düştüğünde testin sonlandırıldığı standart hataya dayalı durdurma kuralı kullanılmıştır. Sabit soru sayısı durdurma kuralı için testteki madde sayısının 30, 40 ve 50 olduğu koşullar belirlenmiştir. Ayrıca bu çalışmada standart hata durdurma kuralından telafi-edici ve telafi-edici olmayan standart hata durdurma kuralları uygulanmıştır. Her iki standart hata durdurma kuralı için standart hata değerleri 0,20 0,25 ve 0,30 olarak belirlenmiştir.

#### **1.8.7.6. İçerik Ağırlıklandırılması**

Geleneksel kâğıt-kalem testlerinde, test geliştirme sürecinde testin uygulanacağı alanın içeriğinin özellikleri göz önünde bulundurularak testi oluşturacak maddeler belirlenir. BOB testlerinde bireyin kestirilen yetenek parametresine ait en yüksek bilgiyi veren maddeler seçildiğinden test ile ölçülmek istenen farklı içeriklere ilişkin madde dağılımı farklılık gösterebilir. Dolayısıyla, bireyler ölçülen her bir içeriğe ya da konu alanına ait farklı sayıda maddelere cevap verir. Örneğin matematik testini alan bireylerden biri sadece aritmetik sorularına cevap verirken, bir diğeri ise sadece geometri sorularına cevap vermek zorunda kalabilir. Bu durum BOB testinin geçerliğini, puanların karşılaştırılabilirliğini tehlikeye sokmakla birlikte, testi alan ve testi uygulayan bireyler açısından sorun oluşturabilir. Özellikle lisans ve sertifika testleri için, test sürecinde sorulan sorular test ile ölçülmek istenen bütün alanlara ilişkin maddeler içermediği durumda hukuki sorunlara neden olabilir (Kinsbury and Zara, 1991). Bilgisayar ortamında bireyselleştirilmiş testlerde, yukarıda belirtilen problemleri ortadan kaldırmak için ilk olarak Green ve arkadaşları (1984) tarafından içerik ağırlıklandırması (content balancing) görüşü ortaya atılmış ve ilerleyen yıllarda farklı içerik ağırlıklandırması yöntemleri geliştirilmiştir.

Wainer ve Kiely (1987), maddeleri tek tek sormak yerine maddelerin madde kümeleri (testlets) formatında sorulmasını; Kingsbury ve Zara (1989), geleneksel testteki içerik alanlarına ilişkin madde dağılımları dikkate alınarak en çok bilgi veren maddelerin seçilmesine olanak sağlayan sınırlandırılmış BOB testi yöntemini (constrained CAT) geliştirmiştir. Leung, Chang ve Hau (2000), Kingsbury ve Zara'nın geliştirmiş olduğu yönteme dayanan sınırlandırılmış BOB testi yöntemini ve Chen ve Ankenmann (2004) ise içerik ağırlıklandırmasına olanak sağlayan Modifiye edilmiş multinominal model (MMM) yöntemini geliştirmiştir (Song, 2010).

### **1.8.8. BOB Testlerinin Avantaj ve Dezavantajları**

Bireyselleştirilmiş testlerin en önemli avantajlarından biri bireyin yetenek düzeyine uygun soruları belirleyerek sormasıdır. BOB testi yöntemi, maddenin özellikleri ile bireyin yeteneğini eşleştirilerek puanların güvenilirliği artırır ve testin uzunluğunu azaltır (Lord, 1977; McBride Sz Martin, 1983; Weiss ve Kingsbury, 1984). Diğer önemli avantajı ise test koşullarının standartlaştırılabilir olmasıdır. Eğer BOB testi analizinde testi durdurma kuralı olarak sabit test bilgisi ya da yeteneğe ait standart hata durdurma kuralı seçilmişse bireylerin test sürecinde cevaplamış olduğu maddeler ve testteki madde sayısı farklı olmasına karşın testin güvenilirliği her bir birey için aynı kalır.

Bireye uyarlanmış test uygulamasının önemli üstünlükleri şöyle sıralanabilir:

- i. Testin uygulanma süresini kısaltır ve her öğrenci kendi seviyesine özgü bir test almasına olanak sağlar (Weiss,1983). Dolayısıyla test süresinin uzunluğundan kaynaklı sıkılma, yorgunluk ve performans düzeyinin düşmesi gibi faktörlerin etkisi azalır.
- ii. Geliştirilmesi, uygulanması ve test sonuçlarının rapor edilmesi daha kolaydır. Ayrıca bu sürece hata karışma olasılığı daha düşüktür (Song,2010)
- iii. Testler bireysel olduğu için testi alan bireylerin diğer bölümlerdeki sorulara cevap vermesi için testi alan diğer bireylerle birlikte hareket etmesine gerek yoktur (Rudner, 1998).
- iv. Kâğıt-kalem testleri genellikle ortalama yetenek düzeyine sahip bireyler için daha güvenilir sonuçlar verirken, BOB testleri geniş bir ranjdaki yetenek düzeyi için aynı güvenilirlikte ölçümler yapılmasına olanak sağlar.

- v. Test kâğıdı gereksinimlerini ve bundan kaynaklı güvenilirlik sorunlarını ortadan kaldırır.
- vi. Farklı formattaki maddelerin birlikte sorulmasına olanak sağlar. Ayrıca sorular görsel ve işitsel araçlarla desteklenebilir.
- vii. Madde havuzundan istenmeyen sorunların çıkartılması kolaydır. Soru seçiminde esnekliği artırır.

Bu üstünlüklerinin yanında bir takım dezavantajlar da mevcuttur:

- i. Her konu ya da yetenek için uygulanamaz
- ii. Bilgisayar donanımındaki sınırlılıkları ve maliyeti sorun olabilir,
- iii. Bilgisayar korkusu bu tür uygulamalarda sorunlar doğurabilir,
- iv. Her bir birey farklı soruları içeren testleri cevaplandıracağından, farklı testlerden kaynaklı eşitsizlik algısı oluşabilir.
- v. BOB testi uygulamalarında bireylerin genellikle testteki maddeleri tekrar incelemesi ve yanıtlarını değiştirmesine izin verilmez.
- vi. Ölçülen özellik ve boyut sayısına bağlı olarak madde havuzundaki madde sayısının yeterli olması gerekir (Chae, Kang, Jeon, ve Linacre, 2000; Rudner, 1998).

### **1.8.9 Bireyselleştirilmiş Testlerin Eğitimdeki Uygulamaları**

Yurtdışında bilgisayar ortamında bireyselleştirilmiş (BOB) testi uygulamaları konusunda çok geniş bir alan yazını bulunmaktadır. Ancak, geleneksel kâğıt-kalem testleri yerine BOB testlerinin kullanılabilmesi için, öncelikle bu test uygulamalarının daha verimli ve avantajlı olduğu, testin psikometrik özelliklerinin ve test sonuçlarına bağlı olarak yapılan değerlendirmelerin ve alınan kararların geçerliliğinin yapılan araştırmalarla desteklenmesi gerekir. Nitekim son yıllarda yapılan Monte Carlo simülasyon çalışmaları, gerçek veriye dayalı simülasyon (post-hoc simulation) çalışmaları ve bu simülasyona çalışmalarının sonuçlarına bağlı olarak geliştirilen gerçek BOB testi uygulamalarının sonuçları, BOB testlerinin kâğıt-kalem testleri yerine uygulanabilirliğini destekler niteliktedir (Weiss, 2004).

Bilgisayar ortamında bireyselleştirilmiş testler günümüzde GMAT (Graduate Management Admission Test), GRE (Graduate Record Examination), TOEFL (Test of

English as a Foreign Language) uygulamalarında kullanılmaktadır. Bu testlerde bireylere önceden hazırlanmış ve parametreleri bilinen sorulardan oluşan bir havuzdan sorular sorulmaktadır. Her soru bireyin önceki sorulardaki performansına göre belirlenmektedir. Bu testler belirli test merkezlerinde önceden randevu alınarak uygulanmaktadır.

GRE'de sayısal (quantitative) ve sözel (verbal) kısımların uygulanması için 58 soru kullanılmakta ve uygulama toplam 75 dakika sürmektedir. TOEFL uygulaması ise 165 ila 210 dakika almaktadır, ama bu süreye dinleme ve yazma kısımlarının uygulanması da dâhildir. Benzer biçimde GMAT uygulaması da 210 dakika sürmektedir. Bu sınavlar çok daha az soru kullanarak kısa sürelerde ölçtükleri yetenek düzeyi hakkında yüksek güvenilirliğe sahip bilgiler sağlamaktadırlar. Bu yaygın olarak kullanılan BOB testleri dışında, yurtdışında farklı yetenekleri ölçmek için geliştirilen BOB test uygulamaları arasında şunlar sayılabilir:

ADEPT (Adaptive English Proficiency Test for Web) testi CALEAP şirketi tarafından geliştirilmiş BOB testine dayalı İngilizce sınavıdır. ADEPT testinde maddeler, MTK'ya dayalı madde seçim yöntemleri ile maddeler seçilir ve test sürecinde madde kalibrasyonu yapan CBAT-2 algoritmasını kullanılır.

ACE (Adjustable Competence Evaluation) BOB testi mantıksal ve analitik çıkarım yapma yeteneğini ölçmeyi amaçlayan dokuz farklı BOB testi algoritmasından oluşmaktadır. Altı farklı dile uyarlanabilme özelliğine sahip ACE testi, işçi seçiminde ve işçilerin yeteneklerini geliştirme amaçlı kullanılmaktadır(<http://www.mastermanagement.net/Default.aspx?ID=294>).

CAA (Alberta Computer Adaptive Assessment System) BOB testi Alberta öğretmenlerinin kullandığı eğitim programına özgü, çevrimiçi değerlendirme olanağı sağlayan ve bireyselleştirilmiş testlerin öğretme ve öğrenme süreçlerinde kullanılmasını amaçlayan bir BOB testidir. Öğrencilerin dil, sanat, matematik, fen ve sosyal becerilerini ölçmek için geliştirilmiştir.

AICPA (American Institute of Certified Public Accountants) BOB testi Amerika'da muhasebe biriminde çalışacak bireylerin seçiminde kullanılan bireyselleştirilmiş testtir. AICPA testi bireyselleştirilmiş (adaptive) ve içerik ağırlıklandırılmasının yapıldığı madde kümelerinden (content-balanced testlets) oluşan çok aşamalı test yöntemini

kullanır. Bu BOB testinde seçme ve karar aşaması 20-25 maddeden oluşan madde kümelerine cevap verdikten sonra başlar (bk, [www.cpa-exam.org](http://www.cpa-exam.org)).

ASVAB (Armed Services Vocational Aptitude Test Battery ) BOB testi Amerikan askeri savunma birimi tarafından geliştirilmiş olup Savunma Biriminde çalışacak elemanların seçiminde kullanılmaktadır. ASVAB yetenek testleri; genel yetenek, kelime bilgisi, aritmetik mantık yürütme, okuduğunu kavrama, sayısal işlemler ve matematik bilgisi gibi yetenekleri ölçmeyi amaçlar (bk, <http://www.military.com/join-armed-forces/asvab>)

BULATS (Business Language Testing Service) BOB testi Cambridge Üniversitesi ESOL Sınavları, Alliance Française, Goethe-Enstitüsü ve Salamanca Üniveristesinin birlikte geliştirmiş olduğu bir bireyselleştirilmiş testtir. BULATS testi özellikle iş dünyasında bireylerin bireylerin İngilizce, Almanca, Fransızca veya İspanyolca dil becerilerini ölçmeyi amaçlayan kapsamlı bir bireyselleştirilmiş testtir. Özellikle son yıllarda bir çok kurum tarafından kullanılan bir dil yeterlik testidir (bk, [http://www.bulats.org/about/about\\_computer\\_test.html](http://www.bulats.org/about/about_computer_test.html))

The CAT (the CAT of Written English for Spanish Speakers) testi İspanyolca konuşan bireylerin İngilizce yazma becerilerini ölçmek amacıyla geliştirilmiş bir bireyselleştirilmiş testtir. The CAT testi Autonoma Madrid Üniversitesindeki bir grup araştırmacı tarafından geliştirilmiştir (bk, <http://www.iic.uam.es/en/solutions-and-services/competencies-assessment/ecat>).

Cito, Hollanda Eğitimde Ölçme Enstitüsü (Institute for Educational Measurement in The Netherlands) eğitimde kullanılmak amacıyla birçok BOB testi geliştirmiştir. Şu anda Cito tarafından geliştirilen ve uygulanmakta olan BOB testleri bireylerin matematik becerilerini, temel becerilerini ve coğrafya bilgisini ölçmektedir. Ayrıca, 4-6 yaş grubundaki öğrencilerin aritmetik becerilerini, dil becerilerini (kelime ve dinleme) ölçmek amacıyla dört farklı OB testi geliştirilmiştir. Bazı BOB testleri ise çocuklarda görülen disleksi ve erken okur-yazarlığı teşhis etmek için kullanılmaktadır (bk, [http://www.cito.com/en/research\\_and\\_development/computer\\_webbased\\_testing.aspx](http://www.cito.com/en/research_and_development/computer_webbased_testing.aspx))

J-CAT (Japanese Computerized Adaptive Test) BOB testi Japonca öğrenen bireylerin Japonca yabancı dil yeterliliklerini belirlemek amacıyla geliştirilmiş ağ-tabanlı bir testtir.

J-CAT, dinleme, kelime, dilbilgisi ve okuma olmak üzere dört bölümden oluşmaktadır (bk, [www.j-cat.org/en/](http://www.j-cat.org/en/))

MAP (Measures of Academic Progress ) testi Northwest Evaluation Association tarafından geliştirilmiş bir bireyselleştirilmiş testtir. MAP testi farklı yaş gruplarındaki öğrencilerin matematik, okuma, dil gelişimi ve fen bilimleri alanlarındaki performanslarını ölçmek amacıyla geliştirilmiş bir bireyselleştirilmiş testlerden oluşur. Amerika'da yaklaşık olarak 900 okul-grubu tarafından kullanılmaktadır (bk, [http://www.nwea.org/sites/www.nwea.org/files/resources/Introduction%20to%20MAP%20-%20digital\\_brochure\\_Aug13.pdf](http://www.nwea.org/sites/www.nwea.org/files/resources/Introduction%20to%20MAP%20-%20digital_brochure_Aug13.pdf))

Diğer eğitimde kullanılan ve gelişmekte olan bireyselleştirilmiş testler ise: Adaptive Matrices Test (AMT), CATE (Computerized Adaptive Test of English), ACT tarafından geliştirilen COMPASS test formları serisi, LPCAT (Learning Potential CAT), NAPLEX (North American Pharmacist Licensure Examination), NCLEX (National Council Licensure Examinations), CASAS (Comprehensive Adult Student Assessment System), CABT (Computer Adaptive Baseline Test), SAGE (Student Assessment of Growth and Excellence), STAR matematik, okuma ve erken yaş okuryazarlık testleri, SRI (Scholastic Reading Inventory) şeklindedir (iacat.org, 2015).



## 2. İLGİLİ ARAŞTIRMALAR

Alan yazınına bakıldığında bilgisayar ortamında bireyselleştirilmiş (BOB) testi yöntemlerine ilişkin yurt dışında yapılmış birçok çalışma bulunmasına karşın, ülkemizde yapılan çalışmalar sınırlı sayıdadır. Bu bölümde tek boyutlu ve çok boyutlu BOB testleri ile ilgili yapılan çalışmalara ilişkin bilgi verilmiştir.

### 2.1 BOB Testi İle İlgili Yurtdışında Yapılan Çalışmalar

McBride ve Martin (1983) testlerin kâğıt-kalem testleriyle geçerlilik ve güvenilirlik konularında karşılaştırılmasını içeren bir çalışma yapmıştır. Bu çalışmada, bireyselleştirilmiş testlerin kâğıt-kalem testlerindeki soru sayısının yarısından az soruyla yüksek seviyede güvenilir sonuçlar verdiği ve büyük ölçekli örneklerde bireyselleştirilmiş testlerin geçerliliğinin daha yüksek olduğu bulgusuna ulaşmıştır. Ayrıca, McBride ve Martin (1983) yapmış olduğu çalışmada, aynı uzunluktaki kâğıt-kalem testleriyle karşılaştırıldığında sözel yeteneğin ölçüldüğü 15 soruluk bireyselleştirilmiş testlerin güvenilirliğinin daha yüksek olduğu bulgusuna ulaşmıştır.

Bloxom ve Vale (1987), madde seçimi yöntemlerinde karmaşık hesaplamaları azaltmak ve Owen'ın (1975) aşamalı Bayesyen adaptive yetenek kestirimini çok boyutlu testlere uyarlamak amacıyla bir çalışma yapmıştır. Yapılan çalışma sonucunda, tek boyutlu BOB testleri için geliştirilen Owen'ın aşamalı Bayesyen adaptive yetenek kestirim yöntemi çok boyutlu testlerde kullanılabilir şekilde geliştirilmiştir.

Tam (1992), Bloxom ve Vale'in geliştirmiş olduğu madde seçim ve yetenek kestirim yöntemleri ile diğer yöntemleri ölçmenin güvenirliliği, test bilgisi ve hesaplama süresi açısından karşılaştırmıştır. Sonuç olarak, Bloxom ve Vale'in madde seçim ve yetenek kestirim yöntemlerinin güvenilirlik ve hesaplama süresi açısından üstünlüğünün madde havuzundaki maddeler yeterli olduğunda ve yetenek kestirim yöntemlerinden ağırlıklandırılmış maksimum olabilirlik yöntemi kullanıldığında telafi edildiği bulgusuna ulaşılmıştır.

Segall (1996) BOB testleri için Bayesyen istatistiklerine dayalı yetenek kestirimi ve madde seçim yöntemleri geliştirmiştir. Luecht (1996) ise Segall'in geliştirmiş olduğu yöntemi çok boyutlu testlerde içerik ağırlıklandırılmasının yapıldığı simülasyon

verilerini kullanarak çok boyutlu BOB testlerine uyarlamıştır. Ayrıca, içerik ağırlıklandırmasının yöntem üzerindeki etkisini incelemiştir.

van der Linden (1999) ise ölçülen her bir boyuta ilişkin ayrı yetenek parametreleri kestirmek yerine boyutlara ilişkin kestirilen yetenek parametrelerinin lineer kombinasyonundan oluşan bir yetenek parametresi veren ve maksimum olabilirlik kestirimlerinin varyansını minimize eden bir algoritma geliştirmiştir.

Eggen ve Straetmans (2000) bireylerin uygun dersleri seçmesi ve farklı gruplara sınıflandırılması için kullanılan yerleştirme testlerinin kâğıt-kalem testleri yerine BOB testlerinin kullanılmasını amaçlayan bir çalışma yapmıştır. Bu amaç doğrultusunda kâğıt-kalem testi ile tek boyutlu BOB testi performanslarını karşılaştırmıştır. Sonuç olarak, kâğıt-kalem testi yerine BOB testi kullanıldığında testi oluşturan madde sayısında %20 ile %40 oranında bir azalma olmasına karşın güvenilirlik katsayılarında anlamlı bir değişim olmadığı bulgusuna ulaşılmıştır.

Miller (2003) yapmış olduğu tez çalışmasında bireylerin başarılarını ölçmek amacıyla geliştirilen kâğıt-kalem testleri ile BOB testlerinin performanslarını karşılaştırmıştır. Bu amaç doğrultusunda kâğıt-kalem formatındaki başarı testini alan bireylerin yetenekleri BOB testi yöntemleri ile ölçülmüştür. Sonuç olarak kâğıt-kalem testi sonucu kestirilen yetenek parametreleri ile BOB testi sonucu kestirilen yetenek parametreleri arasında yüksek ve anlamlı bir korelasyonun olduğu bulgusuna ulaşılmıştır. Ayrıca, bireylerin test formatı tercihinde anlamlı bir farklılığın olmadığı bulgusuna ulaşılmıştır.

Wang ve Chen (2004) farklı koşulların kullanıldığı simülasyon çalışmasında tek boyutlu ve çok boyutlu BOB testlerinin performanslarını karşılaştırmıştır. Tek boyutlu BOB testleri ile karşılaştırıldığında, test çok boyutlu olduğunda ve boyutlar arasındaki korelasyon yüksek olduğunda güvenilirliğin daha yüksek olduğu bulgusuna ulaşılmıştır. Benzer şekilde, Luecht (1996) ve Segall (1996) tek boyutlu ve çok boyutlu BOB testi yöntemlerini karşılaştırdığı çalışmada, çok boyutlu BOB testi yöntemlerinin %25 ile %40 oranında testteki madde sayısını azalttığı bulgusuna ulaşılmıştır.

Diao ve Reckase (2009) yetenek kestirim yöntemlerinden MLE (Segall 1996, 2000) ve Bayesyen MAP (Segall 1996, 2000) yöntemlerini, madde seçim yöntemlerinden A-optimality ve D-optimality ve Kullback-Leibler bilgisine dayalı madde seçim yöntemlerinin kullanıldığı çok boyutlu BOB testi yöntemlerini karşılaştırmıştır. Bu çalışmada, farklı yetenek kestirim ve madde seçim yöntemlerinin kullanıldığı çok

boyutlu BOB testleri test uzunluđu, kullanılan önsel bilgilerin dağılımı, boyutların belirlenme şekilleri koşulları altında karşılaştırmıştır. Sonuç olarak, testteki madde sayısı yeterli olduğunda farklı madde seçim yöntemlerinin ve yetenek kestirim yöntemlerinin kullanıldığı BOB testlerinin benzer sonuçlar verdiği bulgusuna ulaşılmıştır. Diğer taraftan madde sayısı az olduğunda Bayesyen yetenek kestirim yöntemlerinin daha güvenilir ve tutarlı sonuçlar verdiği ve D-optimality madde seçim yönteminin A-optimality madde seçim yöntemine göre daha iyi sonuçlar verdiği bulgusuna ulaşılmıştır.

Song (2010) tez çalışmasında çok boyutlu verinin tek boyutlu MTK modeline uyumunun içerik ağırlıklandırılmasının yapıldığı çok boyutlu BOB testi yöntemleri üzerindeki etkisini incelemiştir. İçerik ağırlıklandırılmasının yapılmadığı BOB testi ölçüt alınmış ve üç farklı içerik ağırlıklandırması yöntemleri (the constrained CAT-CCAT, the modified multinomial model-MMM ve the modified constrained CAT-MCCAT) ölçmenin güvenilirliği, madde havuzu kullanımı ve madde kullanım sıklığı açısından karşılaştırılmıştır. Analiz sonucunda, özellikle konu alanlarının ortalama güçlük düzeyleri farklılık gösterdiğinde içerik ağırlıklandırması yöntemlerinin benzer sonuçlar verdiği bulgusuna ulaşılmıştır. Ancak, madde havuzu kullanımı ve madde kullanım sıklığı açısından içerik ağırlıklandırma yöntemleri arasında anlamlı bir farklılığın bulunmadığı belirtilmiştir. Son olarak, MMM içerik ağırlıklandırma yönteminin madde havuzunu en verimli bir şekilde kullanılmasını sağladığı bulgusuna ulaşılmıştır.

Yoo (2011) yapmış olduğu çalışmada testlerin ölçtüğü boyut sayısının, boyutlar arasındaki korelasyonun derecesinin, madde seçim yöntemlerinin ve test desenlerinin çok-boyutlu BOB testleri üzerindeki etkisini incelemiştir. Bu çalışmanın bağımlı değişkenlerini ise yetenek kestirimlerindeki yanlılık değeri, yetenek sınıflandırmasının doğruluđu, bilinen yetenek parametrelerindeki iyileşme derecesi değişkenleri oluşturmaktadır. Diğer çok boyutlu BOB testi çalışmalarının aksine testteki boyut sayısındaki artışın analiz sonuçlarının etkilemediği bulgusuna ulaşılmıştır. Boyutlar arasındaki korelasyon arttıkça yetenek kestirimlerindeki iyileşmenin arttığı ve farklı koşullar altında Bayesyen madde seçim yöntemlerinin en iyi sonucu verdiği bulgusuna ulaşılmıştır. Ayrıca boyutlar arasındaki korelasyon ve testteki madde sayısı arttıkça Bayesyen ve Fisher madde seçim yöntemlerinin benzer sonuçlar verdiği bulgusuna ulaşılmıştır.

Lin (2012) çalışmasında farklı madde seçim yöntemlerinin ve çoktan seçmeli ve çok kategorili maddelerden oluşan farklı test desenlerinin çok boyutlu BOB testi yöntemleri üzerindeki etkisini incelemiştir. Bu amaç doğrultusunda, Fisher bilgisine dayalı madde seçim yöntemlerinden D-optimality, Kullback-Liebler bilgi indeksi, CEM (continuous entropy method) ve ortak bilgi (mutual information-MI) yöntemleri ile farklı test desenlerinin performansları karşılaştırılmıştır. Sonuç olarak, D-optimality madde seçim yönteminin farklı koşullar altında en iyi sonucu verdiği ve test yeterli uzunlukta olduğunda madde seçim yöntemlerinin benzer sonuçlar verdiği bulgusuna ulaşılmıştır. Ayrıca çok kategorili maddelerin çok boyutlu BOB testlerinin başında kullanılmasının ve madde havuzunun büyüklüğünün yetenek kestirimini etkilemediği bulgusuna ulaşılmıştır.

Makransky ve Glas (2013) yapmış olduğu çalışmada, geleneksel kâğıt-kalem testleri ile tek boyutlu ve çok boyutlu BOB testi yöntemlerini; güvenilirlik katsayıları, test uzunluğu ve boyutlar arasındaki korelasyon açısından karşılaştırmıştır. Analiz sonucunda, geleneksel yöntem ile karşılaştırıldığında çok boyutlu BOB testi yöntemlerinin testin güvenilirliğini arttırdığı, testteki madde sayısını azalttığı ve test sürecinde madde seçimini kolaylaştırdığı bulgusuna ulaşılmıştır. Ayrıca çok boyutlu test yöntemlerinin avantajlarının özellikle test kısa olduğunda açık bir şekilde görüldüğünü vurgulamaktadır. Ayrıca, güvenilirlik katsayısı aynı kalmasına karşın çok boyutlu BOB testinin testteki madde sayısını kâğıt-kalem testi yöntemine göre 3 veya 4 kat azalttığı bulgusuna ulaşılmıştır.

## **2.2 BOB Testi İle İlgili Yurtiçinde Yapılan Çalışmalar**

Ülkemizde bireyselleştirilmiş test uygulamaları hakkında yapılmış çalışmalara bakılırsa, Kaptan (1993) yapmış olduğu çalışmada daha önce uygulanan ÖSS sınavına yönelik geliştirilmiş BOB testi ile kâğıt-kalem testinin performanslarını karşılaştırmıştır. Çalışma sonucunda yetenek kestirim yöntemleri arasında anlamlı bir farklılık bulunmamasına karşın, BOB testi uygulamasının testteki madde sayısını ve testin uygulama süresini azalttığı bulgusuna ulaşılmıştır.

İşeri (2002) OKSYS (Ortaöğretim Kurumları Seçme Yerleştirme Sınavı) sınavına katılan öğrencilerin matematik başarılarını ölçmek amacıyla geliştirilmiş farklı BOB testi koşulları ile kâğıt-kalem testi formatının performanslarını karşılaştırmıştır. Sonuç olarak, BOB testi yönteminin kâğıt-kalem testi ile karşılaştırıldığında daha az madde ile daha güvenilir sonuçlar verdiği bulgusuna ulaşılmıştır. Ayrıca, hem sabit madde

sayısı durdurma kuralı hem de standart hata durdurma kuralı için Bayesyen yetenek kestirim yönteminin MLE yetenek kestirim yöntemine göre daha güvenilir sonuçlar verdiği belirtilmektedir.

Cömert (2008) farklı BOB testi yöntemlerinden çok düzeyli uygulama, iki aşamalı uygulama, değişilmemeli uygulama ve piramitsel çok aşamalı uygulama yöntemlerinin kullanılmasına olanak sağlayan bir yazılım geliştirilmiştir. Çalışma sonucunda geleneksel kâğıt-kalem testi ile karşılaştırıldığına BOB testi yöntemlerinin bireylerin yetenek düzeylerini daha az madde ile kestirdiği bulgusuna ulaşılmıştır. Ayrıca, çok düzeyli BOB test uygulamasında, testteki ortalama madde sayısının diğer BOB testi yöntemlerine göre daha fazla olduğu, puanlara ilişkin sapmaların ise daha az olduğu bulgusuna ulaşılmıştır. Piramitsel çok aşamalı BOB testi yöntemine ilişkin bulgulara bakıldığında, testteki madde sayısı yeterli olmadığında özellikle orta düzeydeki bireylerin yetenek kestirimlerine daha fazla hata karıştığı belirtilmiştir.

Erdoğan (2009), BOB testine ait madde havuzunu oluşturan maddelerin madde güçlük parametrelerinin farklı yöntemlerle kestirilmesine ilişkin bir çalışma yapmıştır. Bu amaç doğrultusunda, doğru yanıtlanma olasılığı, norm referanslı madde analizi ve madde tepki kuramı (MTK) yöntemleri ile madde güçlük parametrelerinin kestirilmesine olanak sağlayan bir program geliştirilmiş ve her bir yöntemle ait kestirilen parametreler karşılaştırılmıştır. Analiz sonucunda, BOB testi yöntemlerine ilişkin en iyi madde güçlük parametresi kestirim yönteminin norm referanslı madde analizi yöntemi olduğu bulgusuna ulaşılmıştır.

Tercan (2010) tarafından yapılan bir çalışmada BOB testi uygulamasına ve klasik test kuramına (KTK) dayalı bir testin elektronik ortamda uygulanmasına yönelik bir çalışma yapılmıştır. Ayrıca her iki yöntemin gerçek test uygulamalarında ve uzaktan eğitimde kullanılmasına ilişkin öğretmen görüşleri incelenmiştir. Araştırma sonucunda, öğretmen adaylarının BOB testi uygulamaları hakkında yeterli bilgilerinin olmamasına karşın, sınıf-içi etkinliklerinde BOB testi yöntemlerini kullanmak istediklerini belirtmişlerdir.

Kalender (2011), Öğrenci Seçme Sınavı (ÖSS) fen alt testinin BOB testi yöntemleri ile uygulanmasına yönelik bir çalışma yapmıştır. Bu çalışmada farklı yetenek kestirim ve test sonlandırma kurallarının kullanıldığı BOB testi yöntemleri ile fen alt testinin kâğıt-kalem formatı sonuçları kestirilen yetenek parametreleri, test uzunluğu, testin güvenilirliği açısından karşılaştırmıştır. ÖSS fen alt testi için en uygun BOB testi

yöntemine karar vermek için gerçek verilere dayalı simülasyon yönteminden yararlanılmıştır. Araştırma sonucunda, ÖSS fen alt testi için en uygun BOB testi algoritmasının hata varyansı durdurma kuralının 0.30, yetenek kestirim yöntemlerinden EAP yönteminin kullanıldığı BOB testi yönteminin olduğu bulgusuna ulaşılmıştır. Ayrıca, BOB testi uygulamasının testte madde sayısını %50 oranında azalttığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun katsayısının 0.74 olduğu bulgusuna ulaşılmıştır.

Kaskatı (2011), Romatoid Artrit hastalarının özürlülük düzeylerinin belirlenmesi amacıyla Rasch modellerine dayalı BOB testi geliştirmiştir. Romatoid Artrit hastalarının özürlülük düzeylerini ölçmek için geliştirilmiş maddelere ait parametreler Rasch modeli kestirilerek madde havuzu oluşturulmuştur. Kâğıt kalem testi ile Rasch modeline dayalı BOB yönteminin performansını karşılaştırmak amacıyla 29 hastaya hem kâğıt kalem formunda hem de BOB testi uygulanmıştır. Analiz sonucunda, Rasch modeline dayalı BOB testinin Romatoid Artrit hastalarının özürlülük düzeylerinin ölçülmesi için uygun olduğu ve kâğıt-kalem testi ile BOB testi uygulamaları sonucu kestirilen yetenek parametreleri arasında yüksek korelasyon olduğu gözlenmiştir.

Bulut ve Kan (2012) Akademik Personel ve Lisansüstü Eğitime Giriş Sınavı'nın (ALES) BOB testi yöntemi ile uygulanmasına ilişkin gerçek verilere dayalı simülasyon çalışması (post-hoc simülasyon) yapmışlardır. 2008 yılında uygulanan ALES sınavdan 10000 kişi rastgele olarak seçilmiş, madde ve yetenek parametreleri ise MTK'nın üç parametrelili lojistik modeli ile kestirilmiştir. Analiz sonucunda ALES sınavı için geliştirilen en uygun BOB testi yönteminin yetenek kestirim yöntemlerinden EAP yöntemi ve standart hata durdurma kurallarından 0.25, 0.30 ve 0.40 hata eşik değerinin kullanıldığı BOB testi algoritması olduğu bulgusuna ulaşılmıştır. Ayrıca, BOB testi uygulamalarının testteki madde sayısında %70'e varan azalma sağladığı ve kâğıt-kalem testi ile BOB testi uygulamaları arasında 0.93'ün üzerinde korelasyonunun olduğu bulgusuna ulaşılmıştır.

Kezer (2013) yapmış olduğu tez çalışmasında farklı başlama kuralı, farklı yetenek kestirim ve durdurma kuralının kullanıldığı BOB testi yöntemleri ile bireyin kelime bilgisini ölçmek için geliştirilmiş geleneksel kâğıt kalem test yönteminin performanslarını karşılaştırmıştır. Bu amaç doğrultusunda araştırmacı gerçek BOB testi uygulaması için çevirim içi ortam geliştirmiştir. Sonuç olarak, kâğıt-kalem testleri ile BOB testi sonucu kestirilen yetenek parametreleri arasındaki korelasyonun yüksek

olduđu ve BOB testi uygulamalarının daha az madde ile benzer gvenirlikte lmeler yaptığı bulgusuna ulařılmıştır. Ayrıca farklı BOB testi kořulları altında kestirilen yetenek parametreleri arasındaki korelasyonun da yksek olduđu ve durdurma kuralı olarak standart hatanın 0,50 olarak belirlendiđinde MLE yetenek kestirim yntemleri ile Bayesyen MAP ve EAP yetenek kestirim yntemlerinin benzer sonular verdiđi bulgusuna ulařılmıştır.

Sulak (2013) BOB testlerine iliřkin yapmış olduđu tez alıřmasında, BOB testi uygulamalarında kullanılan madde seme yntemleri, yetenek kestirim yntemleri ve test durdurma kurallarını karřılařtırmalı olarak incelenmiştir. Bu alıřmada toplamda 30 kořul karřılařtırılmıştır. MLE yetenek kestirimi kullanıldıđında elde edilen ortalama madde sayısı, EAP yetenek kestirimi kullanıldıđında elde edilen ortalama madde sayısından daha yksek bulunmuřtur. BOB testi uygulamalarında EAP yetenek kestiriminin test uzunluđunu kısalttığı sonucuna varılmıştır. Ayrıca, bu alıřmada BOB testlerinde kullanılan madde seme yntemlerinin madde havuzu kullanımına iliřkin iyi bir denge gstermediđi ve yksek  $\alpha$ -parametresine sahip maddeleri daha ok kullandıđı sonucuna varılmıştır.

Boztun-ztrk (2014) BOB testlerine iliřkin yapmış olduđu tez alıřmasında, eřitli madde kullanım sıklığı kontrol yntemlerinin madde seme yntemi ve madde havuzu zelliklerine gre lme kesinliđine ve test gvenliđine olan etkisinin incelemiřtir. Bu alıřmada madde kullanım sıklığı kontrol yntemleri olarak Bilgi Kořullu Tesadfi, Sympson Hetter ve Azalarak Kaybolma yntemleri kullanılmıştır. Madde seim yntemi olarak da Fisher'in En oklu Bilgi,  $\alpha$ -Tabakalama ve Ařamalı En ok Bilgi Oranı yntemleri kullanılmıştır. Sonu olarak, madde kullanım sıklığı kontrol yntemleri kullanıldıđında lme kesinliđi belirlemede kullanılan deđerlerde byk bir farklılık grlmediđi bulgusuna ulařılmıştır. Ayrıca, test gvenirliđi aısından ise Azalarak Kaybolma madde kullanım sıklığı kontrol ynteminin madde havuzunun kullanımındaki arpıklık ve test akıřmasının azaltılması aısından daha iyi sonu verdiđi bulgusuna ulařılmıştır.

## 3. YÖNTEM

### 3.1. Araştırmanın Yöntemi

Bu araştırmada, yabancı dil sınavına giren bireylerin yabancı dil yeteneklerinin telafi-edici Çok Boyutlu Bilgisayara ortamında bireyselleştirilmiş (Çok-Boyutlu BOB) test yöntemleri ile kestirilmesi ve çok boyutlu BOB testi yöntemlerinin performanslarının karşılaştırılması amaçlanmaktadır. Araştırmada var olan yöntem ve tekniklerin gerçek veri üzerinden sınanması amaçlandığından araştırma temel araştırma niteliğindedir (Karasar, 2009).

### 3.2. Çalışma Grubu

#### 3.2.1. Çalışma Grubunun Özellikleri

Araştırmanın çalışma grubunu Hacettepe Üniversitesinde İngilizce Yeterlik Sınavı (İYS)'na giren bireyler oluşturmaktadır. Her bir dönem sonunda ve dönem içerisinde İYS'ye giren öğrenci sayısı 1200 ile 2000 arasında değişmektedir. Araştırmanın örneklemini ise 2008-2013 eğitim-öğretim yıllarında ilkbahar ve sonbahar dönemlerinde Hacettepe Üniversitesinde İYS'ye giren bireyler oluşturmaktadır. Analiz sürecinde gerçek madde parametrelerine bağlı olarak bireylere ilişkin yetenek parametreleri türetilmiş ve her bir analiz için aynı yetenek parametreleri kullanılmıştır. Çok-boyutlu BOB testi sürecinde her bir koşul için yapılan analizlerde testi alan birey sayısı 500 ile sınırlandırılmıştır. Bireylerin her bir boyuta ilişkin yetenek parametreleri  $[-3, +3]$  aralığında çok değişkenli normal dağılım göstermektedir.

### 3.3. Araştırma Verileri

Araştırmanın verilerini, 2008-2013 eğitim-öğretim yıllarında ilkbahar ve sonbahar dönemlerinde Hacettepe Üniversitesinde İYS'ye ait testlerdeki maddeler ve sınava giren bireylere ait cevap örüntülerinin yer aldığı veri seti oluşturmaktadır.

İYS Hacettepe üniversitesini kazanan öğrencilerin İngilizce dil düzeylerini belirlemek amacıyla öncelikle sonbahar döneminin başlangıcına denk gelen eylül ayında yapılır. Daha sonra 1 yıl İngilizce hazırlık alacak öğrencilere sonbahar ve ilkbahar dönemlerinin sonunda olmak üzere kâğıt-kalem testi formatında iki kez uygulanır. Sınav dinleme (listening), okuduğunu anlama (reading) ve dilbilgisi (grammar) olmak üzere üç temel bölümden oluşup her bir testteki ortalama madde sayısı ise toplam 65'dir.



Sınavda aynı maddelerin farklı yerlerde olduğu iki kitapçık türü (A, B) kullanılmaktadır. Testteki maddeler çoktan seçmeli olup beş seçenektir. Araştırmada kullanılan veriler düzenlenirken, iki farklı kitapçık türüne göre elde edilen kişilerin cevapları “A” kitapçık türü temel alınarak 1-0 puan matrisine çevrilip birleştirilmiştir. Verilerin düzenlenmesinde herhangi bir hata yapıp yapılmadığı her bir bireyin teste ait toplam doğru sayısı hesaplanıp, elde edilen doğru sayıları ile karşılaştırılarak kontrol edilmiştir.

### 3.4. Verilerin Analizi

Verilerin analizi 2 aşamadan oluşmaktadır. İlk aşamada çok boyutlu bilgisayar ortamında bireyselleştirilmiş testlerde kullanılacak maddelere ait madde parametrelerinin maddeler-arası boyutluluk ve madde-içi boyutluluk modellerine dayalı tanımlayıcı çok boyutlu MTK modelleri ile kestirilmiş ve her bir modele ait madde havuzu oluşturulmuştur. Ayrıca her bir çok boyutlu model için model veri uyumu istatistikleri hesaplanarak testin tek boyutlu ve çok boyutlu olduğu durumdaki uyum istatistikleri karşılaştırılmıştır. İkinci aşamada ise post-hoc simülasyon yöntemi kullanılarak, her bir modele ilişkin en iyi sonuç veren çok boyutlu BOB testi algoritması belirlenmesi amacıyla farklı çok boyutlu BOB testi algoritmalarına ilişkin analiz sonuçları karşılaştırılmıştır.

#### 3.4.1 Veri Analizinin Birinci Aşaması

##### 3.4.1.1 Madde Parametrelerinin Kestirilmesi

İlk aşamada, okuma (reading), dinleme (listening) ve dilbilgisi (grammar) olmak üzere üç boyuttan oluşan son 5 yıla ait İngilizce Yeterlik Sınavı (İYS) veri setlerinde yer alan her bir boyutla ilişkili maddelerin boyutlarla ilişkili olup olmadığı doğrulayıcı faktör analizi ile belirlenmiştir. Daha sonra veri setlerinin üç boyutlu olarak kabul edildiği durum için R yazılımında tanımlı “mirt” paket programı aracılığı ile testteki maddeler telafi-edici çok boyutlu madde tepki kuramına (Compensatory 2PL-MIRT) dayalı madde-içi boyutluluk modeli ile kalibre edilerek her bir boyut için tanımlanan **a<sub>i</sub> - parametreleri** (**a<sub>1</sub>, a<sub>2</sub>, a<sub>3</sub>** -madde ayırt ediciliği ) ve maddelere ait **d-parametresi** ( madde kolaylığı) kestirilerek madde havuzu oluşturulmuştur. 2008-2013 yılları arasında uygulanan ve üç boyuttan oluşan İYS sınavına ait 10 testteki toplam 628 madde, telafi-edici çok boyutlu madde tepki kuramına (Compensatory 2PL-MIRT) dayalı madde-içi boyutluluk modeli ile kalibre edilmiştir. Boyutların en az birine ait

madde ayırt edicilik parametresi 0,5'in altında olan ve d-parametresi [-4,4] aralığının dışında olan toplam 69 madde madde havuzunda çıkartılmıştır. Sonuç olarak, telafi-edici çok boyutlu MTK'ya dayalı madde-içi boyutluluk modeline ait madde havuzu, dinleme boyutunda 110, dilbilgisi boyutunda 250 ve okuma boyutunda 199 madde olmak üzere toplam 559 maddeden oluşmuştur.

Benzer şekilde, testlere ait maddeler telafi-edici çok boyutlu MTK modellerinden maddeler-arası boyutluluk modeli ile kalibre edilerek telafi-edici çok boyutlu maddeler-arası boyutluluk modeline ait madde havuzu oluşturulmuştur. 2008-2013 yılları arasında uygulanan ve üç boyuttan oluşan İYS sınavına ait 10 testteki toplam 628 madde, telafi-edici çok boyutlu madde tepki kuramına (Compensatory 2PL-MIRT) dayalı maddeler-arası boyutluluk modeli ile kalibre edilmiştir. Boyutların en az birine ait madde ayırt edicilik parametresi 0,5'in altında olan ve d-parametresi [-4,4] aralığının dışında olan toplam 73 madde havuzunda çıkartılmıştır. Sonuç olarak, telafi-edici çok boyutlu MTK'ya dayalı maddeler-arası boyutluluk modeline ait madde havuzu, dinleme boyutunda 115, dilbilgisi boyutunda 240 ve okuma boyutunda 200 madde olmak üzere toplam 555 maddeden oluşmuştur.

### **3.4.1.2 Model-Veri Uyumu İstatistiklerinin İncelenmesi**

Çok boyutlu BOB testi ile analiz yapılmadan önce her bir çok-boyutlu MTK modeli için model veri-uyumu incelenerek her bir modele ilişkin uyum istatistiklerinden AIC, BIC, log-olabilirlik (log-likelihood) ve ki-kare ( $\chi^2$ ) istatistikleri hesaplanmıştır. Öncelikle, testin tek boyutlu olduğu duruma ait uyum istatistikleri hesaplanmıştır. Daha sonra ise, her bir teste ait veriler, testin üç boyutlu (okuma, dinleme ve dilbilgisi) olduğu durumda madde ve yetenek parametreleri çok boyutlu MTK modellerinden madde-içi boyutluluk ve maddeler arası boyutluluk modelleri ile kestirilerek model-veri uyumu istatistikleri hesaplanmıştır. Her bir çok-boyutlu modele ait model-veri uyumu istatistikleri, testin tek boyutlu olduğu durumdaki istatistiklerle karşılaştırılarak AIC, BIC, log-olabilirlik (log-likelihood) ve ki-kare ( $\chi^2$ ) değerlerindeki düşüşün anlamlı olup olmadığına bakılmıştır. Ayrıca çok boyutlu telafi-edici modellerden madde-içi ve maddeler-arası boyutluluk modellerine ait model-veri uyumu istatistikleri karşılaştırılmıştır.

Model-veri uyumuna ilişkin bulgular Ek 1, Ek 2 ve Ek 3'te verilmiştir. Testin tek boyutlu olduğu duruma ve çok boyutlu modellerden madde-içi boyutluluğa ilişkin model-veri uyumu istatistiklerinin karşılaştırılmasına ilişkin bulgular Ek1'de verilmiştir. Testin tek

boyutlu olduđu duruma ve maddeler-arası boyutluluk modeline ait model-veri uyumu istatistiklerinin karşılaştırılmasına ilişkin bulgular Ek2’de verilmiştir. Ayrıca, çok boyutlu telafi-edici modellerden madde-içi ve maddeler arası boyutluluk modellerine ait model-veri uyumu istatistiklerinin karşılaştırılmasına ilişkin bulgular ise Ek3’te verilmiştir.

Model-veri uyumu istatistiklerine bakıldığında, her bir testin tek boyutlu kabul edildiği durum ile karşılaştırıldığında, çok boyutlu modellere ait model-veri uyumu istatistiklerinin daha düşük olduğu görülmektedir. Ayrıca testlerin tek boyutlu ve çok boyutlu olduđu duruma ait uyum istatistikleri arasındaki farkın ise istatistiksel olarak anlamlı olduđu görülmektedir ( $p<0.05$ ).

Çok boyutlu modellerden madde-içi boyutluluk ve maddeler-arası boyutluluk modellerine ait model veri uyumu istatistiklerine bakıldığında ise madde-içi boyutluluk modeline ait model veri uyumu istatistiklerinin daha düşük olduğu Ek3’te görülmektedir. Ayrıca analizde kullanılan her bir test için madde-içi boyutluluk ve maddeler-arası boyutluluk modellerine ilişkin model-veri uyumu istatistikleri arasındaki farkın ise istatistiksel olarak anlamlı olduđu görülmektedir ( $p<0.05$ ). Bu bulgular ışığında her bir testin tek boyutlu model ile karşılaştırıldığında çok boyutlu modellere daha iyi uyum gösterdiği görülmektedir. Çok boyutlu modellere ilişkin model-veri uyumu bulgularına bakıldığında ise madde-içi boyutluluk modelinin diğer modellerle karşılaştırıldığında istatistiksel olarak model-veri uyumunun daha iyi olduđu görülmektedir.

### **3.4.2. Veri Analizinin İkinci Aşaması**

İkinci aşamada madde havuzu oluşturulduktan sonra post-hoc simülasyon yöntemi uygulanarak simülasyon veri seti yerine İYS’ye ait gerçek veri seti kullanılarak bireylere ait yetenek kestirilmiştir. Bireylere ait yetenek kestirimi yapılırken çok boyutlu BOB testi analizi için R-yazılımında tanımlı “MAT” paket programı (Choi ve King, 2011) kullanılmıştır. Telafi-edici çok boyutlu BOB testi yöntemleri ile analiz yapılırken, farklı yetenek kestirme yöntemleri, madde seçim yöntemleri ve durdurma kuralları kullanılarak her bir boyuta ilişkin yetenek kestirimi yapılmıştır. Telafi-edici çok boyutlu BOB testi yöntemlerinin performansları; her bir boyuta ilişkin güvenilirlik indeksi, gerçek yetenek puanı ile çok boyutlu BOB testi sonucu kestirilen yetenek puanlarının tutarlığını veren RMSD değeri, kestirilen yetenek puanlarına ilişkin ölçmenin standart hatası (SEM) ve yetenek kestirimi için gerekli ortalama madde sayısı açısından karşılaştırılmıştır.

Bu çalışmada, çok boyutlu BOB testi yöntemlerinde bireylerin test sürecinde cevaplamış oldukları maddeler değişkenlik gösterdiğinden her bir boyuta ilişkin güvenilirlik katsayısı, birey parametreleri üzerinden güvenilirliği hesaplayan birey ayırma güvenilirlik indeksi (Person Separation Reliability Index-PSI) katsayısı ile hesaplanmıştır. Birey ayırma indeksi her bir boyuta ilişkin kestirilen yetenek parametrelerinin güvenilirliğini verir. Birey ayırma güvenilirlik indeksi:

$$PSI = \frac{\sigma^2(\theta) - MSE(\theta)}{\sigma^2(\theta)}$$

formülü ile hesaplanır. Burada  $\sigma^2(\theta)$  kestirilen yetenek parametresine ait varyansı ve  $MSE(\theta)$  Yetenek parametresine ait hata karelerinin ortalamasına eşittir. Prieto ve arkadaşları (1998) birey ayırma indeksi ile güvenilirlik indeksi arasında yüksek bir ilişki olduğunu belirtmişlerdir.

Bireylere ait gerçek yetenek puanı ile çok boyutlu BOB testi sonucu kestirilen yetenek puanlarının tutarlığı ise her bir boyuta ilişkin yetenek parametrelerine ait RMSD (Root Mean Squared Difference) değerler hesaplanarak belirlenmiştir. Her bir boyuta ilişkin RMSD değerleri:

$$RMSD = \sqrt{\frac{\sum_{j=1}^n (\hat{\theta} - \theta)^2}{n}}$$

Formülü ile hesaplanır. Burada  $\hat{\theta}$  kestirilen yetenek parametresini  $\theta$  gerçek yetenek parametresini ve  $n$  ise testi alan birey sayısını temsil etmektedir.

Bu çalışmada, çok boyutlu BOB testi yöntemlerinin performanslarını karşılaştırmak için kullanılan bir diğer ölçüt ise, her bir boyuta ilişkin kestirilen yetenek parametrelerine ait ölçmenin standart hatası (Standard Error of Measurement-SEM) değeridir. Her boyuta ilişkin ölçmenin standart hatası:

$$SEM = S_{\hat{\theta}} \sqrt{1 - r^2}$$

Formülü ile hesaplanmaktadır. Burada  $S_{\hat{\theta}}$  her bir boyuta ilişkin kestirilen yetenek parametresinin standart sapmasını,  $r$  ise her bir boyuta ilişkin güvenilirlik katsayısını göstermektedir.

Post-Hoc simülasyon yöntemlerinin kullanılmasının temel amacı, gerçek çok boyutlu BOB testi (Live- MCAT) analizinde kullanılacak olan en uygun yetenek kestirme

yöntemine, madde seçim yöntemine ve durdurma kuralına karar vermektir. Böylece yabancı dil sınavları için geliştirilecek olan çok boyutlu BOB testi analizinde kullanılacak en uygun yetenek kestirme yöntemine, madde seçim yöntemine, durdurma kuralına ve testin boyutuna karar verilmiş olacaktır.

Telafi-edici çok boyutlu BOB testi analizlerinde post-hoc simülasyon yöntemleri uygulanırken;

- Yetenek kestirim yöntemleri olarak çok boyutlu yöntemler için Fisher'in puanlama yöntemi ( Fisher' scoring method) ve Bayesian yöntemlerden Maximum A Posteriori (MAP) kullanılmıştır.
- Test durdurma kurallarından *hata varyansı durdurma kuralı* ile *sabit soru sayısı* kullanılmıştır. Hata varyansı durdurma kuralı için her bir boyuta ilişkin hata varyansının sırasıyla 0,20, 0,25 ve 0,30 olduğunda test sonlandırılmıştır. Sabit soru sayısı durdurma kuralı için ise sırasıyla testteki madde sayısı 30 40 ve 50 olduğunda test sonlandırılmıştır.
- Çok boyutlu BOB testi yöntemi için Madde seçim yöntemleri olarak optimal madde seçim yöntemlerinden;
  - D-optimality,
  - A-optimality,
  - Seçkisiz (Random) madde seçim

(Segall, 1996; van der Linden, 1999; Mulder ve van der Linden, 2009) yöntemleri kullanılmıştır.

Her iki analiz yöntemi için maddenin kullanım sıklığını (item exposure) kontrol etmeye olanak sağlayan randomesque yöntemi kullanılmıştır. Bu yöntem ile analizlerde madde havuzundan madde seçilirken test bilgisini maksimum yapan ilk on madde arasından birinin seçkisiz olarak atanması kuralı uygulanmıştır. Böylece testteki maddelerin kullanım sıklığı kontrol altında tutularak sadece yüksek bilgi veren maddelerin test sürecinde seçilmesi engellenerek maddelerin testi alan bireyler tarafından ezberlenmesinin önüne geçilmiş olur.

#### 4. BULGULAR VE TARTIŞMA

Bu bölümde farklı madde seçim yöntemi, yetenek kestirim yöntemi ve durdurma kuralının kullanıldığı çok boyutlu MTK modellerinden madde-içi ve maddeler-arası boyutluluk modellerine dayalı çok-boyutlu BOB testi analiz bulguları verilmiştir. Bulgular verilirken alt problemler tekrar yazılmış ve bu problemlere ilişkin olarak çözümler aranmaya çalışılmıştır.

Gerçek veri setine dayalı çok boyutlu BOB testi simülasyon çalışması (post-hoc simulation study) sonucu bireylerin yabancı dil yeteneğini ölçmeyi amaçlayan en uygun çok-boyutlu BOB testi yöntemini belirlemek amacıyla farklı madde seçim yöntemi, yetenek kestirim yöntemi ve durdurma kuralının kullanıldığı çok boyutlu BOB testine ait her bir boyuta ilişkin kestirilen yetenek parametrelerine ait RMSD değerleri, ölçmenin standart hatası (ÖSH), güvenilirlik katsayısı ile gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerleri hesaplanmıştır. Analiz bulguları her bir çok-boyutlu model için madde seçim yöntemlerine göre sınıflandırılmıştır.

**Alt Problem 1: A-optimality madde seçim yönteminin kullanıldığı madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB test analizlerine ilişkin farklı yetenek kestirim yöntemleri ve durdurma kuralları altında her bir koşula ait güvenilirlik katsayısı, standart hata, yetenek parametrelerine ait korelasyon, testin uzunluğu ve RMSD değerleri nasıldır?**

Bu bölümde A-optimality madde seçim yönteminin kullanıldığı madde-içi ve maddeler-arası boyutluluk modeline dayalı farklı yetenek kestirim yöntemi ve durdurma kurallarına dayalı çok-boyutlu BOB testi analizlerine ilişkin bulgulara yer verilmiştir.

**a) Maddeler-arası boyutluluk modeline ait sabit madde sayısı durdurma kuralı ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.1'de her bir çok-boyutlu model için madde seçim yöntemlerinden A-optimality, durdurma kurallarından sabit soru sayısı ve yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait analiz bulgularına yer verilmiştir. Her bir model için sabit madde sayısı durdurma kuralı kullanılarak analizler bireyin cevapladığı madde sayısı sırasıyla 30, 40 ve 50'ye ulaştığında testler sonlandırılmış ve her bir model için analiz bulguları karşılaştırılmıştır.

**Tablo 4.1: A-optimality Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin BOB Testi Bulguları**

Çok-Boyutlu Model	Yetenek Kestirim yöntemi	Durdurma kuralı (sabit soru sayısı)	Güvenirlilik Katsayısı			Ölçmenin Standart Hatası			Korelasyon			RMSD Değerleri		
			K	G1	G2	G3	SH1	SH2	SH3	Kor1	Kor2	Kor3	RMSD.1	RMSD.2
Maddeler-Arası Boyutluluk Modeli	Fisher	30	0,84	0,94	0,91	0,383	0,251	0,290	0,882	0,918	0,896	0,462	0,393	0,441
		<b>40</b>	<b>0,88</b>	<b>0,95</b>	<b>0,94</b>	<b>0,363</b>	<b>0,238</b>	<b>0,272</b>	<b>0,802</b>	<b>0,831</b>	<b>0,823</b>	<b>0,645</b>	<b>0,610</b>	<b>0,620</b>
		50	0,86	0,94	0,92	0,340	0,220	0,251	0,960	0,985	0,975	0,268	0,166	0,209
	Bayesyen	30	0,86	0,94	0,92	0,384	0,250	0,290	0,808	0,829	0,827	0,611	0,595	0,590
		<b>40</b>	<b>0,84</b>	<b>0,94</b>	<b>0,92</b>	<b>0,359</b>	<b>0,231</b>	<b>0,266</b>	<b>0,949</b>	<b>0,982</b>	<b>0,971</b>	<b>0,300</b>	<b>0,183</b>	<b>0,226</b>
		50	0,86	0,94	0,92	0,340	0,220	0,251	0,960	0,986	0,976	0,267	0,157	0,208
Madde-İçi Boyutluluk Modeli	Fisher	30	0,96	0,88	0,87	0,249	0,393	0,409	0,790	0,767	0,761	0,751	0,745	0,747
		<b>40</b>	<b>0,96</b>	<b>0,87</b>	<b>0,86</b>	<b>0,224</b>	<b>0,358</b>	<b>0,376</b>	<b>0,915</b>	<b>0,903</b>	<b>0,892</b>	<b>0,433</b>	<b>0,443</b>	<b>0,458</b>
		50	0,96	0,89	0,87	0,212	0,337	0,356	0,929	0,918	0,911	0,396	0,408	0,417
	Bayesyen	<b>30</b>	<b>0,96</b>	<b>0,88</b>	<b>0,87</b>	<b>0,232</b>	<b>0,374</b>	<b>0,385</b>	<b>0,847</b>	<b>0,810</b>	<b>0,828</b>	<b>0,604</b>	<b>0,641</b>	<b>0,602</b>
		40	0,96	0,86	0,85	0,210	0,349	0,362	0,985	0,956	0,956	0,180	0,295	0,288
		50	0,96	0,88	0,87	0,199	0,333	0,348	0,989	0,963	0,964	0,155	0,270	0,261

Tablo 4.1’de maddeler-arası boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden MLE yöntemine dayalı Fisher’in puanlama yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,84 ile 0,94 aralığında, ölçmenin standart hatasının 0,25 ile 0,38 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,88 ile 0,92 aralığında ve RMSD değerlerinin ise 0,39 ile 0,46 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30’dan 40’a çıkarıldığı durumda ise boyutlara ilişkin güvenilirlik katsayılarının değişmediği, standart hataların ve RMSD azaldığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı görülmektedir. Diğer test uzunlukları ile karşılaştırıldığında testteki ortalama madde sayısı 50 olduğunda durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en düşük ve korelasyon değerlerinin ise en yüksek olduğu görülmektedir. Ayrıca, testteki ortalama madde sayısının değişkenlik göstermesine rağmen, birinci boyuta ait güvenilirlik katsayılarının eşit olduğu ve diğer boyutlara ilişkin güvenilirlik katsayılarında önemli bir artış olmadığı görülmektedir.

Maddeler-arası boyutluluk modeline ait BOB testi analiz bulgularına bakıldığında, genel olarak her bir durdurma kuralı için birinci boyuta (dinleme boyutuna) ilişkin güvenilirlik katsayılarının en düşük ve ikinci boyuta (dilbilgisi boyutuna) ilişkin güvenilirlik katsayılarının ise en yüksek değerlere sahip olduğu görülmektedir. Madde seçim yöntemlerinden A-optimality, yetenek kestirim yöntemlerinden Fisher’in puanlama yöntemi ve durdurma kuralı olarak sabit madde sayısı kullanıldığında, testteki ortalama madde sayısının 40 olduğu durumun daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Tablo 4.1’de maddeler-arası boyutluluk modeline ilişkin yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,86 ile 0,94 aralığında, ölçmenin standart hatasının 0,25 ile 0,38 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,81 ile 0,83 aralığında ve RMSD değerlerinin ise 0,59 ile 0,61 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30’dan 40’a çıkarıldığı durumda ise her bir boyuta ilişkin güvenilirlik katsayılarının değişmediği, standart hataların ve RMSD değerlerinin azaldığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı görülmektedir.



Diğer test uzunlukları ile karşılaştırıldığında testteki ortalama madde sayısı 50 olduğunda durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en düşük ve korelasyon değerlerinin ise en yüksek olduğu görülmektedir. Fakat testteki ortalama madde sayısının artması, boyutlara ilişkin güvenilirlik katsayılarında anlamlı bir farklılığa neden olmamaktadır.

Maddeler-arası boyutluluk modeline ait BOB testi analiz bulgularına bakıldığında, genel olarak her bir durdurma kuralı için birinci boyuta (dinleme boyutuna) ilişkin güvenilirlik katsayılarının en düşük ve ikinci boyuta (dilbilgisi boyutuna) ilişkin güvenilirlik katsayısının ise en yüksek değerlere sahip olduğu görülmektedir. Madde seçim yöntemlerinden A-optimality, yetenek kestirim yöntemlerinden MLE yöntemine dayalı Fisher'in puanlama yöntemi ve durdurma kuralı olarak sabit madde sayısı kullanıldığında, testteki ortalama madde sayısının 40 olduğu durumda özellikle RMSD değerlerinde önemli bir düşüş olduğu ve daha güvenilir sonuçlar verdiği söylenebilir.

Maddeler-arası boyutluluk modeli için Bayesyen MAP yetenek kestirim yöntemi kullanıldığında testteki madde sayısının artması, boyutlara ilişkin güvenilirlik katsayısında önemli bir artışa neden olmamaktadır. Maddeler-arası boyutluluk modeli için Bayesyen MAP ve Fisher'in puanlama yöntemine ilişkin bulgular karşılaştırıldığında, her iki yetenek kestirim yöntemi için en uygun madde sayısının 40 olmasına karşın, her bir koşul için Bayesyen MAP yetenek kestirim yönteminin daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

### **b) Madde-içi boyutluluk modeline ait sabit madde sayısı durdurma kuralı ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.1'de madde-içi boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,87 ile 0,96 aralığında, ölçmenin standart hatasının 0,25 ile 0,41 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,76 ile 0,79 aralığında ve RMSD değerlerinin ise 0,74 ile 0,75 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30'dan 40'a çıkarıldığı durumda ise birinci boyuta ilişkin güvenilirlik katsayısının değişmediği (0,96) ve diğer boyutlara ilişkin güvenilirlik katsayılarında ise 0,01 düzeyinde bir düşüş olduğu görülmektedir. Ayrıca, boyutlara ilişkin standart hataların ve RMSD azaldığı ve gerçek

ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı görülmektedir. Testteki ortalama madde sayısı 50 olduğunda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en düşük, korelasyon değerlerinin ise en yüksek olduğu görülmektedir.

Madde-içi boyutluluk modeline ait BOB testi analiz bulgularına bakıldığında, genel olarak her bir durdurma kuralı için birinci boyuta (dinleme boyutuna) ilişkin güvenilirlik katsayılarının en yüksek ve diğer boyutlara ilişkin güvenilirlik katsayılarının ise birbirine yakın değerlere sahip olduğu görülmektedir.

Genel olarak madde-içi boyutluluk modeline ilişkin analiz bulguları incelendiğinde, madde seçim yöntemlerinden A-optimality, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak sabit madde sayısı kullanıldığında, testteki ortalama madde sayısının 40 olduğu durumun daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir. Ayrıca madde sayısı arttıkça boyutlara ilişkin standart hata ve RMSD değerleri azalırken, testin boyutlarına ilişkin güvenilirlik katsayılarında anlamlı bir değişkenlik olmadığı görülmektedir.

Madde-içi boyutluluk modelinde yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,87 ile 0,96 aralığında, ölçmenin standart hatasının 0,23 ile 0,39 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,81 ile 0,87 aralığında ve RMSD değerlerinin ise 0,60 ile 0,64 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30'dan 40'a çıkarıldığı durumda ise birinci boyuta ilişkin güvenilirlik katsayısının aynı kaldığı ve diğer boyutlara ilişkin güvenilirlik katsayısının ise 0,02 düzeyinde bir azalma olduğu görülmektedir. Buna karşı, standart hataların ve RMSD değerlerinin azaldığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı görülmektedir. Testteki ortalama madde sayısı 50 olduğu durumda ise her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en düşük, korelasyon değerlerinin ise en yüksek olduğu görülmektedir.

Madde-içi boyutluluk modeline ait BOB testi analiz bulgularına bakıldığında, genel olarak her bir durdurma kuralı için birinci boyuta (dinleme boyutuna) ilişkin güvenilirlik katsayılarının en yüksek ve diğer iki boyuta (dilbilgisi ve okuduğunu anlama boyutları) ilişkin güvenilirlik katsayılarının ise yüksek ve birbirine yakın değerler aldığı görülmektedir. Genel olarak madde seçim yöntemlerinden A-optimality, yetenek

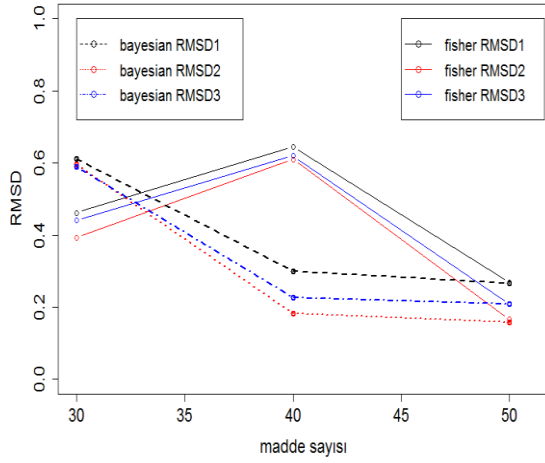
kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak sabit madde sayısı kullanıldığında, testteki ortalama madde sayısının 30 olduğu durumun daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Madde-içi boyutluluk modeli için Bayesyen MAP yetenek kestirim yöntemi kullanıldığında testteki madde sayısının artması, boyutlara ilişkin güvenilirlik katsayısında önemli bir artışa neden olmamaktadır. Maddeler-arası boyutluluk modeli için Bayesyen MAP ve Fisher'in puanlama yöntemine ilişkin bulgular karşılaştırıldığında, her bir koşul için MAP yetenek kestirim yönteminin daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

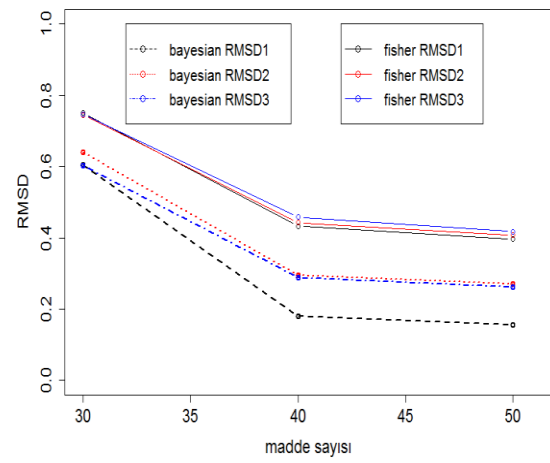
Maddeler-arası ve madde –içi boyutluluk modeline ilişkin farklı çok boyutlu BOB testi koşullarına ilişkin bulgular karşılaştırıldığında, her iki model için aynı koşullarda Bayesyen yetenek kestirim yöntemine ait güvenilirlik katsayıları ve yetenek parametreleri arasındaki korelasyonun Fisher'in puanlama yöntemine göre daha yüksek ve standart hata ve RMSD değerlerinin ise daha düşük olduğu görülmektedir. Ayrıca, madde-içi boyutluluk ve maddeler-arası boyutluluk modeline ilişkin Fisher'in puanlama yetenek kestirim ve Bayesyen MAP yetenek kestirim yöntemine ilişkin çok boyutlu BOB testi bulguları karşılaştırıldığında, madde-içi boyutluluk modelinin her iki yetenek kestirim yöntemi için daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği yorumu yapılabilir.

Şekil 4.1'de maddeler-arası ve madde-içi çok boyutlu telafi-edici modellere ilişkin madde seçim yöntemlerinden A-optimality, testi durdurma kurallarından sabit madde sayısının ve yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testi analizlerine ait grafikler verilmiştir. Şekilde kesikli çizgiler Bayesyen yetenek kestirim yöntemine ait bulguları temsil ederken, düz çizgiler ise Fisher'in puanlama yetenek kestirim yöntemine ilişkin bulguları temsil etmektedir. Ayrıca grafikte yer alan siyah renkli çizgiler birinci boyuta (dinleme boyutuna) ait bulguları, kırmızı renkli çizgi ikinci boyuta (dilbilgisi boyutuna) ait bulguları ve mavi renkli çizgi ise üçüncü boyuta (okuduğunu anlama boyutuna) ait bulguları temsil etmektedir.

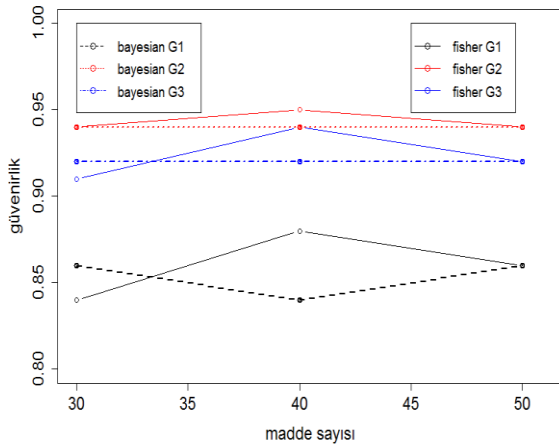
a.1 RMSD- Madde sayısı ilişkisi



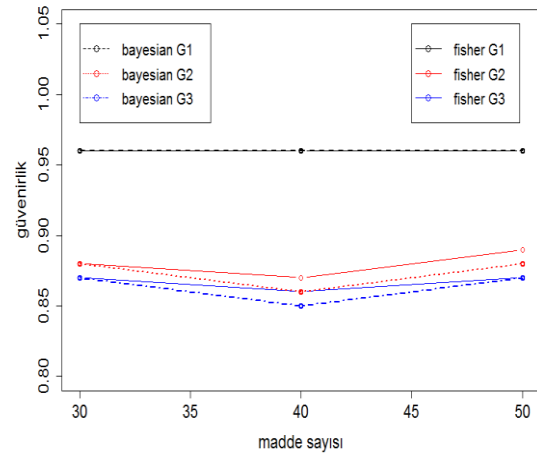
a.2 RMSD- Madde sayısı ilişkisi



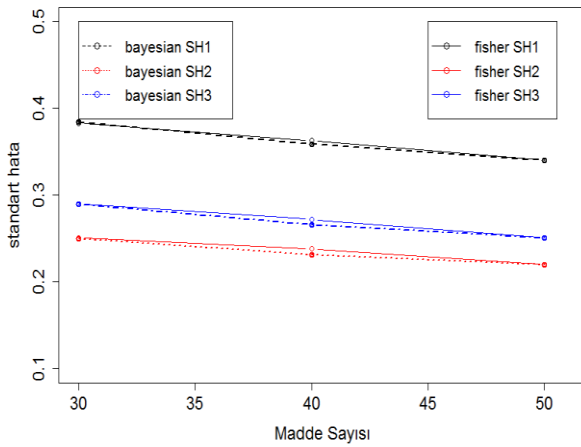
b.1 Güvenirlilik- Madde sayısı ilişkisi



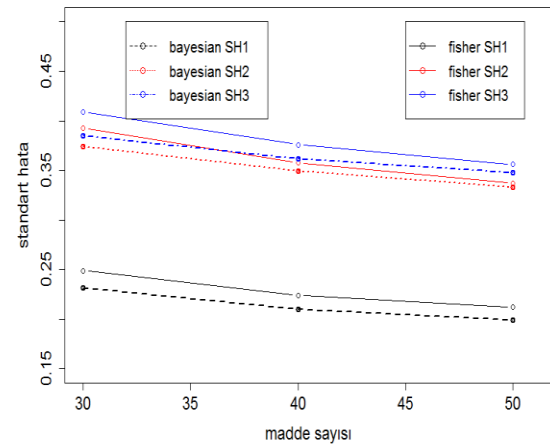
b.2 Güvenirlilik- Madde sayısı ilişkisi



c.1 Ölçmenin standart hatası- Madde sayısı



c.2 Ölçmenin standart hatası-Madde sayısı



**Şekil 4.1: A-optimality Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin Grafikler**

Şekil 4.1'de yer alan a.1 ve a.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin RMSD değerleri ile testteki madde sayısı arasındaki ilişkiyi vermektedir. Şekil 4.1 a.1'e bakıldığında maddeler-arası boyutluluk modelinde, testteki madde sayısı arttıkça Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin azaldığı görülmektedir. Ancak, Fisher'in puanlama yönteminin kullanıldığı durumda testteki madde sayısı 30'dan 40 çıktığında RMSD değerlerinin arttığı ve madde sayısı 50 olduğunda ise Bayesyen yetenek kestirim yöntemi ile benzer sonuçlar verdiği görülmektedir.

Madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi analizlerine ilişkin RMSD değerleri ile testteki madde sayısı arasındaki ilişkiyi veren Şekil 4.1 a.2'e bakıldığında, testteki madde sayısı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin azaldığı görülmektedir. Ayrıca her iki çok-boyutlu model için, Bayesyen yetenek kestirim yöntemine ait RMSD değerlerinin daha düşük olduğu ve testteki madde sayısı 50'ye ulaştığında bu farkın azaldığı görülmektedir.

Şekil 4.1'de yer alan b.1 ve b.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin güvenilirlik katsayıları ile testteki madde sayısı arasındaki ilişkiyi vermektedir. Şekil 4.1 a.1'e bakıldığında maddeler-arası boyutluluk modelinde, Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin güvenilirlik katsayılarının testteki sayısı 30'dan 40'a çıktığında ikinci ve üçüncü boyuta ilişkin güvenilirlik katsayılarının değişmediği ve birinci boyuta ilişkin güvenilirlik katsayısının 0,01 düzeyinde bir azalma olduğu görülmektedir. Ancak, Fisher'in puanlama yönteminin kullanıldığı durumda testteki madde sayısı 30'dan 40 çıktığında her bir boyuta ilişkin güvenilirlik katsayılarının arttığı ve madde sayısı 50 olduğunda ise Bayesyen yetenek kestirim yöntemi ile benzer sonuçlar verdiği görülmektedir.

Madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi analizlerine ilişkin güvenilirlik katsayıları ile testteki madde sayısı arasındaki ilişkiyi veren Şekil 4.1 b.2'ye bakıldığında, testteki madde sayısı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait birinci boyuta ilişkin güvenilirlik katsayılarının değişmediği ve diğer boyutlara ilişkin güvenilirlik katsayılarının ise testteki madde sayısı 40 olduğunda biraz düştüğü görülmektedir. Ayrıca her iki çok-boyutlu model için,

Bayesyen yetenek kestirim yöntemi ve Fisher'in puanlama yönteminin kullanıldığı çok-boyutlu BOB testi uygulamalarının benzer güvenilirlik katsayılarına sahip olduğu görülmektedir.

Şekil 4.1'de yer alan c.1 ve c.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testlerine ait her bir boyuta ilişkin ölçmenin standart hatası ile testteki madde sayısı arasındaki ilişkiyi vermektedir. Şekil 4.1 c.1'e bakıldığında maddeler-arası boyutluluk modelinde, Bayesyen MAP yetenek kestirimi ve Fisher'in puanlama yöntemine ait her bir boyuta ilişkin standart hatanın testteki madde sayısı arttıkça azalma eğiliminde olduğu görülmektedir. Ayrıca maddeler arası boyutluluk modelinde hem Bayesyen hem de Fisher'in puanlama yönteminin kullanıldığı BOB testine ait standart hatanın her bir koşul için bir birine yakın değerler aldığı görülmektedir.

Madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi analizlerine ilişkin ölçmenin standart hatası ile testteki madde sayısı arasındaki ilişkiyi veren Şekil 4.1 c.2'ye bakıldığında, testteki madde sayısı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin ölçmenin standart hatasının azaldığı ve birinci boyuta ait standart hata değerinin diğer boyutlarla kıyaslandığında daha düşük olduğu görülmektedir. Ayrıca madde-içi boyutluluk modelinde Bayesyen yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testi uygulamalarının daha düşük standart hata değerlerine sahip olduğu görülmektedir.

### **c) Maddeler-arası boyutluluk modeline ait farklı hata varyansı durdurma kuralları ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.2'de her bir çok-boyutlu model için madde seçim yöntemlerinden A-optimality, durdurma kurallarından hata varyansı ve yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait analiz bulgularına yer verilmiştir. Her bir model için standart hata durdurma kuralı kullanılarak analizler bireylerin kestirilen yetenek parametrelerine ait hata varyansının sırasıyla 0,20, 0,25 ve 0,30'un altına düştüğünde testler sonlandırılmıştır. Bireyin yabancı dil yeteneğini ölçmeyi amaçlayan en uygun çok-boyutlu BOB testi yöntemine ilişkin en uygun yetenek kestirim yöntemi, madde seçim yöntemi ve durdurma kuralına karar vermek için her bir çok-boyutlu BOB testi koşuluna ilişkin analiz bulguları karşılaştırılmıştır.

**Tablo 4.2: A-Optimality Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralı Yöntemine İlişkin BOB Testi Bulgular**

Çok Boyutlu Model	Yetenek kestirim yöntemi	Durdurma kuralı (s. hata)	Ortalama madde sayısı	Güvenirlilik Katsayısı			Ölçmenin Standart Hata			Korelasyon			RMSD Değerleri		
				S. Hata	K	G1	G2	G3	SH1	SH2	SH3	Kor1	Kor2	Kor3	RMSD1
Maddeler-Arası Boyutluluk Modeli	Fisher	0,20	39,6	0,84	0,94	0,92	0,359	0,229	0,265	0,951	0,983	0,974	0,294	0,173	0,216
		<b>0,25</b>	<b>31,4</b>	<b>0,95</b>	<b>0,84</b>	<b>0,82</b>	<b>0,231</b>	<b>0,379</b>	<b>0,395</b>	<b>0,981</b>	<b>0,946</b>	<b>0,945</b>	<b>0,197</b>	<b>0,326</b>	<b>0,320</b>
		0,30	15,1	0,74	0,89	0,84	0,437	0,293	0,354	0,900	0,963	0,930	0,415	0,259	0,346
	Bayesyen	0,20	39,7	0,84	0,94	0,92	0,359	0,228	0,265	0,948	0,983	0,971	0,305	0,174	0,226
		<b>0,25</b>	<b>22,8</b>	<b>0,83</b>	<b>0,94</b>	<b>0,91</b>	<b>0,404</b>	<b>0,266</b>	<b>0,312</b>	<b>0,806</b>	<b>0,855</b>	<b>0,850</b>	<b>0,604</b>	<b>0,545</b>	<b>0,540</b>
		0,30	15,2	0,79	0,91	0,87	0,437	0,297	0,353	0,786	0,849	0,816	0,626	0,540	0,591
Madde-İçli Boyutluluk Modeli	Fisher	0,20	49,58	0,96	0,88	0,86	0,208	0,335	0,354	0,988	0,962	0,963	0,162	0,276	0,264
		<b>0,25</b>	<b>31,40</b>	<b>0,95</b>	<b>0,84</b>	<b>0,82</b>	<b>0,231</b>	<b>0,379</b>	<b>0,395</b>	<b>0,981</b>	<b>0,946</b>	<b>0,945</b>	<b>0,197</b>	<b>0,326</b>	<b>0,320</b>
		0,30	20,53	0,95	0,87	0,86	0,269	0,425	0,438	0,763	0,726	0,730	0,809	0,820	0,807
	Bayesyen	0,20	43,71	0,96	0,89	0,88	0,208	0,343	0,357	0,981	0,944	0,946	0,201	0,332	0,317
		<b>0,25</b>	<b>26,05</b>	<b>0,94</b>	<b>0,83</b>	<b>0,82</b>	<b>0,232</b>	<b>0,381</b>	<b>0,393</b>	<b>0,972</b>	<b>0,922</b>	<b>0,929</b>	<b>0,244</b>	<b>0,387</b>	<b>0,362</b>
		0,30	15,74	0,92	0,78	0,78	0,268	0,426	0,428	0,913	0,890	0,886	0,437	0,474	0,478

Tablo 4.2’de maddeler-arası boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher’in puanlama yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,84 ile 0,92 aralığında, ölçmenin standart hatasının 0,23 ile 0,36 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,95 ile 0,98 aralığında ve RMSD değerlerinin ise 0,17 ile 0,29 aralığında değiştiği görülmektedir. Standart hata durdurma kuralının 0,20 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 39,6’dır. Standart hata durdurma kuralı 0,20’den 0,25’e çıkarıldığında ise birinci boyuta ilişkin güvenilirlik katsayısı (0,95) artarken diğer boyutlara ilişkin güvenilirlik katsayısının azaldığı görülmektedir. Benzer şekilde, birinci boyuta ilişkin standart ve RMSD değerinin azaldığı, diğer boyutlara ilişkin standart hata ve RMSD değerlerinin arttığı görülmektedir. Ayrıca birinci boyuta ilişkin gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı ve diğer boyutlara ilişkin korelasyon değerlerinin azaldığı görülmektedir. Standart hata durdurma kuralının 0,25 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 31,4’tür. Diğer standart hata durdurma kuralları ile karşılaştırıldığında, standart hata durdurma kuralının 0,30 olduğunda durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en yüksek ve korelasyon değerlerinin ise en düşük değere sahip olduğu görülmektedir.

Maddeler-arası boyutluluk modeline ait çok-boyutlu BOB testi analiz bulgularına bakıldığında, genel olarak, madde seçim yöntemlerinden A-optimality, yetenek kestirim yöntemlerinden Fisher’in puanlama yöntemi ve standart hata durdurma kuralının 0,25 olduğu koşula ait testteki ortalama madde sayısının 31,4 olup daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Tablo 4.2’de maddeler-arası boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,84 ile 0,94 aralığında, ölçmenin standart hatasının 0,23 ile 0,36 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,95 ile 0,98 aralığında ve RMSD değerlerinin ise 0,17 ile 0,31 aralığında değiştiği görülmektedir. Standart hata durdurma kuralının 0,20 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 39,7’dir.



Standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında ise birinci ve üçüncü boyuta ilişkin güvenilirlik katsayıları 0,01' düzeyinde azalırken ikinci boyuta ilişkin güvenilirlik katsayısının değişmediği görülmektedir. Benzer şekilde, birinci boyuta ilişkin standart hata ve RMSD değerinin azaldığı, diğer boyutlara ilişkin standart hata ve RMSD değerlerinin arttığı görülmektedir. Standart hata durdurma kuralının 0,25 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 22,8'dir. Diğer standart hata durdurma kuralları ile karşılaştırıldığında, standart hata durdurma kuralının 0,30 olduğunda durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en yüksek ve korelasyon değerlerinin ise en düşük değere sahip olduğu görülmektedir.

Maddeler-arası boyutluluk modeline ilişkin analiz sonuçlarına bakıldığında, madde seçim yöntemlerinden A-optimality, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve standart hata durdurma kuralı kullanıldığında, standart hatanın 0,25'e eşitlendiği ve ortalama madde sayısının 22,8 olduğu durumun daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Maddeler-arası boyutluluk modeli için Bayesyen MAP yetenek kestirim yöntemi kullanıldığında, kestirilen yetenek parametrelerine ait standart hatanın azalması, boyutlara ilişkin güvenilirlik katsayılarında artışa neden olmaktadır. Maddeler-arası boyutluluk modeli için Bayesyen MAP ve Fisher'in puanlama yöntemine ilişkin bulgular karşılaştırıldığında, her bir koşul için MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

#### **d) Madde-içi boyutluluk modeline ait farklı standart hata durdurma kuralları ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.2'de yer alan madde-içi boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,86 ile 0,96 aralığında, ölçmenin standart hatasının 0,21 ile 0,35 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,96 ile 0,99 aralığında ve RMSD değerlerinin ise 0,16 ile 0,28 aralığında değiştiği görülmektedir. Standart hata durdurma kuralının 0,20 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 49,6'dır. Standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında ise her bir boyuta ilişkin

güvenirlik katsayısının ve yetenek parametrelerine ilişkin korelasyon değerlerinin azaldığı; boyutlara ilişkin standart hata ve RMSD değerlerinin arttığı görülmektedir. Standart hata durdurma kuralının 0,25 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 31,4'tür.

Madde-içi boyutluluk modeline ait analiz bulgularına bakıldığında, genel olarak, madde seçim yöntemlerinden A-optimality, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve standart hata durdurma kuralının 0,25 olduğu koşula ait testteki ortalama madde sayısının 31,4 olup daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Tablo 4.2'de madde-içi boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,88 ile 0,96 aralığında, ölçmenin standart hatasının 0,21 ile 0,36 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,95 ile 0,98 aralığında ve RMSD değerlerinin ise 0,20 ile 0,33 aralığında değiştiği görülmektedir. Standart hata durdurma kuralının 0,20 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 43,7'dir. Standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında ise boyutlara ait güvenilirlik katsayılarının ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin azalırken, standart hata ve RMSD değerlerinin ise arttığı görülmektedir. Ayrıca, standart hata durdurma kuralının 0,25 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 26 olarak hesaplanmıştır. Diğer standart hata durdurma kuralları ile karşılaştırıldığında, standart hata durdurma kuralının 0,30 olduğunda durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en yüksek ve korelasyon ve güvenilirlik katsayıların ise en düşük değere sahip olduğu görülmektedir. Bunun temel sebebi ise testteki madde sayısının artmasıdır.

Madde-içi boyutluluk modeline ait çok-boyutlu BOB testi analiz bulgularına bakıldığında, genel olarak her bir standart hata durdurma kuralı için birinci boyuta (dinleme boyutuna) ilişkin güvenilirlik katsayılarının en yüksek ve diğer boyutlara ilişkin güvenilirlik katsayılarının ise yüksek ve birbirine yakın değerler aldığı görülmektedir. Ayrıca, madde seçim yöntemlerinden A-optimality, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve standart hata durdurma kuralı kullanıldığında, standart hatanı 0,25'e eşitlendiği durumda ortalama madde sayısının 26'ya eşit olup diğer

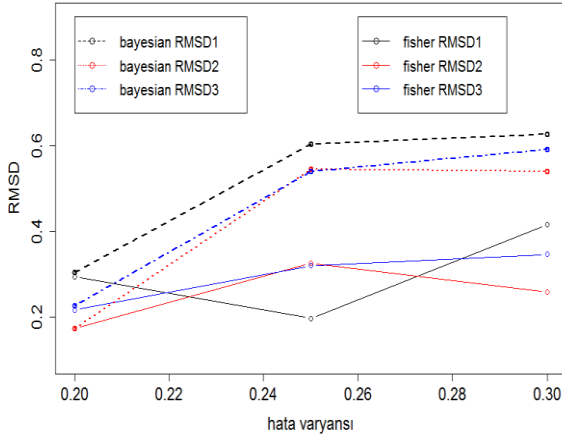
durumlarla karşılaştırıldığında daha az madde ile güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Madde-içi boyutluluk modeli için Bayesyen MAP yetenek kestirim yöntemi kullanıldığında, kestirilen yetenek parametrelerine ait standart hatanın azalması, boyutlara ilişkin güvenilirlik katsayılarında artışa neden olmaktadır. Madde-içi boyutluluk modeli için Bayesyen MAP ve MLE yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemine ilişkin bulgular karşılaştırıldığında, her bir koşul için MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir. Diao ve Reckase (2009) yapmış oldukları çalışmada benzer sonuçlar elde etmiştir. Ayrıca, Bayesyen MAP yetenek kestiriminin MLE yetenek kestirim yöntemlerine göre gerçek  $\theta$  değerine daha çabuk yakınsadığı belirtilmiştir. Dolayısıyla, testteki madde sayısı az olduğunda MLE yetenek kestirim yöntemlerinin yakınsaya mamasından dolayı kestirilen yetenek parametrelerine ait güvenilirlik katsayılarının daha düşük olduğu söylenebilir. Bu çalışmadaki bulgulara paralel olarak, Diao ve Reckase (2009) testteki madde sayısı 50 olduğunda her iki yetenek kestirim yöntemlerinin benzer sonuçlar verdiğini ifade etmektedir.

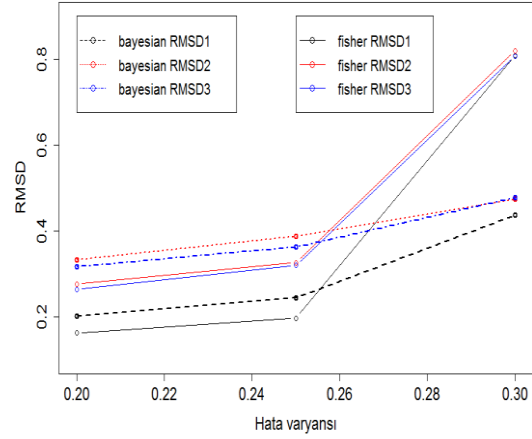
Maddeler-arası ve madde –içi boyutluluk modellerine ait farklı çok-boyutlu BOB testi koşullarına ilişkin bulgular karşılaştırıldığında, her iki model için aynı koşullarda Bayesyen yetenek kestirim yöntemine ait güvenilirlik katsayıları ve yetenek parametreleri arasındaki korelasyonun Fisher'in puanlama yöntemine göre daha yüksek; standart hata ve RMSD değerlerinin ise daha düşük olduğu görülmektedir. Ayrıca, standart hata durdurma kuralını kullanıldığı madde-içi boyutluluk ve maddeler-arası boyutluluk modeline ait farklı yetenek kestirim yöntemlerine ilişkin çok boyutlu BOB testi bulguları karşılaştırıldığında, madde-içi boyutluluk modelinin her iki yetenek kestirim yöntemi için daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği yorumu yapılabilir.

Şekil 4.2'de maddeler-arası ve madde-içi çok boyutlu telafi-edici modellere ilişkin madde seçim yöntemlerinden A-optimality, testi durdurma kurallarından boyutlara ilişkin hata varyansı ve yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testi analizlerine ait grafikler verilmiştir.

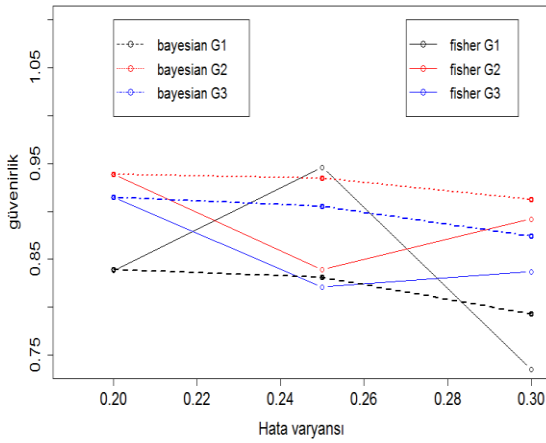
a.1 RMSD- Hata varyansı ilişkisi



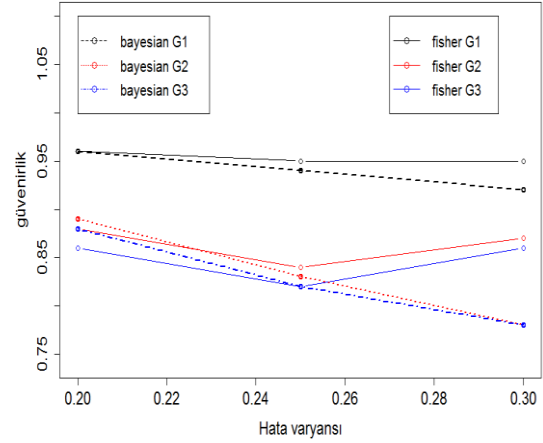
a.2 RMSD- Hata varyansı ilişkisi



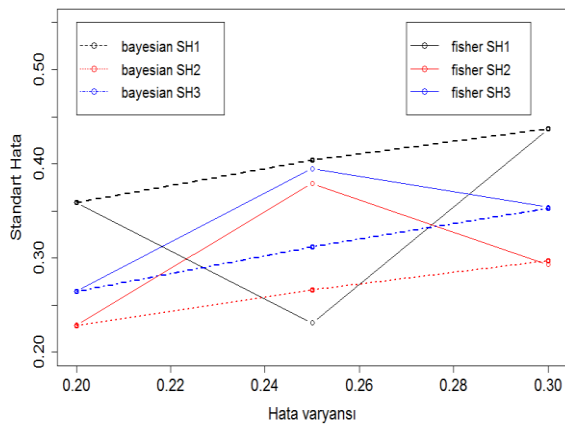
b.1 Güvenirlilik- Hata varyansı ilişkisi



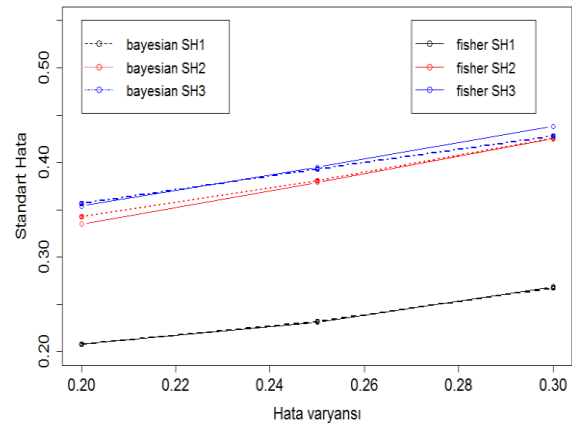
b.2 Güvenirlilik- Hata varyansı ilişkisi



c.1 Ölçmenin standart hatası- Hata varyansı



c.2 Ölçmenin standart hatası- Hata varyansı



**Şekil 4.2: A-Optimality Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralına İlişkin Grafikler**

Şekil 4.2'de yer alan a.1 ve a.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin RMSD değerleri ile yetenek parametrelerine ilişkin hata varyansları arasındaki ilişkiyi vermektedir. Şekil 4.2 a.1'e bakıldığında maddeler-arası boyutluluk modelinde, yetenek parametrelerine ilişkin hata varyansı arttıkça Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin arttığı görülmektedir. Ancak, Fisher'in puanlama yönteminin kullanıldığı durumda hata varyansı 0,20'den 0,25'e çıktığında birinci boyutla ilişkin RMSD değerlerinin azalırken diğer boyutlara ilişkin RMSD değerlerinin arttığı görülmektedir. Dolayısıyla Bayesyen yetenek kestirimi yönteminin daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi analizlerine ilişkin RMSD değerleri ile yetenek parametrelerine ait hata varyansı arasındaki ilişkiyi veren Şekil 4.2 a.2'ye bakıldığında, kestirilen yetenek parametrelerine ilişkin hata varyansı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin arttığı görülmektedir. Ayrıca maddeler-arası boyutluluk modeli için farklı hata varyansı durdurma kuralı kullanıldığında Bayesyen yetenek kestirim yöntemine ait RMSD değerlerinin daha yüksek olduğu ve yetenek parametrelerine ilişkin hata varyansı arttıkça Bayesyen ve Fisher'in puanlama yöntemine ait RMSD değerleri arasındaki farkın arttığı görülmektedir.

Şekil 4.2'de yer alan b.1 ve b.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin güvenilirlik katsayıları ile yetenek parametrelerine ait hata varyansları arasındaki ilişkiyi vermektedir. Şekil 4.2 b.1'e bakıldığında maddeler-arası boyutluluk modelinde, yetenek parametrelerine ait hata varyansları arttıkça, Bayesyen MAP yetenek kestirimine yöntemine ait her bir boyuta ilişkin güvenilirlik katsayılarının düzenli bir şekilde azaldığı görülmektedir. Ancak, Fisher'in puanlama yönteminin kullanıldığı durumda hata varyansı arttıkça boyutlara ilişkin güvenilirlik katsayılarının artıp azaldığı görülmektedir.

Şekil 4.2 b.2'ye bakıldığında, her bir boyuta ilişkin hata varyansı arttıkça Bayesyen MAP yetenek kestirimine ait güvenilirlik katsayılarının azaldığı görülmektedir. Ancak, Fisher'in puanlama yetenek kestirim yöntemi kullanıldığında, hata varyansı 0,20'den 0,25'e çıktığında güvenilirlik katsayıları azalırken, hata varyansı 0,30'a çıktığında ise

tekrar artmaktadır. Dolayısıyla madde-içi boyutluluk modeli için hata varyansı durdurma kuralı kullanıldığında Bayesyen yetenek kestirim metodunun daha güvenilir sonuçlar verdiği yorumu yapılabilir.

Şekil 4.2'de yer alan c.1 ve c.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testlerine ait her bir boyuta ilişkin ölçmenin standart hatası ile hata varyansı arasındaki ilişkiyi vermektedir. Şekil 4.2 c.1'e bakıldığında, Bayesyen MAP yetenek kestirimi yöntemi kullanıldığında, yetenek parametrelerine ilişkin hata varyansı arttıkça, her bir boyuta ilişkin ölçmenin standart hatasının da arttığı görülmektedir. Buna karşın, Bayesyen MAP yetenek kestirimi yöntemi yerine Fisher'in puanlama yöntemi kullanıldığında, boyutlara ilişkin hata varyansı 0,20'den 0,25'e çıktığında ikinci ve üçüncü boyuta ait standart hatanın artarken, birinci boyuta ait standart hata azalmaktadır. Ayrıca madde-içi boyutluluk modelinde testi durdurma kuralı olarak hata varyansı 0,30'a eşitlendiğinde hem Bayesyen hem de Fisher'in puanlama yönteminin kullanıldığı BOB testine ait standart hatanın, her bir koşul için birbirine yakın değerler aldığı görülmektedir. Bunun temel sebebi her bir boyuta ilişkin yetenek parametrelerine ait hata varyansı 0,30'a çıktığında, Fisher'in puanlama yönteminin kullanıldığı BOB testine ait ortalama madde sayısını 49 ve Bayesyen MAP yöntemine ait ortalama madde sayısı ise 43'e eşit olmasıdır. Diğer bir ifadeyle, her iki yöntem için madde sayısı arttıkça birbirine yakın sonuçlar verdiği görülmektedir.

Şekil 4.2 c.2'ye bakıldığında, boyutlara ilişkin hata varyansı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin ölçmenin standart hatasının da arttığı görülmektedir. Ayrıca madde-içi boyutluluk modelinde Fisher'in puanlama ve Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin ölçmenin standart hatasının birbirine yakın değerler aldığı görülmektedir.

Genel olarak hata varyansı durdurma kuralının kullanıldığı madde-içi ve maddeler-arası çok boyutlu BOB testi analiz yöntemlerine ilişkin bulgular karşılaştırıldığında, hem Bayesyen MAP hem de Fisher'in puanlama yöntemlerine ait en uygun hata varyansı durdurma kuralının 0,25 olduğu görülmektedir. Ayrıca Fisher'in puanlama yönteminin madde düzeyindeki boyutluluk modellerinden ve testte yer alan madde sayısından etkilendiği ve madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi uygulamalarında daha güvenilir ve kararlı sonuçlar verdiği görülmektedir. Diğer

taftan, Bayesyen yetenek kestirim yönteminin madde düzeyindeki boyutluluk modellerinden etkilenmediği ve özellikle madde-içi boyutluluk modelinin kullanıldığı durumda daha az madde ile daha güvenilir sonuçlar verdiği yorumu yapılabilir.

**Alt Problem 2: D-optimality madde seçim yönteminin kullanıldığı madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB test analizlerine ilişkin farklı yetenek kestirim yöntemleri ve durdurma kuralı altında her bir koşula ait güvenilirlik katsayısı, standart hata, yetenek parametrelerine ait korelasyon, testin uzunluğu ve RMSD değerleri nasıldır?**

Bu bölümde madde seçim yöntemlerinden D-optimality yöntemine ilişkin madde-içi ve maddeler-arası boyutluluk modeline dayalı farklı yetenek kestirim ve durdurma kurallarının kullanıldığı çok-boyutlu BOB testi analizlerine ait güvenilirlik katsayısı, standart hata, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon ve RMSD değerlerine ilişkin bulgulara yer verilmiştir.

**a) Maddeler-arası boyutluluk modeline ait sabit madde sayısı durdurma kuralı ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.3'te her bir çok-boyutlu model için madde seçim yöntemlerinden D-optimality, durdurma kurallarından sabit soru sayısı, yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait analiz bulgularına yer verilmiştir. Her bir model için sabit madde sayısı durdurma kuralı kullanılarak analizler bireyin cevapladığı madde sayısı sırasıyla 30, 40 ve 50'ye ulaştığında testler sonlandırılmış ve her bir model için analiz bulguları karşılaştırılmıştır

**Tablo 4.3: D-Optimality Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin BOB Testi Bulgular**

Çok- Boyutlu Model	Yetenek Kestirim yöntemi	Durdurma kuralı (sabit soru sayısı)	Güvenirlilik Katsayısı			Ölçmenin Standart Hatası			Korelasyon			RMSD Değerleri		
			K	G1	G2	G3	SH1	SH2	SH3	Kor1	Kor2	Kor3	RMSD1	RMSD2
Maddeler- Arası Boyutluluk Modeli	Fisher	30	0,66	0,75	0,58	0,563	0,500	0,603	0,699	0,744	0,671	0,746	0,704	0,757
		40	0,74	0,82	0,67	0,536	0,469	0,581	0,698	0,745	0,665	0,789	0,740	0,798
		50	0,78	0,84	0,70	0,513	0,448	0,565	0,679	0,741	0,650	0,835	0,769	0,835
	Bayesyen	30	0,88	0,95	0,93	0,386	0,255	0,294	0,769	0,780	0,788	0,719	0,730	0,708
		<b>40</b>	<b>0,88</b>	<b>0,95</b>	<b>0,94</b>	<b>0,361</b>	<b>0,235</b>	<b>0,269</b>	<b>0,815</b>	<b>0,854</b>	<b>0,849</b>	<b>0,612</b>	<b>0,560</b>	<b>0,564</b>
		50	0,88	0,95	0,94	0,342	0,224	0,255	0,851	0,875	0,861	0,536	0,504	0,526
Madde-İçi Boyutluluk Modeli	Fisher	30	0,85	0,78	0,76	0,347	0,409	0,423	0,942	0,928	0,924	0,348	0,376	0,376
		<b>40</b>	<b>0,88</b>	<b>0,82</b>	<b>0,79</b>	<b>0,318</b>	<b>0,378</b>	<b>0,397</b>	<b>0,963</b>	<b>0,943</b>	<b>0,935</b>	<b>0,281</b>	<b>0,336</b>	<b>0,350</b>
		50	0,90	0,85	0,83	0,292	0,357	0,375	0,965	0,953	0,946	0,271	0,306	0,318
	Bayesyen	<b>30</b>	<b>0,95</b>	<b>0,84</b>	<b>0,83</b>	<b>0,224</b>	<b>0,372</b>	<b>0,383</b>	<b>0,926</b>	<b>0,912</b>	<b>0,909</b>	<b>0,403</b>	<b>0,424</b>	<b>0,427</b>
		40	0,96	0,86	0,85	0,209	0,350	0,363	0,982	0,948	0,949	0,194	0,321	0,310
		50	0,96	0,89	0,88	0,201	0,336	0,351	0,985	0,955	0,955	0,175	0,298	0,290



Tablo 4.3'teki maddeler-arası boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, Fisher'in puanlama yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,58 ile 0,75 aralığında, ölçmenin standart hatasının 0,50 ile 0,60 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,67 ile 0,74 aralığında ve RMSD değerlerinin ise 0,70 ile 0,76 aralığında değiştiği görülmektedir. Buna karşın, RMSD değerlerinin testteki madde sayısı arttıkça düşmesi beklenirken, tam tersine arttığı görülmektedir. Ayrıca, testteki ortalama madde sayısı 50 olduğu durumda bile, genel olarak güvenilirlik katsayılarının diğer koşullar ile karşılaştırıldığında düşük olduğu görülmektedir.

Maddeler-arası boyutluluk modeline ait BOB testi analiz bulgularına bakıldığında, genel olarak her bir durdurma kuralı için boyutlara ilişkin güvenilirlik katsayılarının ve korelasyon değerlerinin düşük, boyutlara ilişkin ölçmenin standart hatasının ve RMSD değerlerinin yüksek olduğu görülmektedir. Madde seçim yöntemlerinden D-optimality, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak sabit madde sayısı kullanıldığında, testteki ortalama madde sayısının 50 olduğu durumda bile güvenilir ve tutarlı sonuçlar elde edilmediği görülmektedir.

Tablo 4.3'te maddeler-arası boyutluluk modeline ilişkin yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak madde sayısının 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,88 ile 0,95 aralığında, ölçmenin standart hatasının 0,23 ile 0,36 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,77 ile 0,79 aralığında ve RMSD değerlerinin ise 0,71 ile 0,73 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30'dan 40'a çıkarıldığı durumda ise standart hataların ve RMSD değerlerinin azaldığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı görülmektedir. Testteki ortalama madde sayısı 50 olduğu durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en düşük ve korelasyon değerlerinin ise en yüksek olduğu görülmektedir. Fakat testteki ortalama madde sayısının artması, boyutlara ilişkin güvenilirlik katsayılarında anlamlı bir farklılığa neden olmamaktadır.

Maddeler-arası boyutluluk modeline ait BOB testi analiz bulgularına bakıldığında, genel olarak her bir durdurma kuralı için birinci boyuta (dinleme boyutuna) ilişkin

güvenirlik katsayılarının en düşük, ikinci boyuta (dilbilgisi boyutuna) ilişkin güvenilirlik katsayısının ise en yüksek değerlere sahip olduğu görülmektedir. Madde seçim yöntemlerinden D-optimality, yetenek kestirim yöntemlerinden Bayesyen MAP puanlama yöntemi ve durdurma kuralı olarak sabit madde sayısı kullanıldığında, testteki ortalama madde sayısının 40 olduğu durumun daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Bayesyen MAP yetenek kestirim yöntemi kullanıldığında testteki madde sayısının artması, boyutlara ilişkin güvenilirlik katsayısında önemli bir artışa neden olmamaktadır. Maddeler-arası boyutluluk modeli için Bayesyen MAP ve Fisher'in puanlama yöntemine ilişkin bulgular karşılaştırıldığında, MAP yetenek kestirim yönteminin daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

**b) Madde-içi boyutluluk modeline ait sabit madde sayısı durdurma kuralı ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.3'teki madde-içi boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,76 ile 0,85 aralığında, ölçmenin standart hatasının 0,23 ile 0,42 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,92 ile 0,94 aralığında ve RMSD değerlerinin ise 0,35 ile 0,38 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30'dan 40'a çıkarıldığı durumda ise boyutlara ilişkin güvenilirlik katsayılarının arttığı görülmektedir. Ayrıca, boyutlara ilişkin standart hataların ve RMSD azaldığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı görülmektedir. Testteki ortalama madde sayısı 50 olduğunda durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en düşük ve korelasyon değerlerinin ise en yüksek olduğu görülmektedir.

Genel olarak madde-içi boyutluluk modeline ilişkin analiz bulguları incelendiğinde, madde seçim yöntemlerinden D-optimality, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak sabit madde sayısı kullanıldığında, testteki ortalama madde sayısının 40 olduğu durumun daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir. Ayrıca madde sayısı arttıkça boyutlara ilişkin standart hata ve RMSD değerleri azalırken, testin boyutlarına ilişkin güvenilirlik katsayılarının ise arttığı görülmektedir.

Tablo 4.3'teki madde-içi boyutluluk modeline ilişkin analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,83 ile 0,95 aralığında, ölçmenin standart hatasının 0,22 ile 0,38 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,91 ile 0,93 aralığında ve RMSD değerlerinin ise 0,40 ile 0,43 aralığında değiştiği görülmektedir.. Ayrıca testteki madde sayısı arttıkça standart hataların ve RMSD değerlerinin azaldığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı görülmektedir. Diğer test uzunlukları ile karşılaştırıldığında testteki ortalama madde sayısı 50 olduğu durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en düşük ve korelasyon değerlerinin ise en yüksek olduğu görülmektedir.

Genel olarak her bir durdurma kuralı için birinci boyuta (dinleme boyutuna) ilişkin güvenilirlik katsayılarının en yüksek ve diğer iki boyuta (dilbilgisi ve okuduğunu anlama boyutları) ilişkin güvenilirlik katsayılarının ise yüksek ve birbirine yakın değerler aldığı görülmektedir. Ayrıca, madde seçim yöntemlerinden D-optimality, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak sabit madde sayısı kullanıldığı durumda, boyutlara ilişkin güvenilirlik katsayılarının oldukça yüksek ve standart hataların düşük olmasından dolayı testi sonlandırmak için gerekli madde sayısı 30 olarak belirlenebilir.

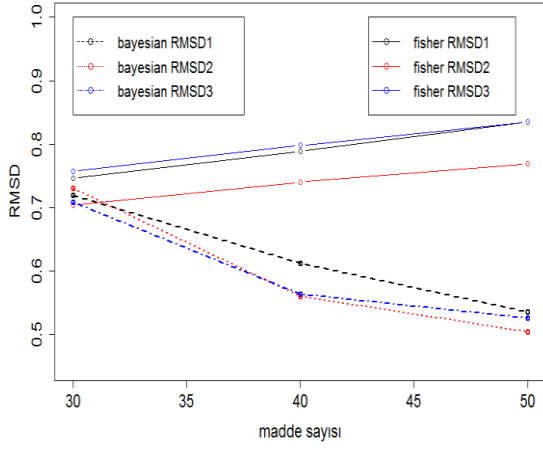
Madde-içi boyutluluk modeli için Bayesyen MAP yetenek kestirim yöntemi kullanıldığı koşullarda boyutlara ilişkin güvenilirlik katsayıları yüksek olduğundan testteki madde sayısının artması, güvenilirlik katsayılarında önemli bir artışa neden olmamaktadır. Maddeler-arası boyutluluk modeli için Bayesyen MAP ve Fisher'in puanlama yöntemine ilişkin bulgular karşılaştırıldığında ise, her bir koşul için MAP yetenek kestirim yönteminin daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Maddeler-arası ve madde –içi boyutluluk modeline ait farklı çok-boyutlu BOB testi koşullarına ilişkin bulgular karşılaştırıldığında, her iki model için aynı koşullarda Bayesyen yetenek kestirim yöntemine ait güvenilirlik katsayıları ve yetenek parametreleri arasındaki korelasyonun Fisher'in puanlama yöntemine göre daha yüksek ve standart hata ve RMSD değerlerinin ise daha düşük olduğu görülmektedir. Ayrıca, maddeler-arası boyutluluk modeli ile karşılaştırıldığında, madde-içi boyutluluk modelinin her iki yetenek kestirim yöntemi için daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği yorumu yapılabilir.

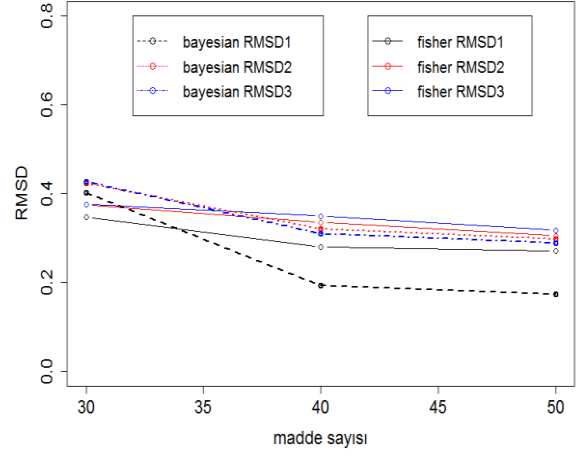
### Maddeler-arası Boyutluluk Modeli

### Madde-İçi Boyutluluk Modeli

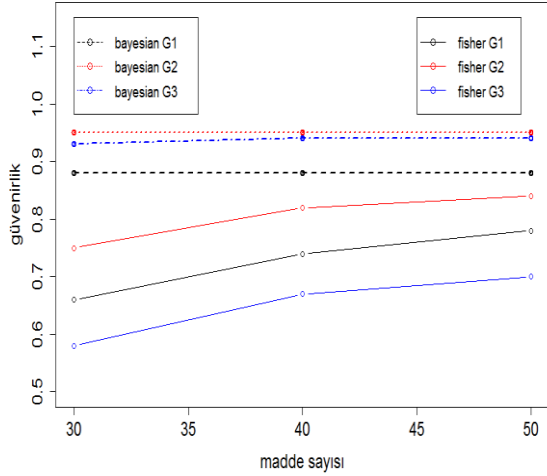
a.1 RMSD- Madde sayısı ilişkisi



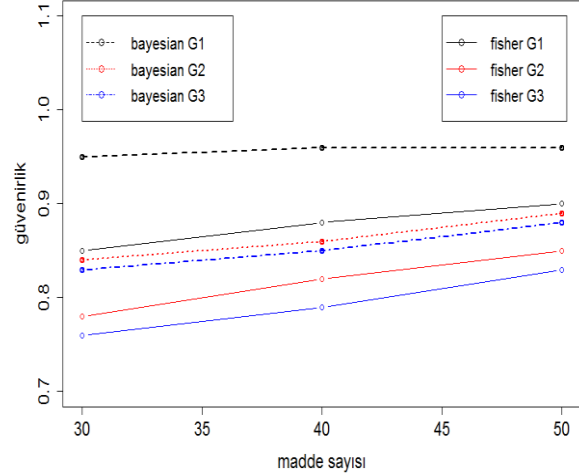
a.2 RMSD- Madde sayısı ilişkisi



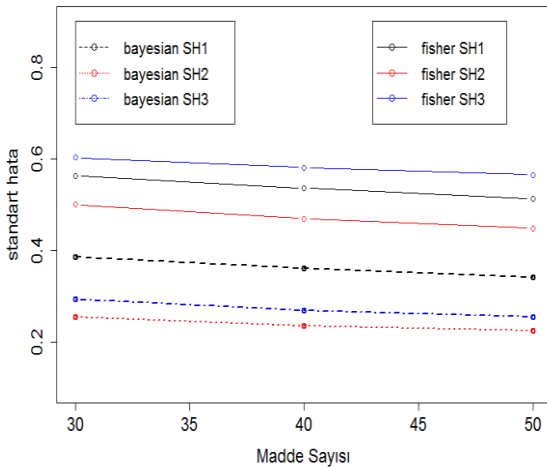
b.1 Güvenirlilik- Madde sayısı ilişkisi



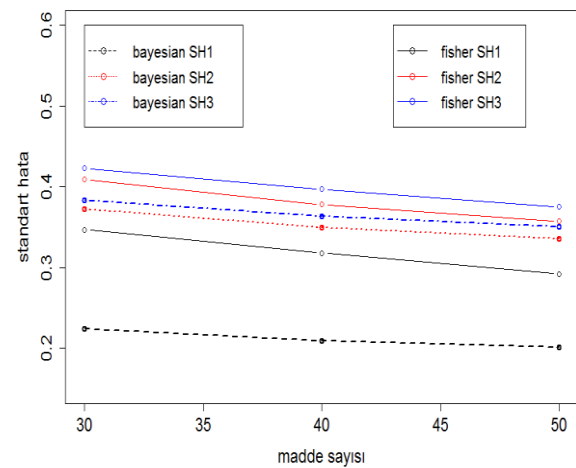
b.2 Güvenirlilik- Madde sayısı ilişkisi



c.1 Ölçmenin standart hatası- Madde sayısı



c.2 Ölçmenin standart hatası- Madde sayısı



Şekil 4.3: D-Optimality Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin Grafikler

Şekil 4.3'te maddeler-arası ve madde-içi çok boyutlu telafi-edici modellere ilişkin madde seçim yöntemlerinden D-optimality, testi durdurma kurallarından sabit madde sayısının kullanıldığı çok-boyutlu BOB testi analizlerine ait grafikler verilmiştir. Şekilde kesikli çizgiler Bayesyen yetenek kestirim yöntemine ait bulguları temsil ederken, düz çizgiler ise Fisher'in puanlama yetenek kestirim yöntemine ilişkin bulguları temsil etmektedir. Ayrıca grafikte yer alan siyah renkli çizgiler birinci boyuta (dinleme boyutuna) ait bulguları, kırmızı renkli çizgi ikinci boyuta (dilbilgisi boyutuna) ait bulguları ve mavi renkli çizgi ise üçüncü boyuta (okuduğunu anlama boyutuna) ait bulguları temsil etmektedir.

Şekil 4.3'te yer alan a.1 ve a.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine dayalı Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin RMSD değerleri ile testteki madde sayısı arasındaki ilişkiyi vermektedir. Şekil 4.3 a.1'e bakıldığında maddeler-arası boyutluluk modelinde, testteki madde sayısı arttıkça Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin azaldığı görülmektedir. Ancak, Fisher'in puanlama yönteminin kullanıldığı durumda, beklenenin aksine testteki madde sayısı arttıkça RMSD değerlerinin arttığı görülmektedir.

Şekil 4.3 a.2'e bakıldığında, testteki madde sayısı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin azaldığı görülmektedir. Ayrıca, maddeler-arası çok-boyutlu model ile karşılaştırıldığında, madde-içi boyutluluk modeline dayalı Bayesyen yetenek kestirim yöntemine ait RMSD değerlerinin daha düşük olduğu görülmektedir.

Şekil 4.3'te yer alan b.1 ve b.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin güvenilirlik katsayıları ile testteki madde sayısı arasındaki ilişkiyi vermektedir. Şekil 4.3 a.1'e bakıldığında, Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin güvenilirlik katsayılarının testteki madde sayısı 30'dan 40'a çıktığında boyutlara ilişkin güvenilirlik katsayılarının değişmediği görülmektedir. Fisher'in puanlama yöntemi ile karşılaştırıldığında, Bayesyen yetenek kestirim yöntemine ait güvenilirlik katsayılarının çok yüksek olduğu görülmektedir.

Şekil 4.3 b.2'ye bakıldığında, testteki madde sayısı arttıkça hem Fisher'in puanlama hemde Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin güvenilirlik katsayılarının arttığı görülmektedir. Ayrıca her iki çok-boyutlu model için, Bayesyen

yetenek kestirim yöntemine ait boyutlara ilişkin güvenilirlik katsayılarının ve Fisher'in puanlama yönteminin kullanıldığı çok-boyutlu BOB testi uygulamalarına ait güvenilirlik katsayılarından daha yüksek olduğu görülmektedir.

Şekil 4.3'da yer alan c.1 ve c.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testlerine ait her bir boyuta ilişkin ölçmenin standart hatası ile testteki madde sayısı arasındaki ilişkiyi vermektedir. Şekil 4.3 c.1'e bakıldığında maddeler-arası boyutluluk modelinde, Bayesyen MAP yetenek kestirimi ve Fisher'in puanlama yöntemine ait her bir boyuta ilişkin standart hatanın testteki madde sayısı arttıkça azalma eğiliminde olduğu görülmektedir. Ayrıca maddeler arası boyutluluk modelinde hem Bayesyen hem de Fisher'in puanlama yönteminin kullanıldığı BOB testine ait standart hatanın her bir koşul için bir birine yakın değerler aldığı görülmektedir.

Şekil 4.3 c.2'ye bakıldığında, testteki madde sayısı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin ölçmenin standart hatasının azaldığı ve birinci boyuta ait standart hata değerinin diğer boyutlarla kıyaslandığında daha düşük olduğu görülmektedir. Ayrıca madde-içi boyutluluk modelinde Bayesyen yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testi uygulamalarının daha düşük standart hata değerlerine sahip olduğu görülmektedir.

### **c) Maddeler-arası boyutluluk modeline ait farklı hata varyansı durdurma kuralları ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.4'te her bir çok-boyutlu model için madde seçim yöntemlerinden D-optimality, durdurma kurallarından hata varyansı ve yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait analiz bulgularına yer verilmiştir. Her bir model için standart hata durdurma kuralı kullanılarak analizler bireylerin kestirilen yetenek parametrelerine ait hata varyansının sırasıyla 0,20, 0,25 ve 0,30'un altına düştüğünde testler sonlandırılmıştır.

**Tablo 4.4: D-Optimality Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralına İlişkin BOB Testi Bulgular**

Çok Boyutlu Model	Yetenek kestirim yöntemi	Durdurma kuralı (s. hata)	Ortalama madde sayısı	Güvenirlilik Katsayısı			Ölçmenin Standart Hatası			Korelasyon			RMSD Değerleri		
				G1	G2	G3	SH1	SH2	SH3	Kor1	Kor2	Kor3	RMSD1	RMSD2	RMSD3
Maddeler-Arası Boyutluluk Modeli	Fisher	0,20	50,00	0,82	0,87	0,74	0,494	0,427	0,551	0,685	0,747	0,668	0,863	0,790	0,840
		0,25	50,00	0,78	0,84	0,70	0,514	0,446	0,565	0,693	0,750	0,666	0,812	0,753	0,812
		0,30	49,43	0,78	0,84	0,70	0,514	0,447	0,565	0,695	0,744	0,667	0,806	0,763	0,811
	Bayesyen	0,20	39,49	0,88	0,95	0,94	0,361	0,232	0,269	0,824	0,854	0,846	0,596	0,558	0,571
		<b>0,25</b>	<b>22,83</b>	<b>0,84</b>	<b>0,94</b>	<b>0,91</b>	<b>0,406</b>	<b>0,265</b>	<b>0,311</b>	<b>0,791</b>	<b>0,822</b>	<b>0,820</b>	<b>0,644</b>	<b>0,620</b>	<b>0,608</b>
		0,30	15,06	0,73	0,89	0,83	0,436	0,293	0,353	0,908	0,964	0,934	0,400	0,257	0,339
Madde-İçi Boyutluluk Modeli	Fisher	0,20	50,00	0,91	0,85	0,83	0,292	0,356	0,376	0,963	0,952	0,943	0,277	0,307	0,325
		0,25	44,42	0,90	0,84	0,83	0,303	0,368	0,386	0,961	0,949	0,943	0,283	0,315	0,327
		<b>0,30</b>	<b>33,08</b>	<b>0,88</b>	<b>0,81</b>	<b>0,79</b>	<b>0,332</b>	<b>0,398</b>	<b>0,411</b>	<b>0,951</b>	<b>0,937</b>	<b>0,932</b>	<b>0,317</b>	<b>0,352</b>	<b>0,353</b>
	Bayesyen	0,20	44,35	0,96	0,88	0,87	0,206	0,342	0,357	0,973	0,919	0,926	0,238	0,396	0,369
		<b>0,25</b>	<b>26,33</b>	<b>0,96</b>	<b>0,88</b>	<b>0,87</b>	<b>0,238</b>	<b>0,384</b>	<b>0,397</b>	0,918	0,897	0,901	0,421	0,455	0,440
		0,30	16,30	0,92	0,78	0,78	0,265	0,424	0,430	<b>0,825</b>	<b>0,803</b>	<b>0,802</b>	<b>0,663</b>	<b>0,670</b>	<b>0,667</b>

Tablo 4.4'te maddeler-arası boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,74 ile 0,87 aralığında, ölçmenin standart hatasının 0,43 ile 0,55 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,67 ile 0,75 aralığında ve RMSD değerlerinin ise 0,79 ile 0,84 aralığında değiştiği görülmektedir. Standart hata durdurma kuralının 0,20 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise, İngilizce yeterlik sınavının kâğıt-kalem testi formatındaki ortalama madde sayısı göz önünde bulundurularak çok-boyutlu BOB testi analizlerinde testteki maksimum madde sayısı 50 ile sınırlandırılmıştır. Böylece, standart hata durdurma kuralının kullanıldığı çok-boyutlu BOB testi analizlerinde testteki madde sayısının 50'den fazla olması engellenerek çok uzun testler ile yetenek kestirimi önlenmiştir.

Standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında ise testteki ortalama madde sayısı değişmediğinden, her bir boyuta ilişkin güvenilirlik katsayılarının az miktarda azaldığı ve standart hataların ise arttığı görülmektedir. Buna karşın, boyutlara ilişkin RMSD değerinin azaldığı ve boyutlara ilişkin gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı görülmektedir. Diğer standart hata durdurma kuralları ile karşılaştırıldığında, standart hata durdurma kuralının 0,30 olduğu durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en yüksek ve korelasyon değerlerinin ise en düşük değere sahip olduğu görülmektedir.

Genel olarak, maddeler-arası boyutluluk modeline ait D-optimality madde seçim yöntemi ve hata varyansı durdurma kuralına ilişkin çok-boyutlu BOB testi analiz bulgularına bakıldığında, her bir koşul için elde edilen güvenilirlik katsayılarının birbirine yakın ve düşük olduğu görülmektedir. Dolayısıyla analiz bulgularının güvenilir ve tutarlı sonuçlar vermediği yorumu yapılabilir.

Tablo 4.4'te yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,88 ile 0,95 aralığında, ölçmenin standart hatasının 0,23 ile 0,36 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,82 ile 0,85 aralığında ve RMSD değerlerinin ise 0,56 ile 0,60 aralığında değiştiği görülmektedir. Standart hata durdurma kuralının 0,20 olduğu durumda



bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 39,5'dir. Standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında ise her bir boyuta ilişkin güvenilirlik katsayılarının azda olsa düştüğü görülmektedir. Ayrıca, standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında, boyutlara ilişkin ölçmenin standart hatasının ve RMSD değerlerinin arttığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin düştüğü görülmektedir. Bunun temel sebebi, standart hata durdurma kuralının 0,25 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısının 39,5'ten 22,8'e düşmesidir. Hata varyansı durdurma kuralının 0,30 olduğunda durumda testteki ortalama madde sayısı 15 olurken, her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en yüksek, güvenilirlik ve korelasyon değerlerinin ise en düşük değere sahip olduğu görülmektedir.

Maddeler-arası boyutluluk modeline ilişkin analiz sonuçlarına bakıldığında, madde seçim yöntemlerinden D-optimality, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve standart hata durdurma kuralı kullanıldığında, standart hatanı 0,25'e eşitlendiği ve ortalama madde sayısının 22,8 olduğu durumun daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Maddeler-arası boyutluluk modeli için Bayesyen MAP ve Fisher'in puanlama yöntemine ilişkin bulgular karşılaştırıldığında, her bir koşul için Bayesyen MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir. Fisher'in puanlama yönteminin kullanıldığı durumda boyutlara ilişkin hata varyanslarının azalması testteki ortalama madde sayısını ve her boyuta ilişkin güvenilirlik katsayılarını etkilemediği görülmektedir.

#### **d) Madde-içi boyutluluk modeline ait farklı standart hata durdurma kuralı ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.4'te yer alan madde-içi boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak hata varyansının 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,83 ile 0,91 aralığında, ölçmenin standart hatasının 0,29 ile 0,38 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,94 ile 0,96 aralığında ve RMSD değerlerinin ise 0,28 ile 0,33 aralığında değiştiği görülmektedir. Hata varyansı durdurma kuralının 0,20 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 50'dir. Standart

hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında ise testteki ortalama madde sayısı 50'den 44,4'e düşmesine rağmen, her bir boyuta ilişkin güvenilirlik katsayılarının ve korelasyon değerlerinin 0,01 düzeyinde azaldığı, boyutlara ilişkin RMSD ve standart hata değerlerinin de 0,01 düzeyinde arttığı görülmektedir. Hata varyansı durdurma kuralının 0,30 olduğunda durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en yüksek, her bir boyuta ilişkin güvenilirlik ve korelasyon değerlerinin ise en düşük değere sahip olduğu görülmektedir. Standart hata durdurma kuralının 0,30 olduğu durumda testteki ortalama madde sayısının ise 33 olduğu görülmektedir.

genel olarak, madde-içi boyutluluk modeli için madde seçim yöntemlerinden D-optimality, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve standart hata durdurma kuralının 0,25 olduğu koşula ait testteki ortalama madde sayısının 31,4 olup daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Tablo 4.4'te madde-içi boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,88 ile 0,96 aralığında, ölçmenin standart hatasının 0,21 ile 0,36 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,92 ile 0,97 aralığında ve RMSD değerlerinin ise 0,24 ile 0,40 aralığında değiştiği görülmektedir. Standart hata durdurma kuralının 0,20 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 44,4'tür. Standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı 44,4'ten 26,3'e düşmesine rağmen boyutlara ait güvenilirlik katsayılarının değişmediği, standart hata ve RMSD değerlerinin ise arttığı görülmektedir. Hata varyansı durdurma kuralının 0,30 olduğunda durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en yüksek ve korelasyon ve güvenilirlik katsayıların ise en düşük değere sahip olduğu görülmektedir.

Madde-içi boyutluluk modeline ilişkin analiz sonuçlarına bakıldığında, madde seçim yöntemlerinden D-optimality, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi kullanıldığında, standart hatanı 0,25'e eşitlendiği durumda ortalama madde sayısının 26'ya eşit olup diğer durumlarla karşılaştırıldığında daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Madde-içi boyutluluk modeli için standart hata durdurma kuralı 0,25 olarak belirlendiği durumda Fisher'in puanlama yetenek kestirim yöntemi kullanıldığında, testteki

ortalama madde sayısı 44,4 olup güvenilirlik katsayıları 0,83 ile 0,94 aralığında değişirken; Bayesyen MAP yetenek kestirim yöntemi kullanıldığında ise testteki ortalama madde sayısı 26,3 olup güvenilirlik katsayıları 0,87 ile 0,96 aralığında değiştiği görülmektedir. Bundan dolayı, madde-içi boyutluluk modeli için hata varyansı durdurma kuralı kullanıldığı durumda, her bir koşul için Bayesyen MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir.

Maddeler-arası ve madde –içi boyutluluk modeline ilişkin farklı çok-boyutlu BOB testi koşullarına ilişkin bulgular karşılaştırıldığında, her iki model için aynı koşullarda Bayesyen yetenek kestirim yöntemine ait güvenilirlik katsayıları ve yetenek parametreleri arasındaki korelasyonun Fisher'in puanlama yöntemine göre daha yüksek; standart hata ve RMSD değerlerinin ise daha düşük olduğu görülmektedir. Ayrıca, maddeler-arası boyutluluk modeli ile karşılaştırıldığında, madde-içi boyutluluk modelinin her iki yetenek kestirim yöntemi için daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği yorumu yapılabilir.

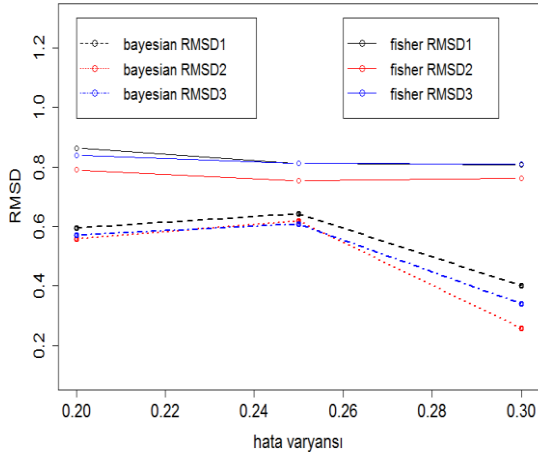
Şekil 4.4'te maddeler-arası ve madde-içi çok boyutlu telafi-edici modellere ilişkin madde seçim yöntemlerinden D-optimality, testi durdurma kurallarından boyutlara ilişkin hata varyansı ve yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testi analizlerine ait grafikler verilmiştir.

Şekil 4.4'te yer alan a.1 ve a.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin RMSD değerleri ile yetenek parametrelerine ilişkin hata varyansları arasındaki ilişkiyi vermektedir.

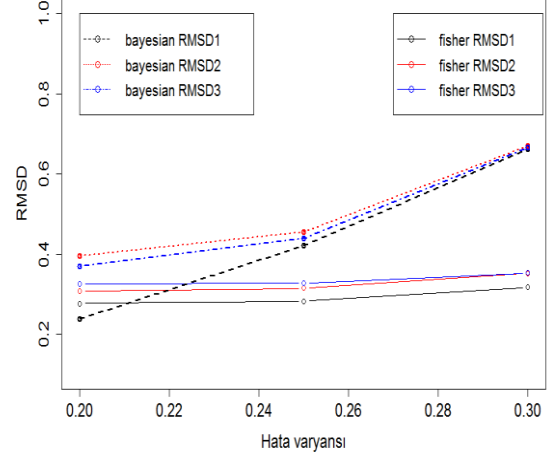
### Maddeler-arası Boyutluluk Modeli

### Madde-İçi Boyutluluk Modeli

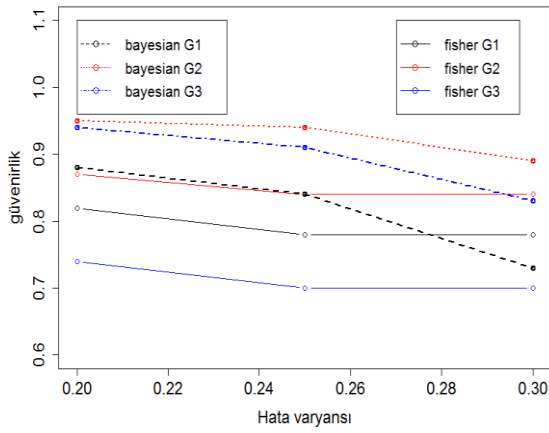
a.1 RMSD- Hata varyansı ilişkisi



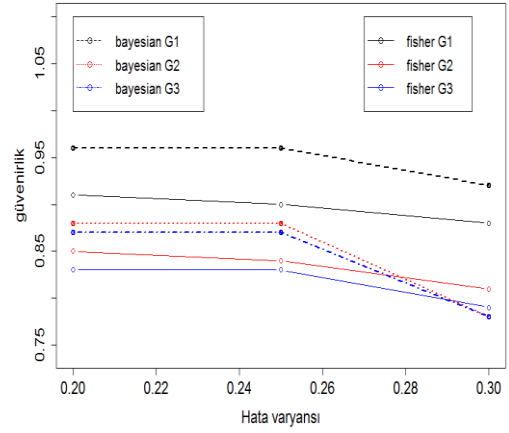
a.2 RMSD- Hata varyansı ilişkisi



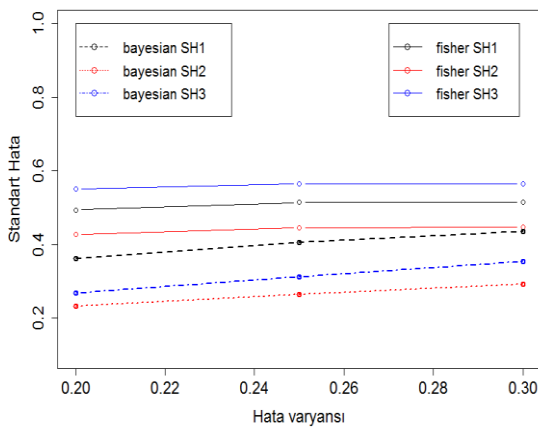
b.1 Güvenirlilik- Hata varyansı ilişkisi



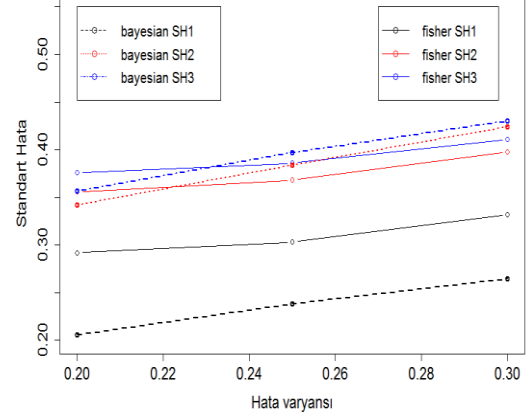
b.2 Güvenirlilik- Hata varyansı ilişkisi



c.1 Ölçmenin standart hatası- Hata varyansı



c.2 Ölçmenin standart hatası- Hata varyansı



Şekil 4.4: D-optimality Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralına İlişkin Grafikler

Şekil 4.4 a.1'e bakıldığında maddeler-arası boyutluluk modelinde, yetenek parametrelerine ilişkin hata varyansı arttıkça Fisher'in puanlama yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin değişmediği görülmektedir. Bunun temel sebebi, boyutlara ilişkin hata varyansları artmasına rağmen, testteki ortalama madde sayısının değişmemesidir. Maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testlerinde, D-optimality madde seçim yönteminin kullanıldığı durumda hem Bayesyen hem de Fisher'in puanlama yetenek kestirimi yönteminin güvenilir ve tutarlı sonuçlar vermediği söylenebilir.

Şekil 4.4 a.2'ye bakıldığında, kestirilen yetenek parametrelerine ilişkin hata varyansı arttıkça Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin arttığı görülmektedir. Buna karşın, kestirilen yetenek parametrelerine ilişkin hata varyansı arttıkça Fisher'in puanlama yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin değişmediği görülmektedir. Dolayısıyla, maddeler-arası boyutluluk modeli ile karşılaştırıldığında, madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testlerinde, D-optimality madde seçim yönteminin kullanıldığı durumda Bayesyen MAP yetenek kestirimi yönteminin daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir. Ayrıca maddeler-arası boyutluluk modeli için yetenek parametrelerine ilişkin hata varyansı arttıkça Bayesen ve Fisher'in puanlama yöntemine ait RMSD değerleri arasındaki farkın arttığı görülmektedir.

Şekil 4.4'te yer alan b.1 ve b.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin güvenilirlik katsayıları ile yetenek parametrelerine ait hata varyansları arasındaki ilişkiyi vermektedir. Şekil 4.4 b.1'e bakıldığında maddeler-arası boyutluluk modelinde, yetenek parametrelerine ait hata varyansları arttıkça, Bayesyen MAP yetenek kestirimine yöntemine ait her bir boyuta ilişkin güvenilirlik katsayılarının düzenli bir şekilde azaldığı görülmektedir. Ancak, Fisher'in puanlama yönteminin kullanıldığı durumda hata varyansı arttıkça boyutlara ilişkin güvenilirlik katsayılarının artıp azaldığı görülmektedir. Ayrıca her bir hata varyansı durdurma kuralı için Bayesyen yetenek kestirim yöntemine ait boyutlara ilişkin güvenilirlik katsayılarının Fisher'in puanlama yetenek kestirim yöntemine ait güvenilirlik katsayılarından daha yüksek olduğu görülmektedir.

Şekil 4.4 b.2'ye bakıldığında, her bir boyuta ilişkin hata varyansı arttıkça Fisher'in puanlama yetenek kestirim yöntemine ait güvenilirlik katsayılarının azaldığı

görülmektedir. Ancak, Bayesyen MAP yetenek kestirimine ait güvenilirlik katsayılarına bakıldığında hata varyansı 0,20'den 0,25'e çıktığında güvenilirlik katsayıları değişmezken, hata varyansı 0,30'a çıktığında ise azaldığı görülmektedir. Maddeler-arası boyutluluk modelinde olduğu gibi, her bir hata varyansı durdurma kuralı için Bayesyen yetenek kestirim yöntemine ait boyutlara ilişkin güvenilirlik katsayılarının Fisher'in puanlama yetenek kestirim yöntemine ait güvenilirlik katsayılarından daha yüksek olduğu görülmektedir. Dolayısıyla, madde-içi boyutluluk modeline dayalı çok boyutlu BOB testi analizinde hata varyansı durdurma kuralı kullanıldığında Bayesyen yetenek kestirim metodunun daha güvenilir sonuçlar verdiği yorumu yapılabilir.

Şekil 4.4'te yer alan c.1 ve c.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testlerine ait her bir boyuta ilişkin ölçmenin standart hatası ile hata varyansı arasındaki ilişkiyi vermektedir. Şekil 4.4 c.1'e bakıldığında, Bayesyen MAP yetenek kestirimi yöntemi kullanıldığında, yetenek parametrelerine ilişkin hata varyansı arttıkça, her bir boyuta ilişkin ölçmenin standart hatasının da arttığı görülmektedir. Ayrıca, maddeler-arası boyutluluk modelinde, her bir hata varyansı durdurma kuralı için Bayesyen yetenek kestirim yöntemine ait boyutlara ilişkin standart hata değerlerinin Fisher'in puanlama yetenek kestirim yöntemine ait standart hata değerlerinden daha düşük olduğu görülmektedir.

Şekil 4.4 c.2'ye bakıldığında, madde-içi boyutluluk modeli için boyutlara ilişkin hata varyansı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin ölçmenin standart hatasının arttığı görülmektedir. Ancak, Bayesyen MAP yetenek kestirimine ait birinci boyuta ilişkin standart hata değerinin, Fisher'in puanlama yetenek kestirim yöntemine ait birinci boyuta ilişkin standart hata değerinden daha düşük olduğu görülmektedir.

Genel olarak D-optimality madde seçim yöntemine ait hata varyansı durdurma kuralının kullanıldığı madde-içi ve maddeler-arası çok boyutlu BOB testi analiz yöntemlerine ilişkin bulgular karşılaştırıldığında, her iki çok boyutlu model için Bayesyen MAP yetenek kestirim yöntemine ait en uygun hata varyansı durdurma kuralının 0,25 olduğu görülmektedir. Fisher'in puanlama yetenek kestirim yöntemine ait en uygun hata varyansı durdurma kuralının ise 0,30 olduğu görülmektedir. Dolayısıyla, Fisher'in puanlama yönteminin madde düzeyindeki boyutluluk modellerinden ve testte yer alan madde sayısından etkilendiği ve madde-içi boyutluluk

modeline dayalı çok-boyutlu BOB testi uygulamalarında daha güvenilir ve kararlı sonuçlar verdiği görülmektedir. Diğer taraftan, Bayesyen yetenek kestirim yönteminin madde düzeyindeki boyutluluk modellerinden etkilenmediği ve özellikle madde-içi boyutluluk modelinin kullanıldığı durumda daha az madde ile daha güvenilir sonuçlar verdiği yorumu yapılabilir.

**Alt Problem 3: Seçkisiz (Random) madde seçim yönteminin kullanıldığı madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB test analizlerine ilişkin her bir koşula ait güvenilirlik katsayısı, standart hata, yetenek parametrelerine ait korelasyon, testin uzunluğu ve RMSD değerleri nasıldır?**

Bu bölümde çok-boyutlu BOB testleri uygulamalarında maddelerin her hangi bir kurala bağlı olmadan rastgele seçildiği seçkisiz (random) madde seçim yöntemine ilişkin madde-içi ve maddeler-arası boyutluluk modeline dayalı farklı yetenek kestirim ve durdurma kurallarının kullanıldığı çok-boyutlu BOB testi analizlerine ait güvenilirlik katsayısı, standart hata, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon ve RMSD değerlerine ilişkin bulgulara yer verilmiştir. Seçkisiz madde seçim yönteminin kullanılması, her boyuta ilişkin kestirilen yetenek parametrelerine ait hata varyansını en aza indirmeye çalışan A-optimality ve D-optimality madde seçim yöntemlerinin boyutlara ilişkin güvenilirlik katsayıları ve hata varyansları üzerindeki etkisinin görülmesine olanak sağladığından önemli görülmektedir.

**a) Maddeler-arası boyutluluk modeline ait sabit madde sayısı durdurma kuralları ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.5'te her bir çok-boyutlu model için madde seçim yöntemlerinden seçkisiz madde seçim yöntemi, durdurma kurallarından sabit soru sayısı, yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait analiz bulgularına yer verilmiştir. Her bir model için sabit madde sayısı durdurma kuralı kullanılarak analizler bireyin cevapladığı madde sayısı sırasıyla 30, 40 ve 50'ye ulaştığında testler sonlandırılmış ve her bir model için analiz bulguları karşılaştırılmıştır.

**Tablo 4.5: Seçkisiz (Random) Madde Seçim Yöntemi ve Sabit Madde Sayısı Durdurma Kuralına İlişkin Analiz Bulguları**

Çok-Boyutlu Model	Yetenek Kestirim yöntemi	Durdurma kuralı (sabit soru sayısı)	Güvenirlilik Katsayısı			Ölçmenin Standart Hatası			Korelasyon			RMSD Değerleri		
			G1	G2	G3	SH1	SH2	SH3	Kor1	Kor2	Kor3	RMSD1	RMSD2	RMSD3
Maddeler-Arası Boyutluluk Modeli	Fisher	K												
		30	0,61	0,69	0,67	0,502	0,460	0,469	0,874	0,892	0,884	0,463	0,433	0,441
		40	0,70	0,78	0,75	0,464	0,415	0,428	0,896	0,914	0,910	0,424	0,387	0,391
	50	0,73	0,80	0,78	0,435	0,381	0,395	0,915	0,923	0,930	0,387	0,368	0,347	
	Bayesyen	30	0,61	0,68	0,67	0,500	0,455	0,466	0,868	0,889	0,877	0,474	0,439	0,454
		40	0,70	0,77	0,75	0,464	0,418	0,427	0,900	0,914	0,912	0,415	0,388	0,387
50		0,74	0,81	0,79	0,437	0,385	0,397	0,916	0,923	0,919	0,382	0,367	0,371	
Madde-İçi Boyutluluk Modeli	Fisher	30	0,78	0,64	0,73	0,438	0,496	0,511	0,547	0,572	0,547	1,198	1,123	1,137
		40	0,79	0,74	0,73	0,394	0,456	0,475	0,601	0,575	0,581	1,161	1,155	1,137
		<b>50</b>	<b>0,85</b>	<b>0,77</b>	<b>0,73</b>	<b>0,354</b>	<b>0,425</b>	<b>0,445</b>	<b>0,944</b>	<b>0,927</b>	<b>0,925</b>	<b>0,342</b>	<b>0,378</b>	<b>0,374</b>
	Bayesyen	30	0,76	0,67	0,64	0,434	0,492	0,509	0,918	0,893	0,888	0,409	0,452	0,450
		40	0,82	0,73	0,70	0,388	0,452	0,471	0,933	0,917	0,911	0,372	0,403	0,406
		<b>50</b>	<b>0,85</b>	<b>0,78</b>	<b>0,74</b>	<b>0,359</b>	<b>0,425</b>	<b>0,446</b>	<b>0,942</b>	<b>0,928</b>	<b>0,924</b>	<b>0,345</b>	<b>0,375</b>	<b>0,374</b>



Tablo 4.5'te seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,61 ile 0,69 aralığında, ölçmenin standart hatasının 0,47 ile 0,50 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,87 ile 0,89 aralığında ve RMSD değerlerinin ise 0,44 ile 0,46 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30'dan 40'a çıkarıldığı durumda ise boyutlara ilişkin güvenilirlik katsayılarının ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı, standart hataların ve RMSD değerlerinin ise azaldığı görülmektedir. Testteki ortalama madde sayısı 50 olduğu durumda ise her bir boyuta ilişkin standart hatalarının en düşük ve güvenilirlik katsayılarının ise en yüksek olduğu görülmektedir. Genel olarak, testteki ortalama madde sayısı 50 olduğu durumda bile, boyutlara ilişkin en yüksek güvenilirlik katsayısının 0,80'e eşit olduğu ve güvenilirlik katsayılarının diğer koşullar ile karşılaştırıldığında düşük olduğu görülmektedir.

Maddeler-arası boyutluluk modeline ait BOB testi analiz bulgularına bakıldığında, her bir durdurma kuralı için boyutlara ilişkin güvenilirlik katsayılarının ve korelasyon değerlerinin düşük, boyutlara ilişkin ölçmenin standart hatasının ve RMSD değerlerinin yüksek olduğu görülmektedir. Maddelerin herhangi bir kurala bağlı olmadan rastgele seçildiği maddeler-arası boyutluluk modeline ilişkin analiz sonuçlarına bakıldığında, Fisher'in puanlama yetenek kestirim yöntemi için testteki ortalama madde sayısının 50 olduğu durumunda bile güvenilir ve tutarlı sonuçlar elde edilmediği görülmektedir.

Tablo 4.5'te maddeler-arası boyutluluk modeline ilişkin yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyuta ilişkin güvenilirlik katsayılarının 0,61 ile 0,68 aralığında, ölçmenin standart hatasının 0,46 ile 0,50 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,87 ile 0,89 aralığında ve RMSD değerlerinin ise 0,44 ile 0,47 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30'dan 40'a çıkarıldığı durumda ise boyutlara ilişkin güvenilirlik katsayıları ile gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı, standart hataların ve RMSD değerlerinin ise azaldığı görülmektedir. Testteki ortalama madde sayısı 50 olduğunda durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en düşük ve korelasyon değerlerinin ise en yüksek olduğu görülmektedir.

Ayrıca, testteki ortalama madde sayısı 50 olduğu durumda bile, boyutlara ilişkin en yüksek güvenilirlik katsayısının 0,81'e eşit olduğu ve güvenilirlik katsayılarının diğer koşullar ile karşılaştırıldığında düşük olduğu görülmektedir.

Seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası boyutluluk modeline ait BOB testi analiz bulgularına bakıldığında, genel olarak her bir durdurma kuralı için birinci boyuta (dinleme boyutuna) ilişkin güvenilirlik katsayılarının en düşük, ikinci boyuta (dilbilgisi boyutuna) ilişkin güvenilirlik katsayısının ise en yüksek değerlere sahip olduğu görülmektedir. Testteki ortalama madde sayısı 50 olduğu koşulda bile hem Fisher'in puanlam hemde Bayesyen MAP yöntemine ilişkin güvenilirlik katsayılarının birbirine yakın ve düşük olduğu söylenebilir.

**b) Madde-içi boyutluluk modeline ait sabit madde sayısı durdurma kuralları ve farklı yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.5'teki seçkisiz madde seçim yönteminin kullanıldığı, madde-içi boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,64 ile 0,78 aralığında, ölçmenin standart hatasının 0,44 ile 0,51 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,55 ile 0,57 aralığında ve RMSD değerlerinin ise 1,12 ile 1,19 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30'dan 40'a çıkarıldığı durumda ise boyutlara ilişkin güvenilirlik katsayıları ile gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin arttığı ve standart hataların azaldığı görülmektedir. Buna karşın, madde sayısı 30'dan 40'a çıktığında boyutlara ilişkin RMSD değerinin azalması beklenirken, düzenli bir azalma göstermediği görülmektedir. Testteki ortalama madde sayısı 50 olduğu durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en düşük, güvenilirlik ve korelasyon değerlerinin ise en yüksek olduğu görülmektedir. Ayrıca, testteki ortalama madde sayısı 50 olduğu durumda bile güvenilirlik katsayılarının diğer koşullar ile karşılaştırıldığında düşük olduğu görülmektedir.

Genel olarak, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi, durdurma kuralı olarak sabit madde sayısı ve seçkisiz madde seçim yönteminin kullanıldığı madde-içi boyutluluk modeline ilişkin analiz bulguları incelendiğinde, testteki ortalama madde sayısının 50 olduğu durumda bile güvenilir ve tutarlı sonuçların elde

edilemediği söylenebilir. Ayrıca madde sayısı arttıkça boyutlara ilişkin standart hata değerleri kısmen azalırken, testin boyutlarına ilişkin güvenilirlik katsayılarının ise arttığı görülmektedir.

Tablo 4.5'teki maddelerin seçkisiz olarak seçildiği madde-içi boyutluluk modeline ilişkin analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak madde sayısını 30 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,64 ile 0,76 aralığında, ölçmenin standart hatasının 0,43 ile 0,51 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,89 ile 0,91 aralığında ve RMSD değerlerinin ise 0,40 ile 0,45 aralığında değiştiği görülmektedir. Testteki ortalama madde sayısı 30'dan 40'a çıkarıldığı durumda ise boyutlara ilişkin güvenilirlik katsayılarının arttığı görülmektedir. Ayrıca testteki madde sayısı arttıkça standart hataların ve RMSD değerlerinin azaldığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun kısmen arttığı görülmektedir. Ayrıca, testteki ortalama madde sayısı 50 olduğu durumda bile, güvenilirlik katsayılarının diğer koşullar ile karşılaştırıldığında düşük olduğu görülmektedir.

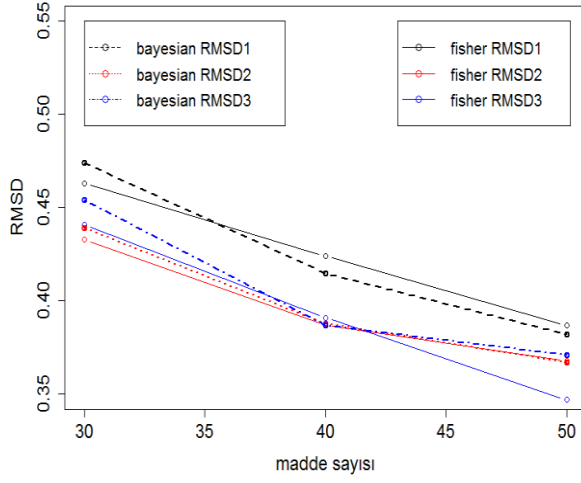
Seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası boyutluluk modeli için Bayesyen MAP ve Fisher'in puanlama yöntemine ilişkin bulgular karşılaştırıldığında, her bir koşul için hem Bayesyen MAP hem de Fisher'in puanlama yöntemine ilişkin güvenilirlik katsayılarının birbirine yakın ve düşük olduğu söylenebilir. Madde-içi boyutluluk modeli için Bayesyen MAP ve Fisher'in puanlama yöntemine ilişkin bulgular karşılaştırıldığında, her iki yetenek kestirim yöntemlerine ilişkin analiz bulgularının birbirine benzer olduğu görülmektedir.

Şekil 4.5 'te maddeler-arası ve madde-içi çok boyutlu telafi-edici modellere ilişkin madde seçim yöntemlerinden seçkisiz madde seçim, testi durdurma kurallarından sabit madde sayısının ve yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testi analizlerine ait grafikler verilmiştir. Şekilde kesikli çizgiler Bayesyen yetenek kestirim yöntemine ait bulguları temsil ederken, düz çizgiler ise Fisher'in puanlama yetenek kestirim yöntemine ilişkin bulguları temsil etmektedir. Ayrıca grafikte yer alan siyah renkli çizgiler birinci boyuta (dinleme boyutuna) ait bulguları, kırmızı renkli çizgi ikinci boyuta (dibilgisi boyutuna) ait bulguları ve mavi renkli çizgi ise üçüncü boyuta (okuduğunu anlama boyutuna) ait bulguları temsil etmektedir.

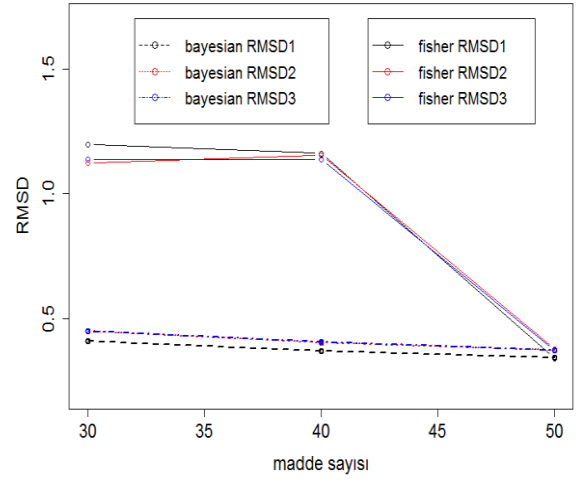
### Maddeler-arası Boyutluluk Modeli

### Madde-içi Boyutluluk Modeli

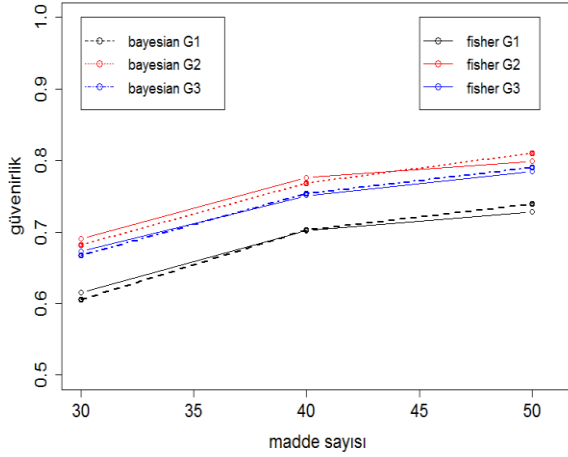
a.1 RMSD- Madde sayısı ilişkisi



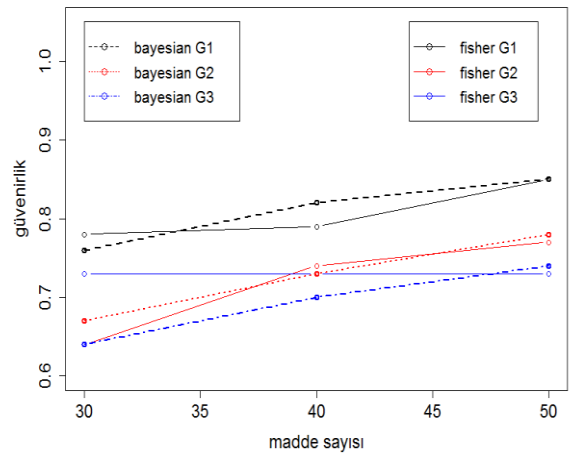
a.2 RMSD- Madde sayısı ilişkisi



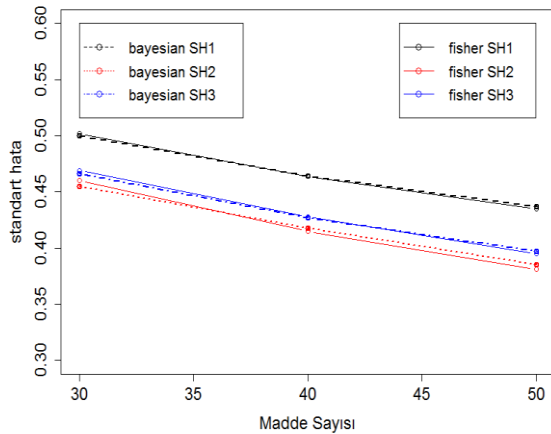
b.1 Güvenirlilik- Madde sayısı ilişkisi



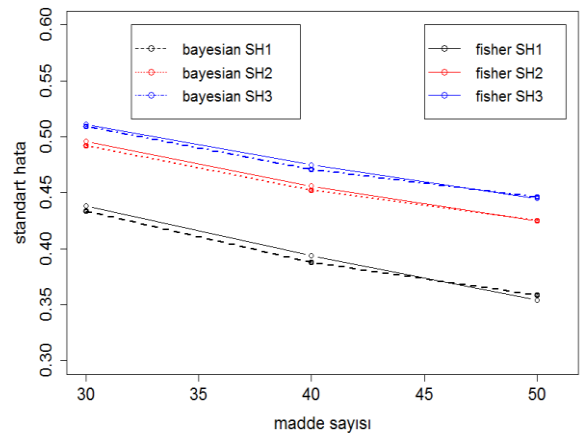
b.2 Güvenirlilik- Madde sayısı ilişkisi



c.1 Ölçmenin standart hatası- Madde sayısı



c.2 Ölçmenin standart hatası- Madde sayısı



Şekil 4.5 : Seçkisiz Madde Seçim Yöntemi Ve Sabit Madde Sayısı Durdurma Kuralına İlişkin Grafikler

Şekil 4.5'te yer alan a.1 ve a.2 grafikleri sırasıyla maddeler-arası ve madde-içi boyutluluk modellerine dayalı Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin RMSD değerleri ile testteki madde sayısı arasındaki ilişkiyi vermektedir. Şekil 4.5 a.1'e bakıldığında maddeler-arası boyutluluk modelinde, testteki madde sayısı arttıkça hem Bayesyen MAP hem de Fisher'in puanlama yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin azaldığı görülmektedir.

Şekil 4.5 a.2'e bakıldığında, testteki madde sayısı arttıkça Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin azaldığı görülmektedir. Fisher'in puanlama yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerine bakıldığında testteki ortalama madde sayısı 30'dan 40'a çıktığında RMSD değerlerinde anlamlı bir değişim olmazken, testteki ortalama madde sayısı 50 olduğunda RMSD değerlerinin düştüğü ve Bayesyen yetenek kestirim yöntemine ait RMSD değerlerine yaklaştığı görülmektedir. Ayrıca, maddeler-arası çok-boyutlu model ile karşılaştırıldığında, madde-içi boyutluluk modeline dayalı Fisher'in yetenek kestirim yöntemine ait RMSD değerlerinin daha yüksek olduğu görülmektedir.

Şekil 4.5'te yer alan b.1 ve b.2 grafikleri sırasıyla seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin güvenilirlik katsayıları ile testteki madde sayısı arasındaki ilişkiyi vermektedir. Şekil 4.5 b.1'e bakıldığında maddeler-arası boyutluluk modelinde, hem Bayesyen MAP hem de Fisher'in puanlama yetenek kestirim yöntemine ait her bir boyuta ilişkin güvenilirlik katsayılarının testteki madde sayısı arttıkça düzenli bir şekilde arttığı görülmektedir. Ayrıca hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirim yöntemine ait her bir boyuta ilişkin güvenilirlik katsayılarının birbirine yakın ve testteki ortalama madde sayısı 50 olduğu durumda 0,80'den düşük değerler aldığı görülmektedir.

Şekil 4.5 b.2'ye bakıldığında, testteki madde sayısı arttıkça Bayesyen MAP yetenek kestirim yöntemine ait her bir boyuta ilişkin güvenilirlik katsayılarının arttığı görülmektedir. Ancak, aynı durum Fisher'in puanlama yetenek kestirim yöntemi için geçerli değildir. Her iki çok-boyutlu model için, Bayesyen yetenek kestirim yöntemine ait boyutlara ilişkin güvenilirlik katsayılarının ve Fisher'in puanlama yönteminin kullanıldığı çok-boyutlu BOB testi uygulamalarına ait güvenilirlik katsayıların birbirine yakın değerlere sahip oldukları görülmektedir. Ayrıca her iki yetenek kestirim yöntemi

için madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi koşullarına ilişkin güvenilirlik katsayılarının, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi koşullarına ilişkin güvenilirlik katsayılarından daha yüksek olduğu söylenebilir.

Şekil 4.5 'de yer alan c.1 ve c.2 grafikleri sırasıyla seçkisiz madde seçim yöntemine ait maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testlerinde her bir boyuta ilişkin ölçmenin standart hatası ile testteki madde sayısı arasındaki ilişkiyi vermektedir. Şekil 4.5 c.1'e bakıldığında maddeler-arası boyutluluk modelinde, Bayesyen MAP yetenek kestirimi ve Fisher'in puanlama yöntemine ait her bir boyuta ilişkin standart hatanın testteki madde sayısı arttıkça azalma eğiliminde olduğu görülmektedir. Ayrıca maddeler arası boyutluluk modelinde hem Bayesyen hem de Fisher'in puanlama yönteminin kullanıldığı BOB testine ait standart hatanın her bir koşul için bir birine yakın değerler aldığı görülmektedir.

Şekil 4.5 c.2 grafiğine bakıldığında, testteki madde sayısı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin ölçmenin standart hatasının azaldığı görülmektedir. Ayrıca madde-içi ve maddeler-arası boyutluluk modellerine dayalı seçkisiz madde seçim yöntemi ve farklı yetenek kestirim yöntemlerinin kullanıldığı çok-boyutlu BOB testi uygulamalarının birbirine yakın değerler aldığı görülmektedir.

### **c) Maddeler-arası boyutluluk modeline ait farklı standart hata durdurma kuralı ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.6'da her bir çok-boyutlu model için madde seçim yöntemlerinden seçkisiz madde seçim yöntemi, durdurma kurallarından hata varyansı, yetenek kestirim yöntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait analiz bulgularına yer verilmiştir. Her bir model için standart hata durdurma kuralı kullanılarak analizler bireylerin kestirilen yetenek parametrelerine ait hata varyansının sırasıyla 0,20, 0,25 ve 0,30'un altına düştüğünde testler sonlandırılmıştır. Bireyin yabancı dil yeteneğini ölçmeyi amaçlayan en uygun çok-boyutlu BOB testi yöntemine ilişkin en uygun yetenek kestirim yöntemi, madde seçim yöntemi ve durdurma kuralına karar vermek için her bir çok-boyutlu BOB testi koşuluna ilişkin analiz bulguları karşılaştırılmıştır.

**Tablo 4.6: Seçkisiz Madde Seçim Yöntemi ve Standart Hata Durdurma Kuralı Yöntemine İlişkin BOB Testi Bulgular**

Çok Boyutlu Model	Yetenek kestirim yöntemi	Durdurma kuralı (s. hata)	Ortalama madde sayısı	Güvenirlilik Katsayısı			Ölçmenin Standart Hatası			Korelasyon			RMSD Değerleri		
				S. Hata	K	G1	G2	G3	SH1	SH2	SH3	Kor1	Kor2	Kor3	RMSD1
Maddeler-Arası Boyutluluk Modeli	Fisher	0,20	50,00	0,75	0,81	0,79	0,439	0,386	0,397	0,563	0,588	0,598	1,126	1,124	1,104
		0,25	49,99	0,74	0,81	0,79	0,437	0,383	0,397	0,909	0,926	0,920	0,397	0,360	0,369
		<b>0,30</b>	<b>48,69</b>	<b>0,73</b>	<b>0,80</b>	<b>0,79</b>	<b>0,438</b>	<b>0,382</b>	<b>0,396</b>	<b>0,922</b>	<b>0,926</b>	<b>0,928</b>	<b>0,370</b>	<b>0,361</b>	<b>0,352</b>
Madde-İçi Boyutluluk Modeli	Bayesyen	0,20	50,00	0,75	0,82	0,80	0,437	0,383	0,396	0,921	0,929	0,924	0,370	0,352	0,362
		0,25	50,00	0,74	0,81	0,80	0,435	0,382	0,393	0,920	0,933	0,921	0,376	0,344	0,371
		<b>0,30</b>	<b>48,79</b>	<b>0,73</b>	<b>0,80</b>	<b>0,78</b>	<b>0,439</b>	<b>0,386</b>	<b>0,399</b>	<b>0,907</b>	<b>0,921</b>	<b>0,915</b>	<b>0,403</b>	<b>0,373</b>	<b>0,382</b>
Madde-İçi Boyutluluk Modeli	Fisher	0,20	50,00	0,85	0,77	0,74	0,355	0,424	0,444	0,947	0,923	0,922	0,334	0,388	0,380
		<b>0,25</b>	<b>50,00</b>	<b>0,85</b>	<b>0,78</b>	<b>0,75</b>	<b>0,355</b>	<b>0,426</b>	<b>0,445</b>	<b>0,944</b>	<b>0,918</b>	<b>0,920</b>	<b>0,341</b>	<b>0,396</b>	<b>0,384</b>
		0,30	48,08	0,84	0,77	0,74	0,364	0,431	0,451	0,434	0,428	0,428	1,871	1,820	1,798
Madde-İçi Boyutluluk Modeli	Bayesyen	0,20	50,00	0,86	0,79	0,75	0,359	0,427	0,447	0,542	0,508	0,541	1,430	1,430	1,380
		<b>0,25</b>	<b>49,98</b>	<b>0,85</b>	<b>0,78</b>	<b>0,75</b>	<b>0,358</b>	<b>0,426</b>	<b>0,446</b>	<b>0,946</b>	<b>0,931</b>	<b>0,922</b>	<b>0,334</b>	<b>0,367</b>	<b>0,378</b>
		0,30	47,79	0,85	0,78	0,75	0,362	0,431	0,450	0,512	0,510	0,512	1,489	1,437	1,409

Tablo 4.6’da seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher’in puanlama yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,75 ile 0,81 aralığında, ölçmenin standart hatasının 0,39 ile 0,44 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,56 ile 0,60 aralığında ve RMSD değerlerinin ise 1,10 ile 1,13 aralığında değiştiği görülmektedir. Standart hata durdurma kuralının 0,20 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise, İngilizce yeterlik sınavının kağıt-kalem testi formatındaki ortalama madde sayısı göz önünde bulundurularak çok-boyutlu BOB testi analizlerinde testteki maksimum madde sayısı 50 ile sınırlandırılmıştır. Standart hata durdurma kuralı 0,20’den 0,25’e çıkarıldığında ise testteki ortalama madde sayısı değişmediğinden, boyutlara ilişkin güvenilirlik katsayılarının ve standart hataların değişmediği görülmektedir. Buna karşın, boyutlara ilişkin RMSD değerinin azaldığı ve boyutlara ilişkin gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonun arttığı görülmektedir. Standart hata durdurma kuralının 0,30 olduğu durumda her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en yüksek ve korelasyon değerlerinin ise en düşük değere sahip olduğu görülmektedir.

Maddeler-arası boyutluluk modeline ait madde seçim yöntemlerinden seçkisiz madde seçim yöntemi, yetenek kestirim yöntemlerinden Fisher’in puanlama yöntemi ve standart hata durdurma kuralının 0,30 olduğu koşula ait testteki ortalama madde sayısının 48,68 olup diğer koşullarla karşılaştırıldığında daha güvenilir ve tutarlı sonuçlar verdiği söylenebilir. Ancak, her bir koşul için uygulanabilecek maksimum madde sayısı 50 ile sınırlandırıldığından, her bir koşul için elde edilen güvenilirlik katsayılarının birbirine yakın ve düşük olduğu görülmektedir. Dolayısıyla analiz bulgularının güvenilir ve tutarlı sonuçlar vermediği yorumu yapılabilir.

Tablo 4.6’da seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,75 ile 0,82 aralığında, ölçmenin standart hatasının 0,38 ile 0,44 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,92 ile 0,93 aralığında ve RMSD değerlerinin ise 0,35 ile 0,37 aralığında değiştiği görülmektedir. Standart hata



durdurma kuralının 0,20 olduğu durumda boyutlara ilişkin hata varyansları 0,20'nin altına düşmediğinden, bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı 50 olduğunda analiz durdurulmuştur. Standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında ise birinci ve ikinci boyuta ilişkin güvenilirlik katsayılarının 0,01 düzeyinde azaldığı ve üçüncü boyuta ilişkin güvenilirlik katsayısının değişmediği görülmektedir. Bunun temel sebebi, standart hata durdurma kuralı 0,20'den 0,25'e çıkarılmasına karşın testteki ortalama madde sayısı değişmemesidir. Ayrıca, standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında, boyutlara ilişkin ölçmenin standart hatası, RMSD değerleri ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinde sadece 0,001 düzeyinde bir değişim olduğundan bu katsayılarında anlamlı bir değişim gözlenmemiştir. Standart hata durdurma kuralının 0,30 olduğu durumda testteki ortalama madde sayısı 48,79 olurken, her bir boyuta ilişkin standart hataların ve RMSD değerlerinin en yüksek, güvenilirlik ve korelasyon değerlerinin ise en düşük değere sahip olduğu görülmektedir.

Seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası boyutluluk modeline ait BOB testi analiz bulgularına bakıldığında, genel olarak her bir standart hata durdurma kuralı için birinci boyuta (dinleme boyutuna) ilişkin güvenilirlik katsayılarının en düşük ve ikinci boyuta (dilbilgisi boyutuna) ilişkin güvenilirlik katsayısının ise en yüksek değere sahip olduğu görülmektedir. Ayrıca seçkisiz madde seçim yöntemi kullanıldığında hem Bayesyen MAP hem de Fisher'in puanlama yetenek kestirim yöntemlerine ait her bir boyuta ilişkin güvenilirlik birbirine yakın ve düşük olduğu görülmektedir. Dolayısıyla, hem Fisher'in puanlama yöntemi hem de Bayesyen MAP yetenek kestirim yönteminin kullanıldığı maddeler-arası çok-boyutlu BOB testlerine ilişkin kestirilen yetenek parametrelerine ait hata varyanslarının azalması testteki ortalama madde sayısını ve her boyuta ilişkin güvenilirlik katsayılarını etkilemediği görülmektedir.

Yetenek kestirim yöntemlerine ait her bir koşula ilişkin RMSD ile gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerleri karşılaştırıldığında, Bayesyen yetenek kestirim yöntemine ait RMSD değerlerinin daha düşük ve korelasyon değerlerinin ise daha yüksek olduğu görülmektedir. Dolayısıyla seçkisiz madde seçim yönteminin kullanıldığı çok-boyutlu BOB testlerinde Bayesyen MAP yetenek kestirim yöntemine ait güvenilir katsayıları düşük olmasına rağmen daha tutarlı sonuçlar verdiği söylenebilir.

**d) Madde-içi boyutluluk modeline ait farklı standart hata durdurma kuralı ve yetenek kestirim yöntemleri altında her bir koşula ilişkin güvenilirlik ve hata istatistikleri nasıldır?**

Tablo 4.6'da yer alan seçkisiz madde seçim yönteminin kullanıldığı madde-içi boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,74 ile 0,85 aralığında, ölçmenin standart hatasının 0,36 ile 0,44 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,92 ile 0,95 aralığında ve RMSD değerlerinin ise 0,33 ile 0,39 aralığında değiştiği görülmektedir. Standart hata durdurma kuralının 0,20 olduğu durumda bireylerin test sürecinde cevaplamış olduğu ortalama madde sayısı ise 50'dir. Standart hata durdurma kuralı 0,20'den 0,25'e çıkarıldığında ise testteki ortalama madde sayısı 50 olup değişmediğinden, birinci boyuta ilişkin güvenilirlik katsayısının değişmediği ve diğer boyutlara ilişkin güvenilirlik katsayılarının 0,01 düzeyinde arttığı görülmektedir. Buna karşın, boyutlara ilişkin RMSD değerlerinin azaldığı ve korelasyon değerlerinin ise arttığı gözlenmektedir. Standart hata durdurma kuralının 0,30 olduğunda durumda her bir boyuta ilişkin güvenilirlik katsayılarında önemli bir değişim olmamasına rağmen, boyutlara ilişkin RMSD değerlerinin arttığı ve boyutlara ilişkin korelasyon değerlerinin ise azaldığı görülmektedir. Standart hata durdurma kuralının 0,30 olduğu durumda testteki ortalama madde sayısının ise 48 olduğu görülmektedir.

Madde-içi boyutluluk modeline ait seçkisiz madde seçim yöntemi ve hata varyansı durdurma kuralına ilişkin çok-boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi kullanıldığında, hata varyansı durdurma kuralı 0,30 olarak belirlendiği durumda bile boyutlara ilişkin güvenilirlik katsayılarının düşük, RMSD ve standart hata değerlerinin oldukça yüksek olduğu görülmektedir.

Tablo 4.6'da seçkisiz madde seçim yönteminin kullanıldığı madde-içi boyutluluk modeline ilişkin çok boyutlu BOB testi analiz bulgularına bakıldığında, yetenek kestirim yöntemlerinden Bayesyen MAP yöntemi ve durdurma kuralı olarak standart hatanın 0,20 olduğu durumda her bir boyutla ilişkin güvenilirlik katsayılarının 0,75 ile 0,86 aralığında, ölçmenin standart hatasının 0,43 ile 0,54 aralığında, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin 0,51 ile 0,54 aralığında ve

RMSD deęerlerinin ise 1,38 ile 1,44 aralıęında deęiřtięi grlmektedir. Standart hata durdurma kuralının 0,20 olduęu durumda bireylerin test srecinde cevaplamıř olduęu ortalama madde sayısı ise 50'dir. Standart hata durdurma kuralının 0,20 olduęu durumda, bireylerin test srecinde cevaplamıř olduęu madde sayısı 50'ye ulařtıęında boyutlara iliřkin hata varyansları 0,20'nin altına dřmemesine raęmen analiz durdurulmuřtur. Standart hata durdurma kuralı 0,20'den 0,25'e ıkarıldıęında ise boyutlara iliřkin gvenirlik katsayılarında nemli bir deęiřim olmadıęı grlmektedir. Bunun temel sebebi, standart hata durdurma kuralı 0,20'den 0,25'e ıkarılmasına karřın testteki ortalama madde sayısı deęiřmemesidir. Ayrıca, standart hata durdurma kuralı 0,20'den 0,25'e ıkarıldıęında, boyutlara iliřkin lmenin standart hatası, RMSD deęerleri ve gerek ve kestirilen yetenek parametreleri arasındaki korelasyon deęerlerinde anlamlı bir deęiřim gzlenmemiřtir. Standart hata durdurma kuralının 0,30 olduęunda durumda testteki ortalama madde sayısı 47,79 olurken, her bir boyuta iliřkin standart hataların ve RMSD deęerlerinin en yksek, gvenirlik ve korelasyon deęerlerinin ise en dřk deęere sahip olduęu grlmektedir.

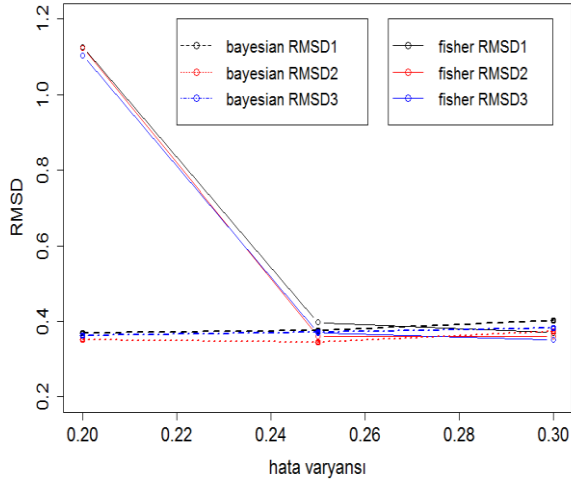
Sekisiz madde seim ynteminin kullanıldıęı maddeler-arası ve madde –ii boyutluluk modeline ait farklı ok-boyutlu BOB testi kořullarına iliřkin bulgular karřılařtırıldıęında, her iki model iin aynı kořullarda, farklı yetenek kestirim ynteminin kullanılması anlamlı bir farklılıęa neden olmadıęı yorumu yapılabilir. Bunun temel sebebi testteki ortalama madde sayısı 50'ye yaklařtıęı her iki yetenek kestirim ynteminin benzer sonulara sahip olduęu sylenebilir. Ayrıca, madde-ii boyutluluk modelinin her iki yetenek kestirim yntemi iin testteki ortalama madde sayısı aynı olmasına raęmen daha gvenilir ve tutarlı sonular verdięi yorumu yapılabilir.

řekil 4.6'da maddeler-arası ve madde-ii ok boyutlu telafi-edici modellere iliřkin sekisiz madde seim yntemi, hata varyansı durdurma kuralı ve yetenek kestirim yntemlerinden Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yntemlerinin kullanıldıęı ok-boyutlu BOB testi analizlerine ait grafikler verilmiřtir. řekil 4.6'daki kesikli izgiler Bayesyen yetenek kestirim yntemine ait bulguları temsil ederken, dz izgiler ise Fisher'in puanlama yetenek kestirim yntemine iliřkin bulguları temsil etmektedir. Ayrıca grafikte yer alan siyah renkli izgiler birinci boyuta (dinleme boyutuna) ait bulguları, kırmızı renkli izgi ikinci boyuta (dilbilgisi boyutuna) ait bulguları ve mavi renkli izgi ise nc boyuta (okuduęunu anlama boyutuna) ait bulguları temsil etmektedir.

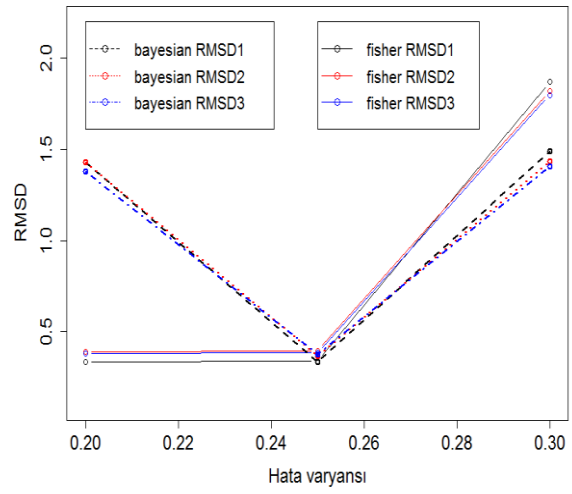
### Maddeler-arası Boyutluluk Modeli

### Madde-içi Boyutluluk Modeli

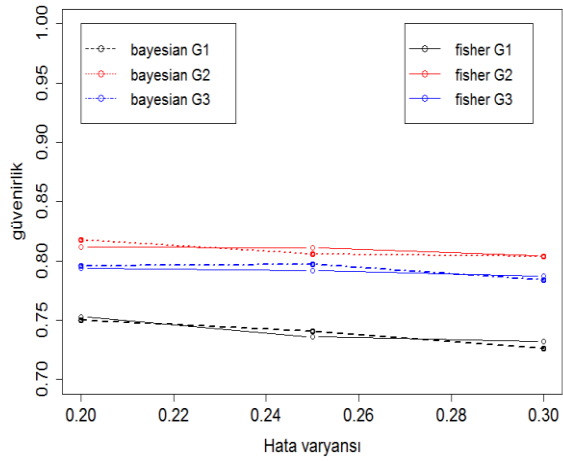
a.1 RMSD- Hata varyansı ilişkisi



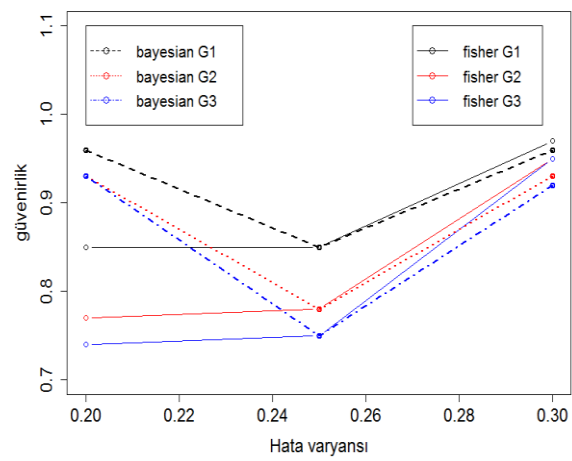
a.2 RMSD- Hata varyansı ilişkisi



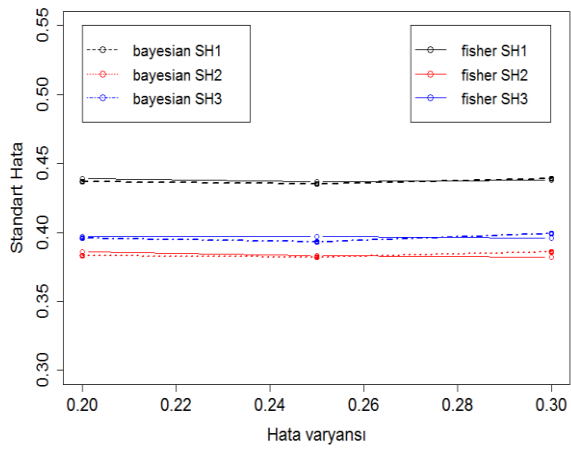
b.1 Güvenirlilik- Hata varyansı ilişkisi



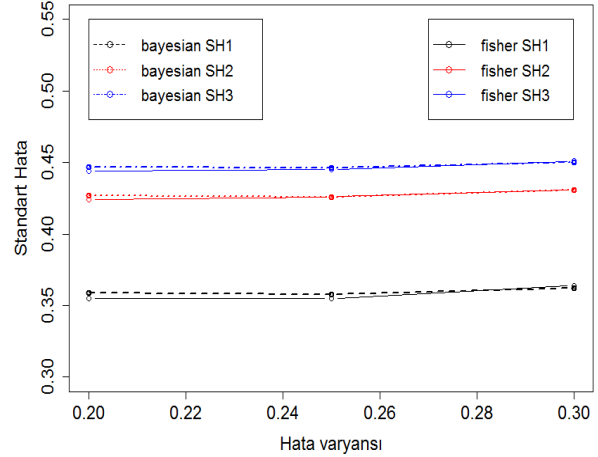
b.2 Güvenirlilik- Hata varyansı ilişkisi



c.1 Ölçmenin standart hatası- Hata varyansı



c.2 Ölçmenin standart hatası- Hata varyansı



Şekil 4.6: Seçkisiz (Random) Madde Seçim Yöntemi ve Hata Varyansı Durdurma Kuralına İlişkin Grafikler

Şekil 4.6'da yer alan a.1 ve a.2 grafikleri sırasıyla seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin RMSD değerleri ile yetenek parametrelerine ilişkin hata varyansları arasındaki ilişkiyi vermektedir.

Şekil 4.6 a.1'e bakıldığında maddeler-arası boyutluluk modelinde, yetenek parametrelerine ilişkin hata varyansı 0,20'den 0,25'e çıktığında Fisher'in puanlama yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin artması beklenirken azaldığı ve 0,30 olduğunda ise değişmediği görülmektedir. Ancak, Bayesyen MAP yetenek kestirim yöntemi kullanıldığı durumda hata varyansı arttıkça boyutlara ilişkin RMSD değerlerinde çok az bir artış olduğu görülmektedir. Dolayısıyla, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testlerinde, seçkisiz madde seçim yönteminin kullanıldığı durumda hem Bayesyen hem de Fisher'in puanlama yetenek kestirimi yöntemlerinin güvenilir ve tutarlı sonuçlar vermediği söylenebilir.

Şekil 4.6 a.2'ye bakıldığında, kestirilen yetenek parametrelerine ilişkin hata varyansı arttıkça Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin hata varyansı 0,20'den 0,25'e çıktığında artması beklenirken azaldığı, hata varyansı 0,30 olduğunda ise tekrar arttığı görülmektedir. Diğer taraftan, hata varyansı durdurma kuralı 0,20'den 0,25'e çıktığında Fisher'in puanlama yetenek kestirimine ait her bir boyuta ilişkin RMSD değerlerinin değişmediği, hata varyansı 0,30 olduğunda ise RMSD değerlerinin arttığı görülmektedir. Dolayısıyla, maddeler-arası boyutluluk modelinde olduğu gibi, madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testlerinde, seçkisiz madde seçim yönteminin kullanıldığı durumda hem Bayesyen hem de Fisher'in puanlama yetenek kestirimi yöntemlerinin güvenilir ve tutarlı sonuçlar vermediği söylenebilir. Bunun temel sebebi ise seçkisiz madde seçim yönteminin kullanıldığı çok-boyutlu BOB testlerinde, her bir koşul için testteki madde sayısı 50'ye yakın olmasına rağmen hata varyansı durdurma kuralı koşullarının sağlanamadan testin sonlandırılmasıdır.

Şekil 4.6'da yer alan b.1 ve b.2 grafikleri sırasıyla seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası ve madde-içi boyutluluk modellerine dayalı Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine ait her bir boyuta ilişkin güvenilirlik katsayıları ile yetenek parametrelerine ait hata varyansları arasındaki ilişkiyi vermektedir. Şekil 4.6 b.1'e bakıldığında maddeler-arası boyutluluk modelinde,

yetenek parametrelerine ait hata varyansları arttıkça, hem Bayesyen MAP yetenek kestirim yönteminin hem de Fisher'in puanlama yönteminin kullanıldığı koşullarda boyutlara ilişkin güvenilirlik katsayılarının da anlamlı bir değişim olmadığı görülmektedir. Ayrıca her bir hata varyansı durdurma kuralı için Bayesyen yetenek kestirim yöntemine ait boyutlara ilişkin güvenilirlik katsayılarının Fisher'in puanlama yetenek kestirim yöntemine ait güvenilirlik katsayılarına eşit olduğu görülmektedir.

Şekil 4.6 b.2'ye bakıldığında, her bir boyuta ilişkin hata varyansı arttıkça hem Bayesyen MAP yetenek kestirim yöntemine hem de Fisher'in puanlama yöntemine ilişkin güvenilirlik katsayılarında anlamlı bir değişim olmadığı görülmektedir. Ayrıca, maddeler-arası boyutluluk modelinde olduğu gibi, madde-içi boyutluluk modelinde de her iki yetenek kestirim yöntemine ait güvenilirlik katsayılarına eşit olduğu görülmektedir. Ayrıca her iki yetenek kestirimi için seçkisiz madde seçim yönteminin kullanıldığı çok-boyutlu BOB testlerinde madde-içi boyutluluk modeline ilişkin güvenilirlik katsayılarının daha yüksek olduğu görülmektedir.

Şekil 4.6'da yer alan c.1 ve c.2 grafikleri sırasıyla seçkisiz madde seçim yönteminin kullanıldığı maddeler-arası ve madde-içi boyutluluk modellerine ilişkin Fisher'in puanlama ve Bayesyen MAP yetenek kestirim yöntemlerine dayalı çok-boyutlu BOB testlerine ait her bir boyuta ilişkin ölçmenin standart hatası ile hata varyansı arasındaki ilişkiyi vermektedir. Şekil 4.6 c.1'e bakıldığında maddeler-arası boyutluluk modelinde, yetenek parametrelerine ilişkin hata varyansı arttıkça, hem Bayesyen MAP yetenek kestirim yönteminin hem de Fisher'in puanlama yönteminin kullanıldığı koşullarda boyutlara ilişkin standart hata değerlerinde anlamlı bir değişim olmadığı görülmektedir. Ayrıca her bir hata varyansı durdurma kuralı için Bayesyen yetenek kestirim yöntemine ait boyutlara ilişkin standart hata değerlerinin Fisher'in puanlama yetenek kestirim yöntemine ait standart hata değerlerine eşit olduğu görülmektedir.

Şekil 4.6 c.2'ye bakıldığında, boyutlara ilişkin hata varyansı arttıkça hem Fisher'in puanlama hem de Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin ölçmenin standart hatasının değişmediği görülmektedir. Ayrıca madde-içi boyutluluk modelinde Fisher'in puanlama ve Bayesyen MAP yetenek kestirimine ait her bir boyuta ilişkin ölçmenin standart hatası karşılaştırıldığında, boyutlara ilişkin standart hata değerlerinin birbirine yakın değerler aldığı görülmektedir. Bunun temel sebebi çok boyutlu BOB testi analizlerine ait her bir koşul için testteki ortalama madde sayısının ve güvenilirlik katsayılarının birbirine yakın olmasıdır.

Genel olarak, seçkisiz madde seçim yöntemine ait hata varyansı durdurma kuralının kullanıldığı madde-içi ve maddeler-arası çok boyutlu BOB testi analiz yöntemlerine ilişkin bulgular karşılaştırıldığında, her iki yetenek kestirim yöntemi için testteki ortalama madde sayısı 50 olmasına rağmen, boyutlara ilişkin güvenilirlik katsayılarının birbirine yakın ve düşük olduğu görülmektedir. Ayrıca, her iki yetenek kestirimi için madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi uygulamalarının daha güvenilir ve kararlı sonuçlar verdiği görülmektedir. Diğer taraftan, seçkisiz madde seçim yöntemi kullanıldığında Bayesyen yetenek kestirim yöntemlerinin madde düzeyindeki boyutluluk modellerinden etkilenmediği ve özellikle madde-içi boyutluluk modelinin kullanıldığı durumda daha güvenilir sonuçlar verdiği yorumu yapılabilir.

#### **Alt Problem 4: Çok boyutlu BOB testi yöntemi için belirlenen madde seçme yöntemlerinden hangisi daha iyi sonuç vermektedir?**

Madde seçim yöntemlerinden A-Optimality, D-Optimality ve seçkisiz (Random) yöntemlerine ait maddeler-arası boyutluluk modelinde dayalı çok-boyutlu BOB testi bulguları karşılaştırıldığında, A-Optimality ve D-Optimality madde seçim yöntemi için Bayesyen MAP yetenek kestirim yönteminin benzer sonuçlar verdiği, buna karşın seçkisiz madde seçim yöntemine ait güvenilirlik katsayılarının düşük ve hata istatistiklerinin ise yüksek olduğu görülmektedir. Yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi kullanıldığında ise, A-Optimality madde seçim yöntemine ait güvenilirlik ve korelasyon katsayılarının D-Optimality ve seçkisiz madde seçim yöntemlerine ait güvenilirlik ve korelasyon katsayılarından daha yüksek olduğu görülmektedir.

Farklı madde seçim yöntemlerinin kullanıldığı her bir koşula ilişkin RMSD ve standart hata değerleri karşılaştırıldığında, A-Optimality madde seçim yönteminin daha düşük RMSD ve standart hata değerine sahip olduğu görülmektedir. Ayrıca seçkisiz madde seçim yöntemine ait güvenilirlik katsayılarının her iki yetenek kestirim yöntemi için düşük olduğu bulgusuna ulaşılmıştır. Dolayısıyla, Bayesyen yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde, A-Optimality ve D-Optimality madde seçim yöntemleri benzer sonuç verirken, Fisher'in puanlama yetenek kestirim yöntemi kullanıldığında ise A-Optimality madde seçim yönteminin D-optimality ve seçkisiz madde seçim yöntemlerine göre daha iyi sonuç verdiği bulgusuna ulaşılmıştır.

Genel olarak, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi analiz sonuçları karşılaştırıldığında, aynı koşullarda A-Optimality madde seçim yöntemlerine

ait güvenilirlik ve kestirilen yetenek parametrelerine ilişkin korelasyon değerlerinin daha yüksek, RMSD ve standart hata değerlerinin ise daha düşük olduğu görülmektedir. Bu bulgular doğrultusunda, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemleri için en uygun madde seçim yönteminin A-Optimality madde seçim yöntemi olduğu söylenebilir.

Madde seçim yöntemlerinden A-Optimality, D-Optimality ve seçkisiz (Random) yöntemlerine ait madde-içi boyutluluk modelinde dayalı çok-boyutlu BOB testi bulguları karşılaştırıldığında, Bayesyen MAP yetenek kestirim yöntemi kullanıldığında A-Optimality ve D-Optimality madde seçim yöntemlerinin benzer sonuçlar verdiği görülmektedir. Yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi kullanıldığında ise, maddeler-arası boyutluluk modelinde olduğu gibi, A-Optimality madde seçim yöntemine ait güvenilirlik ve korelasyon katsayılarının D-Optimality ve seçkisiz madde seçim yöntemlerine ait güvenilirlik ve korelasyon katsayılarından daha yüksek olduğu görülmektedir. Ancak bu fark maddeler-arası boyutluluk modeli ile karşılaştırıldığında daha azdır.

Farklı madde seçim yöntemlerinin kullanıldığı her bir koşula ilişkin RMSD ve standart hata değerleri karşılaştırıldığında A-Optimality madde seçim yönteminin daha düşük RMSD ve standart hata değerine sahip olduğu görülmektedir. Ayrıca seçkisiz madde seçim yöntemine ait güvenilirlik katsayılarının diğer madde seçim yöntemlerine ait güvenilirlik katsayılarından daha düşük olduğu bulgusuna ulaşılmıştır. Dolayısıyla, hem maddeler-arası hem de madde-içi boyutluluk modeli için Bayesyen yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde A-Optimality ve D-Optimality madde seçim yöntemleri benzer sonuç verirken, Fisher'in puanlama yetenek kestirim yöntemi kullanıldığında ise A-Optimality madde seçim yönteminin D-optimality ve seçkisiz madde seçim yöntemlerine göre daha iyi sonuç verdiği bulgusuna ulaşılmıştır.

Genel olarak, her bir boyutluluk modeli için çok-boyutlu BOB testi analiz sonuçları karşılaştırıldığında, aynı koşullarda A-Optimality madde seçim yöntemlerine ait güvenilirlik ve kestirilen yetenek parametrelerine ilişkin korelasyon değerlerinin daha yüksek, RMSD ve standart hata değerlerinin ise daha düşük olduğu görülmektedir. Ayrıca, A-Optimality madde seçim yöntemi madde düzeyindeki boyutluluk modellerinden etkilenmezken, D-Optimality madde seçim yönteminin etkilendiği görülmektedir. Bu bulgular doğrultusunda, maddeler-arası ve madde-içi boyutluluk



modeline dayalı çok-boyutlu BOB testi yöntemleri için en uygun madde seçim yönteminin A-Optimality madde seçim yöntemi olduğu söylenebilir.

Hangi madde seçim yönteminin kullanılması gerektiğini testin amacı belirler. Örneğin, eğer testin ölçtüğü bütün boyutlar ölçülmek istendiğinde, A-optimality ve D-optimality madde seçim yöntemleri en iyi sonucu verir (Mulder ve van der Linden, 2009; Lin, 2012).

Diao (2009) yapmış olduğu çalışmada test sonlandırma kuralı olarak sabit madde sayısının, yetenek kestirim yöntemi olarak MLE yönteminin ve madde seçim yöntemlerinden A-optimality ve D-optimality yöntemlerinin kullandığında testteki madde sayısını 50 olduğunda A-optimality ve D-optimality madde seçim yöntemlerine ait RMSE ve ortalama yanlılık değerlerinin birbirine çok yakın olduğu ve her iki yöntemin benzer sonuçlar verdiği bulgusuna ulaşılmıştır. Nitekim bu çalışmada da testteki madde sayısı arttıkça her iki madde seçim yönteminin benzer sonuçlar verdiği bulgusuna ulaşılmıştır. Diğer taraftan, Diao ve Reckase (2009) Bayesyen MAP yetenek kestirim yöntemi kullanıldığında ise her iki madde seçim yönteminin testteki madde sayısı 20 ile sınırlandırıldığı durumda bile benzer sonuçlar verdiğini belirtmektedir.

Çok-boyutlu BOB testlerine ilişkin yapılan çalışmalara bakıldığında, A-optimality ve D-optimality madde seçim yöntemlerinin karşılaştırıldığı birçok çalışma yapılmıştır (Segall, (1996); Luecht, (1996); van der Linden,1999; Mulder ve van der Linden, 2009; Diao, 2009; Diao ve Reckase; 2009; Yoo, 2011; Lin, 2012). Genel olarak D-optimality yönteminin daha avantajlı olduğu ve daha yaygın olarak kullanıldığı belirtilmektedir (Atkinson ve Donev, 1992; Berger ve Veerkamp, 1996; Passo, 2007). Bunun temel sebebi olarak D-optimality madde seçim yönteminin daha güvenilir ve daha kararlı sonuçlar verdiği belirtilmektedir. Ancak bu yapılan çalışmalara bakıldığında sadece madde-içi boyutluluk modelinin kullanıldığı görülmektedir. Bu çalışmada ise madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testlerinde A-optimality ve D-optimality madde seçim yöntemlerinin benzer sonuçlar verdiği bulgusuna ulaşılmıştır. Maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testlerine ilişkin bulgulara bakıldığında ise hem Fisher'in puanlama hem de Bayesyen MAP yöntemi için A-optimality madde seçim yönteminin daha iyi sonuçlar verdiği görülmektedir. Bunun temel sebebi her iki yöntemde faktör bilgi matrisinin determinantına bağlı olarak madde seçilmesine karşın, A-optimality madde seçim yönteminde bir sonraki madde seçilirken

diğer yetenek parametrelerine ait varyansın da dikkate alınmasıdır (Mulder ve van der Linden, 2009). Bun karşın D-optimality madde seçim yöntemi, birden fazla boyutu ölçen maddelerden ziyade sadece bir boyuta hassas olan maddeleri seçme eğilimindedir.

**Alt Problem 5: Çok boyutlu BOB testi yöntemi için belirlenen yetenek kestirim yöntemlerinden hangisi daha iyi sonuç vermektedir?**

Bayesyen MAP ve MLE yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemlerine ait maddeler-arası boyutluluk modelinde dayalı çok-boyutlu BOB testi bulguları karşılaştırıldığında, Bayesyen MAP yetenek kestirim yönteminin A-Optimality ve D-Optimality madde seçim yöntemi için benzer sonuçlar verdiği, buna karşın seçkisiz madde seçim yöntemine ait güvenilirlik katsayılarının ve hata istatistiklerinin düşük olduğu görülmektedir. Fisher'in puanlama yöntemi kullanıldığında ise, A-Optimality madde seçim yöntemine ait güvenilirlik ve korelasyon katsayılarının D-Optimality ve seçkisiz madde seçim yöntemlerine ait güvenilirlik ve korelasyon katsayılarından daha yüksek olduğu görülmektedir. Özellikle hata varyansı durdurma kuralı kullanıldığında, her bir madde seçim yöntemi için Bayesyen MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir sonuçlar verdiği görülmektedir. Benzer şekilde, farklı yetenek kestirim yöntemlerinin kullanıldığı her bir koşula ilişkin RMSD ve standart hata değerleri karşılaştırıldığında, aynı koşullarda Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok boyutlu BOB testlerine ilişkin RMSD ve standart hata değerlerinin daha düşük olduğu görülmektedir. Dolayısıyla, A-Optimality madde seçim yöntemi ve sabit madde sayısı durdurma kuralının kullanıldığı çok-boyutlu BOB testlerinde, Fisher'in puanlama ve Bayesyen yetenek kestirim yöntemi benzer sonuçlar verdiği görülmektedir. D-Optimality madde seçim yöntemi kullanıldığında ise Bayesyen MAP yetenek kestirim yönteminin daha yüksek güvenilirlik katsayılarına ve daha düşük RMSD ve standart hata değerlerine sahip olduğu görülmektedir. Standart hata durdurma kuralı kullanıldığında ise her bir madde seçim yöntemi için Fisher'in puanlama yönteminin yetenek parametrelerine ait hata varyansını belirlenen hata varyansı koşulunun altına düşürmek için daha fazla maddeye ihtiyaç duyduğu görülmektedir.

Genel olarak, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi analiz sonuçları karşılaştırıldığında, aynı madde seçim yöntemi ve durdurma kuralı koşulları altında Bayesyen MAP yöntemine ait güvenilirlik ve kestirilen yetenek parametrelerine

ilişkin korelasyon değerlerinin daha yüksek, RMSD ve standart hata değerlerinin ise daha düşük olduğu görülmektedir. Dolayısıyla her bir madde seçim ve durdurma kuralı için maddeler-arası boyutluluk modelinde Bayesyen MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir sonuçlar verdiği yorumu yapılabilir. Bu bulgular doğrultusunda, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemlerinde her bir madde seçim yöntemi için en uygun yetenek kestirim yönteminin Bayesyen MAP yetenek kestirim yöntemi olduğu söylenebilir.

Yetenek kestirim yöntemlerinden Bayesyen MAP ve MLE yöntemlerinden Fisher'in puanlama yöntemlerine ait madde-içi boyutluluk modelinde dayalı çok-boyutlu BOB testi bulguları karşılaştırıldığında, Bayesyen MAP yetenek kestirim yönteminin A-Optimality ve D-Optimality madde seçim yöntemi için benzer sonuçlar verdiği görülmektedir. Özellikle hata varyansı durdurma kuralı kullanıldığında, her bir madde seçim yöntemi için Bayesyen MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir sonuçlar verdiği görülmektedir. Benzer şekilde, farklı yetenek kestirim yöntemlerinin kullanıldığı her bir koşula ilişkin RMSD ve standart hata değerleri karşılaştırıldığında, aynı koşullarda Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok boyutlu BOB testlerine ilişkin RMSD ve standart hata değerlerinin daha düşük olduğu görülmektedir. Dolayısıyla, aynı madde seçim yöntemi ve sabit madde sayısı durdurma kuralının kullanıldığı çok-boyutlu BOB testlerinde, Bayesyen MAP yetenek kestirim yönteminin daha yüksek güvenilirlik katsayılarına ve daha düşük RMSD ve standart hata değerlerine sahip olduğu görülmektedir. Standart hata durdurma kuralı kullanıldığında ise her bir madde seçim yöntemi için MLE yöntemlerinden Fisher'in puanlama yönteminin yetenek parametrelerine ait hata varyansını belirlenen hata varyansı koşulunun altına düşürmek için daha fazla maddeye ihtiyaç duyduğu görülmektedir.

Genel olarak, her bir boyutluluk modeli için çok-boyutlu BOB testi analiz sonuçları karşılaştırıldığında, aynı koşullarda Bayesyen MAP yetenek kestirim yöntemlerine ait güvenilirlik ve kestirilen yetenek parametrelerine ilişkin korelasyon değerlerinin daha yüksek, RMSD ve standart hata değerlerinin ise daha düşük olduğu görülmektedir. Ayrıca, Bayesyen MAP yetenek kestirim yönteminin madde seçim yöntemlerinden ve madde düzeyindeki boyutluluk modellerinden daha az etkilendiği yorumu yapılabilir.

Diğer taraftan, Fisher'in puanlama yöntemi A-Optimality madde seçim yöntemi kullanıldığında madde düzeyindeki boyutluluk modellerinden etkilenmezken, diğer

koşullarda hem madde seçim yönteminden hem de boyutluluk modellerinden etkilenmektedir. Bu bulgular doğrultusunda, maddeler-arası ve madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemleri için en uygun yetenek kestirim yönteminin Bayesyen MAP yöntemi olduğu söylenebilir. Bayesyen yetenek kestirim yöntemlerini MLE yetenek kestirim yöntemlerine göre üstün kılan en önemli özelliği, Segall (1996)'in de belirttiği gibi, Bayesyen yöntemlerin boyutlar arasındaki korelasyon ve yetenek parametrelerine ait önsel dağılım bilgisini kullanarak kestirim yapmasıdır (Diao ve Reckase, 2009). Bundan dolayı Bayesyen yöntemleri gerçek  $\theta$  değerine daha çabuk yakınsar ve daha az madde ile daha güvenilirlik ve kararlı kestirimler yapar. Buna karşın, kestirilen yetenek parametreleri hakkında yeterli bilgi olmadığı veya belirlenen önsel parametreler zayıf olduğu durumda, Bayesyen yöntemler ile kestirilen yetenek parametreleri yanlılık gösterir.

Diao (2009) çok-boyutlu BOB testi yöntemleri için testteki madde sayısının az olduğu ve boyutlar arasındaki korelasyonun yüksek olduğu durumlarda Bayesyen MAP yetenek kestirim yönteminin kullanılmasını tavsiye etmektedir. Testteki madde sayısı yeterli düzeyde olduğunda (50 veya daha fazla) MLE ve Bayesyen MAP yöntemlerinin benzer sonuçlar vereceğini belirtmiştir. Bu durumda Bayesyen MAP yetenek kestirim yöntemleri yerine MLE yetenek kestirim yöntemlerinin kullanılması tavsiye edilmektedir. Çünkü testteki madde sayısı yeterli olduğunda MLE yöntemlerinin önsel bilgiye ihtiyaç duymadan kestirim yapar ve kestirilen yetenek parametrelerinin yanlı olma ihtimali Bayesyen yöntemlere göre daha azdır. Reckase (2009) çok-boyutlu BOB testlerinde belirlenen farklı önsel dağılımların Bayesyen yetenek kestirim yöntemleri üzerindeki etkisinin incelenmesi gerektiğini vurgulamaktadır.

**Alt Problem 6: Çok boyutlu BOB testi yöntemi için belirlenen test sonlandırma kurallarından telafi-edici hata varyansı ile sabit soru sayısından hangisi daha iyi sonuç vermektedir?**

Analiz bulguları doğrultusunda maddeler-arası ve madde-içi boyutluluk modellerine dayalı çok-boyutlu BOB testi bulguları incelendiğinde en uygun madde seçim yönteminin A-optimality madde seçim yöntemi olduğu ve en uygun yetenek kestirim yönteminin ise Bayesyen MAP yetenek kestirim yöntemi olduğu bulgusuna ulaşılmıştır. Bundan dolayı bu bölümde, madde düzeyinde boyutluluk modellerine dayalı A-optimality madde seçim yöntemi ve Bayesyen MAP yetenek kestirim yönteminin

kullanıldığı çok-boyutlu BOB testlerine ait sabit madde sayısı ve hata varyansı durdurma kurallarına ilişkin bulgular karşılaştırılmıştır.

Maddeler-arası boyutluluk modeline dayalı A-optimality madde seçim yöntemi ve Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde sabit madde sayısı durdurma kuralı kullanıldığında testi sonlandırmak için en uygun madde sayısının 40 olduğu bulgusuna ulaşılmıştır. Bunun temel sebebi, testteki ortalama madde sayısı 40 olduğunda güvenilirlik katsayılarında önemli bir değişim olmamasına rağmen, RMSD değerlerinin anlamlı bir düzeyde arttığı ve gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerlerinin düştüğü görülmektedir. Buna karşın, madde-içi boyutluluk modeline dayalı A-optimality madde seçim yöntemi ve Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde sabit madde sayısı durdurma kuralı kullanıldığında testi sonlandırmak için en uygun madde sayısının 30 olduğu bulgusuna ulaşılmıştır.

Hata varyansı durdurma kuralına ilişkin maddeler-arası ve madde-içi boyutluluk modeline dayalı A-optimality madde seçim yöntemi ve Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde en uygun hata varyansı durdurma kuralının 0,25 olduğu görülmektedir. Bu durumda maddeler-arası boyutluluk modeli için testi alan her bir bireyin test sürecinde cevaplamış olduğu ortalama madde sayısı 22,8'e eşit olurken madde-içi boyutluluk modelinde testteki ortalama madde sayısı 26'ya eşit olduğu bulgusuna ulaşılmıştır.

Maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testlerinde durdurma kurallarından sabit madde sayısının 40 olduğu ve hata varyansı durdurma kuralının 0,25 olduğu koşullara ilişkin analiz bulguları karşılaştırıldığında, boyutlara ilişkin güvenilirlik ve korelasyon katsayıları arasında 0,01 düzeyinde bir farklılık olduğu görülmektedir. Ayrıca çok-boyutlu BOB testlerine ilişkin RMSD ve standart hata değerleri karşılaştırıldığında ise hata varyansı durdurma kuralına ilişkin RMSD ve standart hata değerlerinin daha yüksek olduğu görülmektedir. Bu bulgular doğrultusunda, testteki ortalama madde sayısını 40'tan 22,8'e düşmesine rağmen boyutlara ilişkin güvenilirlik ve korelasyon katsayıları birbirine yakın değerler almasından dolayı hata varyansının 0,25'e eşitlendiği durdurma kuralının maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testleri için en uygun durdurma kuralı olduğu söylenebilir.

Madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testlerinde durdurma kurallarından sabit madde sayısının 30 olduğu ve hata varyansı durdurma kuralının 0,25 olduğu koşullara ilişkin analiz bulguları karşılaştırıldığında, boyutlara ilişkin güvenilirlik ve korelasyon katsayıları arasında 0,01 düzeyinde bir farklılık olduğu görülmektedir. Ayrıca çok-boyutlu BOB testlerine ilişkin RMSD ve standart hata değerleri karşılaştırıldığında ise hata varyansı durdurma kuralına ilişkin RMSD ve standart hata değerlerinin daha düşük olduğu görülmektedir. Bu bulgular doğrultusunda, testteki ortalama madde sayısını 30'dan 26'ya düşmesine rağmen boyutlara ilişkin güvenilirlik ve korelasyon katsayıları birbirine yakın değerler alması ve hata istatistiklerinin daha düşük olmasından dolayı hata varyansının 0,25'e eşitlendiğin durdurma kuralının madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testleri için en uygun durdurma kuralı olduğu söylenebilir.

Bu çalışmada, madde seçim yöntemlerinden A-optimality ve yetenek kestirim yöntemlerinden Bayesyen MAP yöntemlerinin kullanıldığı maddeler-arası ve madde-içi boyutluluk modellerine dayalı çok-boyutlu BOB testleri için en uygun test sonlandırma kuralının hata varyansının 0,25'e eşitlendiği durum olmasına karşın gerçek uygulamalarda göz önünde tutulması gereken bazı durumlar vardır. Hata varyansı durdurma kuralı daha güvenilir ve etkili olmasına karşın bireylerin test sürecinde farklı sayıda maddelere cevap vermesi testin adil olmadığı algısını oluşturabilir (Gershon, 2005). Bu yüzden eğitimdeki gerçek BOB testi uygulamalarında genellikle sabit madde sayısı durdurma kuralı kullanılır (Yoo, 2011). Sabit madde sayısı durdurma kuralının hata varyansı durdurma kuralına tercih edilmesinin bir diğer nedeni ise test süresi ve testten sıkılma gibi koşulların her bir birey için standartlaştırılmasına olanak sağlamasıdır (Segall, 2004). Buna karşın, Rizavi (2001) sabit madde sayısı durdurma kuralının kullanıldığı durumlarda testteki madde sayısı bireyin yeteneğini güvenilir bir şekilde ölçmek için yeterli olmadığına problemlere neden olacağını ve testin güvenilirlik ve geçerliğini düşüreceğini savunmuştur.

Eğer çok-boyutlu BOB testlerinde hata varyansı durdurma kuralının her bir boyut için sağlanması belirlenmiş ise durdurma kuralı her bir boyut için aynı anda sağlanmadığı sürece test sonlandırılmaz. Bu durumda testteki madde sayısı testin boyut sayısı ve her bir madde havuzundaki maddelerin sahip olduğu bilgiye bağlı olarak beklenenden çok daha fazla olabilir (Yoo, 2011).

## **Alt Problem 7: Analiz bulguları doğrultusunda en uygun çok boyutlu BOB testi algoritması nasıldır?**

Analiz bulguları doğrultusunda maddeler-arası ve madde-içi boyutluluk modellerine dayalı çok-boyutlu BOB testi analiz bulguları incelendiğinde en uygun madde seçim yönteminin A-optimality madde seçim yöntemi olduğu bulgusuna ulaşılmıştır. Seçkisiz madde seçim yöntemine ilişkin analiz bulgularına bakıldığında özellikle hata varyansı durdurma kuralı kullanıldığında yüksek güvenilirlik katsayıları ve düşük hata istatistiklerini elde etmek için testteki ortalama madde sayısının 50'ye yakın olması gerekir. Diğer taraftan, D-optimality madde seçim yönteminin kullanıldığı çok-boyutlu BOB testlerine ilişkin analiz bulguları incelendiğinde, yetenek kestirim yöntemlerinden Bayesyen MAP yetenek kestirim yönteminin kullanıldığı durumda A-optimality madde seçim yöntemi ile benzer sonuçlar vermesine rağmen, Fisher'in puanlama yöntemi kullanıldığında aynı koşullarda güvenilirlik katsayılarının arttığı ve hata istatistiklerinin düştüğü görülmektedir. Dolayısıyla, her iki çok boyutlu model ve yetenek kestirim yöntemi için en uygun madde seçim yöntemi A-optimality yöntemidir.

Analiz bulguları doğrultusunda maddeler-arası ve madde-içi boyutluluk modellerine dayalı çok-boyutlu BOB testi analiz bulguları incelendiğinde en uygun yetenek kestirim yönteminin ise Bayesyen MAP yetenek kestirim yöntemi olduğu bulgusuna ulaşılmıştır. Bunun temel sebebi, Fisher'in puanlama yetenek kestirim yönteminin hem madde düzeyindeki boyutluluk modellerinden hem de madde seçim yöntemlerinden etkilenmesidir.

Ayrıca, madde seçim yöntemlerinden A-optimality ve yetenek kestirim yöntemlerinden Bayesyen MAP yöntemlerinin kullanıldığı madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testine ilişkin RMSD ve standart hata değerlerinin maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testine ilişkin RMSD ve standart hata değerlerinden daha düşük olduğu bulgusuna ulaşılmıştır.

Çok-boyutlu BOB testlerinde kullanılan durdurma kurallarına ilişkin analiz bulguları karşılaştırıldığında, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testleri için en uygun durdurma kuralı hata varyansın 0,25 eşit olduğu koşul (testteki ortalama madde sayısı 22,8) ve sabit madde sayısı durdurma kuralı için ise en uygun koşulun testteki madde sayısının 40 ile sınırlandırıldığı koşulun olduğu bulgusuna ulaşılmıştır. Dolayısıyla, daha az madde ile daha güvenilir sonuçlar verdiği için, maddeler-arası boyutluluk modeli için en uygun durdurma kuralının hata varyansını

0,25'e eşitlendiği hata varyansı durdurma kuralıdır. Diğer taraftan, madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testleri için en uygun durdurma kuralı hata varyansın 0,25 eşit olduğu koşul (testteki ortalama madde sayısı 26) ve sabit madde sayısı durdurma kuralı için ise en uygun koşulun testteki madde sayısının 30 ile sınırlandırıldığı koşulun olduğu bulgusuna ulaşılmıştır. Ancak boyutlara ilişkin güvenilirlik ve korelasyon değerlerinde önemli bir değişim olmamasına rağmen daha az madde ile daha güvenilir sonuçlar verdiği için madde-içi boyutluluk modeli için en uygun durdurma kuralının hata varyansını 0,25'e eşitlendiği hata varyansı durdurma kuralıdır. Dolayısıyla hem madde-içi boyutluluk hem de maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testleri için en uygun durdurma kuralının hata varyansını 0,25'e eşitlendiği hata varyansı durdurma kuralı olduğu söylenebilir.

Sonuç olarak, madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB testleri için en uygun madde seçim yönteminin A-optimality yöntemi, en uygun yetenek kestirim yönteminin Bayesyen MAP yöntemi ve en uygun durdurma kuralının ise hata varyansını 0,25'e eşitlendiği hata varyansı durdurma kuralı olduğu söylenebilir.



## 5. SONUÇ ve ÖNERİLER

Bu bölümde araştırmanın bulgu ve yorumlarına dayalı olarak ulaşılan sonuçların özetine ve bu sonuçlardan yola çıkarak geliştirilen önerilere yer verilmiştir.

### 5.1. Sonuçlar

Bu bölümde farklı madde seçim yöntemi, yetenek kestirim yöntemi ve durdurma kuralının kullanıldığı çok boyutlu MTK modellerinden madde-içi ve maddeler-arası boyutluluk modellerine dayalı çok-boyutlu BOB testi analiz sonuçlarına yer verilmiştir.

Maddeler-arası boyutluluk modelinde dayalı çok-boyutlu BOB testi analizlerinde kullanılan madde seçim yöntemlerinden A-optimality, D-optimality ve seçkisiz (Random) madde seçim yöntemleri karşılaştırıldığında, A-optimality ve D-optimality madde seçim yöntemi için Bayesyen MAP yetenek kestirim yönteminin benzer sonuçlar verdiği görülmektedir, Buna karşın seçkisiz madde seçim yöntemine ait güvenilirlik katsayılarının düşük ve hata istatistiklerinin yüksek olduğu söylenebilir. Yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi kullanıldığında ise, A-optimality madde seçim yöntemine ait güvenilirlik ve korelasyon katsayılarının D-optimality ve seçkisiz madde seçim yöntemlerine ait güvenilirlik ve korelasyon katsayılarından daha yüksek olduğu görülmektedir.

Bayesyen yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde, A-optimality ve D-optimality madde seçim yöntemleri benzer sonuç verirken, Fisher'in puanlama yetenek kestirim yöntemi kullanıldığında ise A-optimality madde seçim yönteminin D-optimality ve seçkisiz madde seçim yöntemlerine göre daha iyi sonuç verdiği bulgusuna ulaşılmıştır.

Genel olarak, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi analiz sonuçları karşılaştırıldığında, aynı koşullarda A-optimality madde seçim yöntemlerine ait güvenilirlik ve kestirilen yetenek parametrelerine ilişkin korelasyon değerlerinin daha yüksek, RMSD ve standart hata değerlerinin ise daha düşük olduğundan, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemleri için en uygun madde seçim yönteminin yetenek kestirim yöntemlerinden etkilenmeyen A-optimality madde seçim yöntemi olduğu söylenebilir.

Madde-içi boyutluluk modelinde dayalı çok-boyutlu BOB testi analizlerinde kullanılan madde seçim yöntemleri karşılaştırıldığında, Bayesyen MAP yetenek kestirim yöntemi

kullanıldığında A-optimality ve D-optimality madde seçim yöntemlerinin benzer sonuçlar verdiği görülmektedir. Yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemi kullanıldığında ise, maddeler-arası boyutluluk modelinde olduğu gibi, A-optimality madde seçim yöntemine ait güvenilirlik ve korelasyon katsayılarının D-optimality ve seçkisiz madde seçim yöntemlerine ait güvenilirlik ve korelasyon katsayılarından daha yüksek olduğu görülmektedir. Ancak bu fark maddeler-arası boyutluluk modeli ile karşılaştırıldığında daha azdır.

Farklı madde seçim yöntemlerinin kullanıldığı her bir koşula ilişkin RMSD ve standart hata değerleri karşılaştırıldığında A-optimality madde seçim yönteminin daha düşük RMSD ve standart hata değerine sahip olduğu görülmektedir. Dolayısıyla, hem maddeler-arası hem de madde-içi boyutluluk modeli için Bayesyen yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde A-optimality ve D-optimality madde seçim yöntemleri benzer sonuç verirken, Fisher'in puanlama yetenek kestirim yöntemi kullanıldığında ise A-optimality madde seçim yönteminin D-optimality ve seçkisiz madde seçim yöntemlerine göre daha iyi sonuç verdiği bulgusuna ulaşılmıştır.

Genel olarak, her iki boyutluluk modeli için çok-boyutlu BOB testi analiz sonuçları karşılaştırıldığında, A-optimality madde seçim yöntemi madde düzeyindeki boyutluluk modellerinden etkilenmezken, D-optimality madde seçim yönteminin etkilendiği görülmektedir. Sonuç olarak, maddeler-arası ve madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemleri için en uygun madde seçim yönteminin A-optimality madde seçim yöntemi olduğu söylenebilir.

Maddeler-arası boyutluluk modelinde dayalı çok-boyutlu BOB testi analizlerinde kullanılan Bayesyen MAP ve MLE yetenek kestirim yöntemlerinden Fisher'in puanlama yöntemleri karşılaştırıldığında, Bayesyen MAP yetenek kestirim yönteminin A-optimality ve D-optimality madde seçim yöntemi için benzer sonuçlar verdiği, buna karşın seçkisiz madde seçim yöntemine ait güvenilirlik katsayılarının daha düşük ve hata istatistiklerinin daha yüksek olduğu görülmektedir.

Fisher'in puanlama yöntemi kullanıldığında ise, A-optimality madde seçim yöntemine ait güvenilirlik ve korelasyon katsayılarının D-optimality ve seçkisiz madde seçim yöntemlerine ait güvenilirlik ve korelasyon katsayılarından daha yüksek olduğu görülmektedir. Özellikle hata varyansı durdurma kuralı kullanıldığında, her bir madde seçim yöntemi için Bayesyen MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir ve tutarlı sonuçlar verdiği görülmektedir.

Farklı yetenek kestirim yöntemlerinin kullanıldığı her bir koşula ilişkin RMSD ve standart hata değerleri karşılaştırıldığında, aynı koşullarda Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok boyutlu BOB testlerine ilişkin RMSD ve standart hata değerlerinin daha düşük olduğu görülmektedir. Ancak testteki madde sayısı 50'ye yaklaştığında her iki yetenek kestirim yönteminin benzer sonuçlar verdiği görülmektedir.

Genel olarak, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi analiz sonuçları karşılaştırıldığında, Bayesyen MAP yöntemine ait güvenilirlik ve kestirilen yetenek parametrelerine ilişkin korelasyon değerlerinin daha yüksek, RMSD ve standart hata değerlerinin ise daha düşük olduğu görülmektedir. Dolayısıyla her bir madde seçim ve durdurma kuralı için maddeler-arası boyutluluk modelinde Bayesyen MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir sonuçlar verdiği yorumu yapılabilir. Bu bulgular doğrultusunda, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemlerinde her bir madde seçim yöntemi için en uygun yetenek kestirim yönteminin Bayesyen MAP yetenek kestirim yöntemi olduğu söylenebilir.

Madde-içi boyutluluk modelinde dayalı çok-boyutlu BOB testi analizlerinde kullanılan Bayesyen MAP ve Fisher'in puanlama yöntemleri karşılaştırıldığında, Bayesyen MAP yetenek kestirim yönteminin A-optimality ve D-optimality madde seçim yöntemi için benzer sonuçlar verdiği görülmektedir. Buna karşın hata varyansı durdurma kuralı kullanıldığında güvenilirlik katsayıları birbirine yakın olmasına rağmen, A-optimality madde seçim yöntemine ait testteki ortalama madde sayısının D-optimality madde seçim yöntemine göre daha düşük olduğu görülmektedir. Fisher'in puanlama yöntemi kullanıldığında ise sabit madde sayısı durdurma kuralı için A-optimality madde seçim yöntemine ait güvenilirlik ve korelasyon katsayılarının D-optimality ve seçkisiz madde seçim yöntemlerine ait güvenilirlik ve korelasyon katsayılarından daha yüksek olduğu görülmektedir. Özellikle hata varyansı durdurma kuralı kullanıldığında, her bir madde seçim yöntemi için Bayesyen MAP yetenek kestirim yönteminin daha az madde ile daha güvenilir sonuçlar verdiği görülmektedir.

Madde-içi boyutluluk modeli için farklı yetenek kestirim yöntemlerinin kullanıldığı her bir koşula ilişkin RMSD ve standart hata değerleri karşılaştırıldığında, aynı koşullarda Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok boyutlu BOB testlerine ilişkin RMSD ve standart hata değerlerinin daha düşük olduğu görülmektedir.

Dolayısıyla, aynı madde seçim yöntemi ve sabit madde sayısı durdurma kuralının kullanıldığı çok-boyutlu BOB testlerinde, Bayesyen MAP yetenek kestirim yönteminin daha yüksek güvenilirlik katsayılarına ve daha düşük RMSD ve standart hata değerlerine sahip olduğu görülmektedir. Standart hata durdurma kuralı kullanıldığında ise her bir madde seçim yöntemi için MLE yöntemlerinden Fisher'in puanlama yönteminin yetenek parametrelerine ait hata varyansını belirlenen hata varyansı koşulunun altına düşürmek için daha fazla maddeye ihtiyaç duyduğu görülmektedir.

Genel olarak, her bir boyutluluk modeli için çok-boyutlu BOB testi analiz sonuçları karşılaştırıldığında, aynı koşullarda Bayesyen MAP yetenek kestirim yöntemlerine ait güvenilirlik ve kestirilen yetenek parametrelerine ilişkin korelasyon değerlerinin daha yüksek, RMSD ve standart hata değerlerinin ise daha düşük olduğu görülmektedir. Ayrıca, Bayesyen MAP yetenek kestirim yönteminin madde seçim yöntemlerinden ve madde düzeyindeki boyutluluk modellerinden daha az etkilendiği yorumu yapılabilir.

Fisher'in puanlama yöntemi A-optimality madde seçim yöntemi kullanıldığında madde düzeyindeki boyutluluk modellerinden etkilenmezken, diğer koşullarda hem madde seçim yönteminden hem de boyutluluk modellerinden etkilenmektedir. Sonuç olarak, maddeler-arası ve madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemleri için en uygun yetenek kestirim yönteminin Bayesyen MAP yöntemi olduğu söylenebilir. Dolayısıyla, çok-boyutlu BOB testi yöntemleri için testteki madde sayısının az olduğu ve boyutlar arasındaki korelasyonun yüksek olduğu durumlarda Bayesyen MAP yetenek kestirim yönteminin kullanılmasını tavsiye etmektedir.

Maddeler-arası ve madde-içi boyutluluk modellerine dayalı çok-boyutlu BOB testi analizleri sonucunda en uygun madde seçim yönteminin A-optimality madde seçim yöntemi olduğu ve en uygun yetenek kestirim yönteminin ise Bayesyen MAP yetenek kestirim yöntemi olduğu bulgusuna ulaşılmıştır.

Maddeler-arası boyutluluk modeline dayalı A-optimality madde seçim yöntemi ve Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde sabit madde sayısı durdurma kuralı kullanıldığında testi sonlandırmak için en uygun madde sayısının 40 olduğu bulgusuna ulaşılmıştır. Buna karşın, madde-içi boyutluluk modeline dayalı A-optimality madde seçim yöntemi ve Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde sabit madde sayısı durdurma kuralı kullanıldığında testi sonlandırmak için en uygun madde sayısının 30 olduğu bulgusuna ulaşılmıştır. Bunun temel sebebi, madde-içi boyutluluk modellerinde seçilen

maddenin her üç boyut hakkında bilgi içerebildiğinden daha az madde ile benzer güvenilirlikte sonuçlar verebilmektedir.

Hata varyansı durdurma kuralına ilişkin maddeler-arası ve madde-içi boyutluluk modeline dayalı A-optimality madde seçim yöntemi ve Bayesyen MAP yetenek kestirim yönteminin kullanıldığı çok-boyutlu BOB testlerinde en uygun hata varyansı durdurma kuralının 0,25 olduğu görülmektedir. Bu durumda maddeler-arası boyutluluk modeli için testi alan her bir bireyin test sürecinde cevaplamış olduğu ortalama madde sayısı 22,8'e eşit olurken madde-içi boyutluluk modelinde testteki ortalama madde sayısı 26'ya eşit olduğu bulgusuna ulaşılmıştır.

Maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testlerinde durdurma kurallarından sabit madde sayısının 40 olduğu ve hata varyansı durdurma kuralının 0,25 olduğu koşullara ilişkin analiz bulguları karşılaştırıldığında, boyutlara ilişkin güvenilirlik ve korelasyon katsayıları arasında 0,01 düzeyinde bir farklılık olduğu görülmektedir. Bu bulgular doğrultusunda, testteki ortalama madde sayısını 40'tan 22,8'e düşmesine rağmen boyutlara ilişkin güvenilirlik ve korelasyon katsayıları birbirine yakın değerler almasından dolayı hata varyansının 0,25'e eşitlendiği durdurma kuralının maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testleri için en uygun durdurma kuralı olduğu söylenebilir.

Madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testlerinde durdurma kurallarından sabit madde sayısının 30 olduğu ve hata varyansı durdurma kuralının 0,25 olduğu koşullara ilişkin analiz bulguları karşılaştırıldığında, boyutlara ilişkin güvenilirlik ve korelasyon katsayıları arasında 0,01 düzeyinde bir farklılık olduğu görülmektedir. Bu bulgular doğrultusunda, testteki ortalama madde sayısını 30'dan 26'ya düşmesine rağmen boyutlara ilişkin güvenilirlik ve korelasyon katsayıları birbirine yakın değerler alması ve hata istatistiklerinin daha düşük olmasından dolayı hata varyansının 0,25'e eşitlendiğın durdurma kuralının madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testleri için en uygun durdurma kuralı olduğu söylenebilir.

Burada kâğıt-kalem formatındaki İngilizce yeterlik sınavındaki (İYS) ortalama madde sayısı 65 olarak kabul edildiğinde, testin maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi formatında uygulandığında testteki ortalama madde sayısının % 65 oranında azaldığı görülmektedir. Benzer şekilde, İYS testi madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi formatında uygulandığında ise testteki ortalama madde sayısının % 60 oranında azaldığı görülmektedir. Sonuç olarak, kâğıt-kalem

testleri ile karşılaştırıldığında hem maddeler-arası hem de madde-içi boyutluluk modellerine dayalı çok-boyutlu BOB testlerinin daha az madde ile daha yüksek güvenilirlikte ölçümler yaptığı söylenebilir.

Sonuç olarak, madde düzeyinde boyutluluk modellerine dayalı çok-boyutlu BOB testleri için en uygun madde seçim yönteminin A-optimality yöntemi, en uygun yetenek kestirim yönteminin Bayesyen MAP yöntemi ve en uygun durdurma kuralının ise hata varyansının 0,25'e eşitlendiği hata varyansı durdurma kuralı olduğu söylenebilir.

## **5.2. Öneriler**

### **5.2.1. Araştırmaya Dönük Öneriler**

Maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi analiz sonuçları bakıldığında, aynı koşullarda A-optimality madde seçim yöntemlerine ait güvenilirlik ve kestirilen yetenek parametrelerine ilişkin korelasyon değerlerinin daha yüksek, RMSD ve standart hata değerlerinin ise daha düşük olduğundan, maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemleri için yetenek kestirim yöntemlerinden etkilenmeyen A-optimality madde seçim yönteminin kullanılması önerilmektedir.

Genel olarak, her iki boyutluluk modeli için çok-boyutlu BOB testi analiz sonuçları karşılaştırıldığında, A-optimality madde seçim yöntemi madde düzeyindeki boyutluluk modellerinden etkilenmezken, D-optimality madde seçim yönteminin etkilendiği görülmektedir. Dolayısıyla, maddeler-arası boyutluluk modelinde olduğu gibi, madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemleri için madde seçim yöntemlerinden A-optimality madde seçim yönteminin kullanılması önerilmektedir.

Madde-içi ve maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemlerinin kullanıldığı durumda madde seçim yöntemlerinden ve madde düzeyindeki boyutluluk modellerinden daha az etkilendiği için Bayesyen MAP yetenek kestirim yönteminin kullanılması önerilmektedir.

Fisher'in puanlama yöntemi A-optimality madde seçim yöntemi kullanıldığında madde düzeyindeki boyutluluk modellerinden etkilenmezken, diğer koşullarda hem madde seçim yönteminden hem de boyutluluk modellerinden etkilenmektedir. Sonuç olarak, maddeler-arası ve madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testi yöntemleri için en uygun yetenek kestirim yönteminin Bayesyen MAP yöntemi olduğu söylenebilir. Dolayısıyla, çok-boyutlu BOB testi yöntemleri için testteki madde sayısının

az olduđu ve boyutlar arasındaki korelasyonun yüksek olduđu durumlarda Bayesyen MAP yetenek kestirim yönteminin kullanılması önerilmektedir.

Testteki ortalama madde sayısını 40'tan 22,8'e düşmesine rağmen boyutlara ilişkin güvenilirlik ve korelasyon katsayıları birbirine yakın değerler almasından dolayı maddeler-arası boyutluluk modeline dayalı çok-boyutlu BOB testleri için sabit madde sayısı durdurma kuralı yerine hata varyansının 0,25'e eşitlendiđi durdurma kuralının kullanılması önerilmektedir. Benzer şekilde, testteki ortalama madde sayısını 30'dan 26'ya düşmesine rağmen boyutlara ilişkin güvenilirlik ve korelasyon katsayıları birbirine yakın değerler almasından dolayı madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testleri için sabit madde sayısı durdurma kuralı yerine hata varyansının 0,25'e eşitlendiđi durdurma kuralının kullanılması önerilmektedir.

Sonuç olarak, kâğıt-kalem testleri ile karşılaştırıldığında hem maddeler-arası hem de madde-içi boyutluluk modellerine dayalı çok-boyutlu BOB testlerinin daha az madde ile daha yüksek güvenilirlikte ölçümler yaptığından benzer formattaki testlerin çok-boyutlu bilgisayar ortamında bireyselleştirilmiş testler ile uygulanması önerilmektedir.

### **5.2.2. Uygulamaya Dönük Öneriler**

Maddeler-arası ve madde-içi boyutluluk modeline dayalı çok-boyutlu BOB testleri için alternatif test sonlandırma kuralı olarak her iki koşul birlikte kullanılabilir. Böylece, boyutlara ilişkin hata varyansı belirlenen hata varyansının altına düşmediđi durumda sabit madde sayısı durdurma kuralı kullanılarak testteki madde sayısı yüksek güvenilirlik ve düşük hata istatistiđi veren maksimum madde sayısı ile sınırlandırılabilir.

Bu çalışmada, çok-boyutlu BOB testlerinde kullanılan madde seçim yöntemlerinden Fisher'in bilgi matrisine dayalı A-optimality, D-optimality ve seçkisiz (random) madde seçim yöntemleri kullanılmıştır. Fisher'in bilgi matrisine dayalı madde seçim yöntemlerinin yanı sıra Kullback-Liebler uzaklık yöntemlerine dayalı farklı madde seçim yöntemleri ve Bayesyen madde seçim yöntemleri kullanılarak, sonuçlar karşılaştırılabilir.

Bu çalışmada, gerçek veriye dayalı simülasyon yöntemi (post-hoc simulation method) kullanılarak, testin üç boyutlu olduđu durumda farklı yetenek kestirimi, madde seçim ve durdurma kurallarına ilişkin çok-boyutlu BOB testi sonuçları karşılaştırılmıştır. Farklı simülasyon çalışmaları yapılarak, testin ölçtüđü boyut sayısının, boyutlara ilişkin

madde havuzu büyüklüğünün ve farklı çok boyutlu modellerin çok-boyutlu BOB testi yöntemleri üzerindeki etkisi incelenebilir.

Bu çalışmada, madde kullanım sıklığı kontrol (item exposure control) yöntemlerinden sadece randomesque yöntemi kullanılmıştır. Alan yazınına bakıldığında bireyselleştirilmiş testlerde kullanılmak üzere geliştirilmiş birçok madde kullanım sıklığı kontrol yöntemleri vardır (Chang ve Ansley, 2003; Georgiadou, Triantafillou ve Economides, 2007; Hetter ve Sympson, 1997; Pastor, Dodd ve Chang, 2002; Stocking ve Lewis, 1998, 2000; Way, 1998). Farklı madde kullanım sıklığı yöntemleri kullanılarak, bu yöntemlerin çok-boyutlu BOB testleri üzerindeki etkisi incelenebilir ve bu yöntemlerin performansları karşılaştırılabilir.

Bu çalışmada, çok-boyutlu BOB testi sürecinde her bir boyut için sorulacak madde sayısını ve testin formatını kontrol altında tutmak için kullanılan içerik ağırlıklandırması (content balance) yöntemleri kullanılmamıştır. Gelecekte yapılacak çalışmalarda, daha önce geliştirilmiş olan içerik ağırlıklandırması yöntemleri (örn. Düzgünleştirilmiş-alfa deseni, Chang, Qian ve Ying, 2001; Sympson ve Hetter's yöntemi, Sympson ve Hetter 1985) kullanılarak BOB testi süreci daha gerçekçi hale getirilebilir. Ayrıca, çok-boyutlu BOB testlerinde kullanılan madde kullanım sıklığı ve içerik ağırlıklandırması yöntemlerinin birbirini nasıl etkilediği üzerinde çalışmalar yapılabilir.

Ülkemizde tek boyutlu BOB testi uygulamalarına yönelik çalışmalar olmasına karşın, çok-boyutlu BOB testlerinin gerçek hayatta uygulamalarına ilişkin bir çalışma henüz yapılmamıştır. Bu çalışma sonuçları ışığında gerçek çok-boyutlu BOB testi uygulamaları yapılması önerilmektedir. Ayrıca bu çalışmada bireylerin sadece İngilizce dil becerilerinin çok-boyutlu BOB testleri ile ölçülmesi amaçlanmıştır. Matematik ve fen bilgisi gibi alanlarda da benzer çalışmalar yapılabilir.



## KAYNAKÇA

- Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and non-compensatory multidimensional items. *Applied Psychological Measurement*, 13, 113-127.
- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological Tests. *MIRT Instructional Module/Educational Measurement: Issues and Practice*, 37–53.
- Bayroff, A. G. (1964). *Feasibility of a programmed testing machine* (U.S. Army Personnel Research Office Research Study 6403). Washington, DC: U.S. Army Behavioral Science Research Laboratory.
- Bayroff, A. G., Thomas, J. J., & Anderson, A. A. (1960). *Construction of an experimental sequential item test* (Research Memorandum 60-1). Washington, DC: Department of the Army, Personnel Research Branch.
- Bejar, I.I., & Weiss, D.J. (1979). *Computer programs for scoring test data with item characteristic curve models* (Research Rep. No. 79-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.
- Berger, M.P.F., & Wong, W.K. (Eds.) (2005). *Applied optimal design*. London: Wiley.
- Binet, A. & Simon, T. A. (1905). Méthode nouvelle pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 11, 191-244.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (p. 397-479). Reading, MA: MIT Press.
- Bloxom, B., & Vale, C.D. (1987). Multidimensional adaptive testing: An approximate procedure for updating. In *Meeting of the psychometric society*. Montreal, Canada, June.
- Bobcock, B.G.E.(2009). *Estimating a Noncompensatory IRT Model Using a modified Metropolis algorithm*. Unpublished Doctoral Dissertation. The University of Minesota.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444.
- Boztunç-Öztürk, N. (2014). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı kontrol yöntemlerinin incelenmesi*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi.
- Bulut, O., & Kan, A. (2012) Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eğitim Arastirmalari- Eurasian Journal of Educational Research*, 49, 61–80.

- Castro, F., Suarez, J., & Chirinos, R. (2010). Competence's initial estimation in computer adaptive testing. Paper presented at the first annual conference of the International Association for Computerized Adaptive Testing, Arnhem, The Netherlands.
- Chen, J. (2012). *Applying item response theory methods to design a learning progression-based science assessment*. Unpublished Doctoral Dissertation. Michigan State University.
- Chen, S., & Ankenmann, R.D. (2004). Effects of practical constraints on item selection rules at the early stages of computerized adaptive testing. *Journal of Educational Measurement, 41*(2), 149-174.
- Choi, S. W. & King D. R. (2011). *MAT: Multidimensional adaptive testing*. [Çevirim içi: <https://cran.r-project.org/web/packages/MAT/MAT.pdf>], Erişim tarihi: 15 Temmuz 2015.
- Cikrikci-Demirtaşlı, N. (1999). Psikometride yeni ufuklar: bilgisayar ortamında bireye uyarlanmış test. *Türk Psikoloji Bülteni, 5*(13), 31-36.
- Cleary, T. A., Linn, R. L., & Rock, D. A. (1968). Reproduction of total test score through the use of sequential programmed tests. *Journal of Educational Measurement, 5*, 183–187.
- Cleary, T. A., Linn, R. L., & Rock, D. A. (1969). An exploratory study of programmed tests. *Educational and Psychological Measurement, 28*, 345–360.
- Cömert, M. (2008). *Bireye uyarlanmış bilgisayar destekli ölçme ve değerlendirme yazılımı geliştirilmesi*. Yayınlanmamış Yüksek Lisans Tezi. Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston
- DeMars, C. E. (2010). Type I error inflation for detecting DIF in the presence of impact. *Educational and Psychological Measurement, 70*, 961-972.
- Diao, Q. (2009). *Comparison of ability estimation and item selection methods in multidimensional computerized adaptive testing*. Unpublished Doctoral Dissertation. Michigan State University.
- Diao, Q. & Reckase, M. (2009). Comparison of ability estimation and item selection methods in multidimensional computerized adaptive testing. In: Weiss DJ (Ed.), Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing. pp. 1-13. [Çevrim-içi: <http://publicdocs.iacat.org/cat2010/cat09diao.pdf>] Erişim tarihi: 15 Temmuz 2015.
- Doğan, N. (2002). Klasik Test Teorisi ve Örtük Özellikler Kuramının Örneklemeler Bağlamında Karşılaştırılması. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Eggen. T.J.H.M. & Straetmans. G.J.J.M. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement. 60*(5), 713-734
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

- Erdođdu, B. (2009). *Computer based testing: evaluation of question classification for computer adaptive testing*. Yayınlanmamış Yüksek Lisans Tezi. Bahçeşehir Üniversitesi.
- Ferguson, R. L. (1969). The development, implementation, and evaluation of a computer-assisted branched test for a program of individually prescribed instruction. *Dissertation Abstracts International*, 30(09), 3856A. (UMI No. 704530).
- Fisher, R.A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22, 700–725.
- Folk, V. G., & Green, B. F. (1989). Adaptive estimation when the unidimensionality assumption of IRT is violated. *Applied Psychological Measurement*, 13, 373-389.
- Fan, M., & Hsu, Y. (1996). Multidimensional computer adaptive testing. *In Annual meeting of the American educational research association*. New York City, NY, April.
- Green, B. G., Bock, R.D., Humphries, L. G., Linn, R.L., & Reckase, M.D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21, 347-360.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Academic Publishers.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. *Journal of Psychology*, 216, 89-101. doi: 10.1027/0044-3409.216.2.89
- Harvey, R., & Hammer, A. (1999). Item Response Theory. *The Counseling Psychologist*, 27(3), 353-383.
- IACAT Official Web Site. [Çevirim-içi: <http://iacat.org/content/cat-software> ]. Erişim tarihi: 15 Temmuz 2015.
- Iseri, A. I. (2002). *Assessment of students' mathematics achievement through computer adaptive testing procedures*. Unpublished Doctoral Dissertation. Middle East Technical University.
- Kalender, İ. (2011). Effects of different computerized adaptive testing strategies on recovery of ability. Unpublished Doctoral Dissertation. Middle East Technical University.
- Kaptan, F. (1993). *Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kağıt-kalem testi uygulamasının karşılaştırılması*. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi.
- Karasar, N. (2010). *Bilimsel araştırma yöntemi*. Ankara. Nobel Yayın Dağıtım.
- Kaskatı, O.T. (2011). *Rasch Modelleri Kullanarak Romatoid Artirit Hastaları Özürüllük Değerlendirmesi İçin Bilgisayar Uyarlamalı test Yöntemi Geliştirilmesi*. Yayınlanmamış Doktora Tezi. Ankara Üniversitesi.

- Kezer, F. (2013). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması*. Yayınlanmamış Doktora Tezi. Ankara Üniversitesi.
- Kinsbury, G.G., & Zara, A.R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, 2, 359-375.
- Kinsbury, G.G., & Zara, A.R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied Measurement in Education*, 4(3), 241-261.
- Kullback, S. (1959). *Information theory and statistics*. New York: Wiley.
- Kullback, S. & Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79–86.
- Leung, C.K., Chang, H., & Hau, K. (2000). *Content balancing in stratified computerized adaptive testing designs*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Lin, H. (2012). *Item selection methods in multidimensional computerized adaptive testing adopting polytomously-scored items under multidimensional generalized partial credit model*. Unpublisjed Doctoral Dissertation. University of Illinois at Urbana-Champaign.
- Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come. MESA Memorandum No. 69. Published in S. Chae, U. Kang, E. Jeon & J. M. Linacre. *Development of computerised middle school achievement test* (in Korean). Seoul, South Korea: Komesa Press.
- Lord, F. M. (1970). Some test theory for tailored testing. In W. H. Holtzman (Ed.), *Computer-assisted instruction, testing, and guidance* (pp. 139–183). New York: Harper & Row.
- Lord, F. M. (1971a). Tailored testing, an approximation of stochastic approximation. *Journal of the American Statistical Association*, 66, 707–711.
- Lord, F. M. (1971b). A theoretical study of the measurement effectiveness of flexilevel tests. *Educational and Psychological Measurement*, 31, 805–813.
- Lord, F. M. (1971c). A theoretical study of two-stage testing. *Psychometrika*, 36, 227–242.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.
- Luecht, R. M. (1996). Multidimensional computerized adaptive testing in a certification or licensure context. *Applied Psychological Measurement*, 20 (4), 389-404.
- Luecht, R.M. (1998). *A framework for exploring and controlling risks associated with test item exposure over time*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Makransky, G., & Glas, C. A. W.. (2013). The applicability of multidimensional computerized adaptive testing to cognitive ability measurement in organizational assessment. *International Journal of Testing*, 13(2), 123-139. doi:10.1080/15305058.2012.672352
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

- McBride, J.R., Martin, J.T. (1983). Reliability and Validity of Adaptive Ability Tests in a military setting. in Weiss D.J. (Ed.) *"New Horizons in Testing"* New York: Academic Press.
- Miller, V. D. (2003). *Assessment of student achievement: a comparative study of student achievement using paper and pencil assessment and computerized adaptive testing*. Unpublished dissertation. Wayne State University. Graduate School.
- Mulder, J., & van der Linden, W. J. (2009). Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika*, 74 (2), 273-296.
- 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.
- Owen, R.J. (1969). *A Bayesian approach to tailored testing* (Research Report 69-92). Princeton, NJ: Educational Testin Service.
- 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.
- Patsula, L.N., & Steffan, M. (1997). *Maintaining item and test security in a CAT environment: A simulation study*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Reckase, M., D. (2009). *Multidimensional item response theory: Statistics for social and behavioral sciences*. New York, NY: Springer.
- Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement*, 25, 193-203.
- Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, 15, 361–373.
- Rudner, L. (1998) An On-line, Interactive, Computer Adaptive Testing Mini-Tutorial. *ERIC Clearinghouse on Assessment and Evaluation*.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.
- Sands, W. A., Waters, B. K., & McBride R. (1997), *Computerized adaptive testing: From inquiry to operation* (pp. 199–205). Washington, DC: American Psychological Association.
- Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61(2), 331-354.
- Segall, D.O. (2000). Principles of multidimensional adaptive testing. In W.J. van der Linden & C.A.W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 53–73). Boston: Kluwer Academic.
- Segall, D. O. (2001). General ability measurement: An application of multidimensional item response theory. *Psychometrika*, 66, 79-97.
- Segall, D. O., Moreno, K. E., & Hetter, R. D. (1997). Item pool development and evaluation. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 117–130). Washington, DC: American Psychological Association.

- Sie, H. (2015). A review of SimuMCAT: A simulation software for multidimensional computerized adaptive testing. *Applied Psychological Measurement* 2015, Vol. 39(3) 241–244. DOI: 10.1177/0146621615569503.
- Sijtsma, K. & Junker, B.W. (2006). Item response theory: past performance. Present developments and future expectations. *Behaviormetrika*, 1, 75-102.
- Silvey, S.D. (1980). *Optimal design*. London: Chapman & Hall.
- Song, T. (2012). *The effect of fitting a tridimensional IRT model to multidimensional data in content-balanced computerized adaptive testing*. Unpublished Doctoral Dissertation. Michigan State University
- Spray, J.A., Davey, T., Reckase, M.D., Ackerman, T.A., & Carlson, J.E. (1990). *Comparison of two logistic multidimensional item response theory models*. (ACT Research Report - ONR 90-8).
- Stout, W., Douglas, B., Junker, B., & Roussos, L. (1999). DIMTEST [computer software]. The William Stout Institute for Measurement, Champaign, IL.
- Stocking, M. L. (1998). A framework for comparing adaptive test designs. Unpublished manuscript
- Stout, W., Froelich, A. G., & Gao, F. (2001). Using resampling to produce and improved DIMTEST procedure. In Boomsma A, van Duijn MAJ, Snijders TAB (eds.) *Essays on item response theory* (pp. 357-375). Springer-Verlag, New York.
- Sulak, S. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında kullanılan madde seçme yöntemlerinin karşılaştırılması*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi.
- Tam, S. S. (1992). *A comparison of methods for adaptive estimation of a multidimensional trait*. Unpublished Doctoral Dissertation, Columbia University.
- Tercan, S.S. (2010). *Çevrimiçi uyarlamalı bir test aracı geliştirilmesi, uygulanması ve öğrenci görüşlerinin belirlenmesi*. Yayınlanmamış Yüksek Lisans Tezi. Sakarya Üniversitesi.
- Thompson, N. A. (2007). A practitioner's guide for variable-length computerized classification testing. *Practical Assessment Research & Evaluation*, 12(1).
- Thompson, N.A., & Weiss, D.J. (2011). A Framework for the Development of Computerized Adaptive Tests *Practical Assessment, Research, and Evaluation*, 16(1). Also presented at the 2011 Innovations in Testing Conference, Phoenix, AZ.
- Urry, V. W. (1970). *A Monte Carlo investigation of logistic test models*. Unpublished Doctoral Dissertation. West Lafayette, IN: Purdue University.
- van der Linden, W.J. (1996). Assembling tests for the measurement of multiple traits. *Applied Psychological Measurement*, 20, 373–388.
- van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.
- van der Linden, W. J. (1998). Bayesian item-selection criteria for adaptive testing. *Psychometrika*, 63, 201-216.

- van der Linden, W.J. (1999). Multidimensional adaptive testing with a minimum error-variance criterion. *Journal of Educational and Behavioral Statistics*, 24, 398–412.
- Van der Linden, W. J., & Glas, C. A. W. (2000). *Computerized adaptive testing: Theory and practice*. Boston: Kluwer.
- 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.
- Veldkamp, B. P. , & van der Linden, W. J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika*, 67(4), 575-588.
- Yao, L. (2012). Multidimensional CAT Item Selection Methods for Domain Scores and Composite Scores: Theory and Applications. *Psychometrika*, 77, 495-523.
- 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 methods for domain scores and composite scores with item exposure control and content constraints. *Journal of Educational Measurement*, 51, 18?38. doi:10.1111/jedm.12032.
- Yao, L., Pommerich, M., & Segall, D. O.. (2014). Using multidimensional CAT to administer a short, yet precise, screening test. *Applied Psychological Measurement*, 38, 614-631. doi:10.1177/0146621614541514.
- Yoo, H. (2011). *Evaluating several multidimensional adaptive testing procedures for diagnostic assessment*. Unpublished Doctoral Dissertation. University of Massachusetts Amherst
- Wainer, H. (1993). Some practical considerations when converting a linearly administered test to an adaptive format. *Educational Measurement: Issues and Practices*, 12, 15-20.
- Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B., Mislevy, R., et al. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Wainer, H., & Kiely, G.L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 6, 473-492.
- Wang, C., Chang, H., & Boughton, K. A. (2011). Kullback-Leibler information and its applications in multi-dimensional adaptive testing. *Psychometrika*, 76 (1), 13-39.
- Wang, C., & Chang, H. (2011). Item selection in multidimensional computerized adaptive tests: Gaining information from different angles. *Psychometrika*, 76(3), 363-384.
- Wang, T., & Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35, 109-135.

- Wang, W. C. & Chen, P.H. (2004). Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Applied Psychological Measurement* 28: 295. DOI: 10.1177/0146621604265938.
- Wang, W., Chen, P., & Cheng, Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods*, 9, 116–136.
- Wang, W.-C., Wilson, M., and Adams, R. (1997). Rasch models for multidimensionality between items and within items. In G. Englehard, Wilson, Mark (Ed.), *Objective Measurement* (Vol. 4, ): Greenwich, CN: Ablex Publishing.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in the item response theory. *Psychometrika*, 54, 427-450.
- Weiss, D. J. (1983). *New horizons in testing: Latent trait test theory and computerized adaptive testing*. New York: Academic Press.
- Weiss, J. D. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, Volume 37, Number 2, July 2004, pp. 70-84.
- Weiss, D. J. (2011). Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences*, 2(1), 1-23.
- Weiss, D. J., & Betz, N. E. (1973). *Ability measurement: Conventional or adaptive? (Research Report 73-1)*. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory.
- Weiss, D. J., & McBride, J. R. (1984). Bias and information of Bayesian adaptive testing. *Applied Psychological Measurement*, 8, 273-285.
- Weiss, D.J., & Kingsbury, G.G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21:4 361-375.
- Weiss, D. J., & Yoes, M. E. (1991). Item response theory. In R.K. Hambleton & J. N. Zaal (Eds.), *Advances in educational and psychological testing: Theory and applications* (pp. 69-95). Boston: Kluwer Academic.
- Zhang, J. M., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika* 64: 213-249.



## **EKLER DİZİNİ**

# EK 1. ETİK KURUL ONAY BİLDİRİMİ

Form: 40

## Tez Çalışması Etik Kurul İzin Muafiyeti Formu

22 / 07 / 2015

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

**Tez Başlığı / Konusu:** MADDE DÜZEYİNDE BOYUTLULUK MODELLERİNİN BİLGISAYAR ORTAMINDA BİREYSELLEŞTİRİLMİŞ TEST YÖNTEMLERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

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 Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik 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.

  
Burhanettin Özdemir  
(Öğrencinin Adı Soyadı, İmzası)

### Öğrenci Bilgileri

Adı Soyadı	Burhanettin ÖZDEMİR
Öğrenci No	H11168579
Anabilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	<input type="checkbox"/> Yüksek Lisans <input type="checkbox"/> Doktora <input checked="" type="checkbox"/> Bütünleşik Dr.

### Danışman Görüşü ve Onayı

Bu çalışmada hazır veriye dayalı simülasyon çalışması olduğundan, tez sürecinde veri toplanmadığından ve ölçek geliştirme çalışması olmadığından etik kurul izninden muaf tutulması uygundur.

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

## Ek-2. Tek Boyutlu ve Çok-Boyutlu Madde-İçi Boyutluluk Modeline İlişkin Uyum İstatistikleri

		<i>AIC</i>	<i>BIC</i>	<i>logLik</i>	<i>X2</i>	<i>df</i>	<i>p</i>
test1	tek boyutlu	53561,32	54156,07	-26654,66			
	üç boyutlu	53380,37	54555,71	-26441,19	426,954	123	0,00
test2	tek boyutlu	80020,2	80653,53	-39884,10			
	üç boyutlu	79833,96	81085,54	-39667,98	432,248	123	0,00
test3	tek boyutlu	13763,55	14187,67	-6755,78			
	üç boyutlu	13761,25	14599,37	-6631,62	248,306	123	0,00
test4	tek boyutlu	34047,26	34541,72	-16905,63			
	üç boyutlu	33906,99	34883,33	-16720,49	370,271	115	0,00
test5	tek boyutlu	77190,55	77848,93	-38467,27			
	üç boyutlu	76868,67	78170,01	-38181,34	571,875	125	0,00
test6	tek boyutlu	43345,46	43893,67	-21546,73			
	üç boyutlu	43320,42	44403,79	-21411,21	271,037	123	0,00
test7	tek boyutlu	82471,21	83116,29	-41107,61			
	üç boyutlu	82236,9	83511,93	-40865,45	484,317	125	0,00
test8	tek boyutlu	36414,95	36958,91	-18081,47			
	üç boyutlu	35720,2	36795,17	-17611,10	940,748	123	0,00
test9	tek boyutlu	14769,94	15189,26	-7258,97			
	üç boyutlu	14793,3	15621,94	-7147,65	222,645	123	0,00
test10	tek boyutlu	13506,55	13906,42	-6625,28			
	üç boyutlu	13502,73	14293,1	-6498,37	253,82	125	0,00

**Ek-3. Tek ve Çok-Boyutlu Maddeler-Arası Boyutluluk Modeline İlişkin Uyum İstatistikleri**

		<i>AIC</i>	<i>BIC</i>	<i>logLik</i>	<i>X2</i>	<i>df</i>	<i>p</i>
test1	tek boyutlu	53561,32	54156,07	-26654,66			
	üç boyutlu	53469,59	54078,49	-26605,79	97,739	3	0,00
test2	tek boyutlu	80020,2	80653,53	-39884,10			
	üç boyutlu	79925,3	80573,71	-39833,65	100,9	3	0,00
test3	tek boyutlu	13763,55	14187,67	-6755,78			
	üç boyutlu	13711,28	14145,49	-6726,64	58,269	3	0,00
test4	tek boyutlu	34047,26	34541,72	-16905,63			
	üç boyutlu	33975,8	34482,83	-16866,90	77,458	3	0,00
test5	tek boyutlu	77190,55	77848,93	-38467,27			
	üç boyutlu	77007,56	77681,37	-38372,78	188,99	3	0,00
test6	tek boyutlu	43345,46	43893,67	-21546,73			
	üç boyutlu	43310,79	43872,06	-21526,40	40,663	3	0,00
test7	tek boyutlu	82471,21	83116,29	-41107,61			
	üç boyutlu	82418,67	83078,87	-41078,34	58,541	3	0,00
test8	tek boyutlu	36414,95	36958,91	-18081,47			
	üç boyutlu	36167,39	36724,31	-17954,70	253,55	3	0,00
test9	tek boyutlu	14769,94	15189,26	-7258,97			
	üç boyutlu	14751,08	15180,38	-7246,54	24,864	3	0,00
test10	tek boyutlu	13506,55	13906,42	-6625,28			
	üç boyutlu	13499,69	13908,93	-6618,85	12,863	3	0,00

#### Ek-4. Çok-Boyutlu Maddeler-Arası ve Madde-İçi Boyutluluk Modeline İlişkin Uyum İstatistikleri

		<i>AIC</i>	<i>BIC</i>	<i>logLik</i>	<i>X2</i>	<i>df</i>	<i>p</i>
test1	maddeler-arası	53469,59	54078,49	-26605,79			
	madde-içi	53380,37	54555,71	-26441,19	329,215	120	0,00
test2	maddeler-arası	79925,3	80573,71	-39833,65			
	madde-içi	79833,96	81085,54	-39667,98	331,348	120	0,00
test3	maddeler-arası	13711,28	14145,49	-6726,64			
	madde-içi	13761,25	14599,37	-6631,62	190,036	120	0,00
test4	maddeler-arası	33975,8	34482,83	-16866,90			
	madde-içi	33906,99	34883,33	-16720,49	292,813	112	0,00
test5	maddeler-arası	77007,56	77681,37	-38372,78			
	madde-içi	76868,67	78170,01	-38181,34	382,885	122	0,00
test6	maddeler-arası	43310,79	43872,06	-21526,40			
	madde-içi	43320,42	44403,79	-21411,21	230,374	120	0,00
test7	maddeler-arası	82418,67	83078,87	-41078,34			
	madde-içi	82236,9	83511,93	-40865,45	425,776	122	0,00
test8	maddeler-arası	36167,39	36724,31	-17954,70			
	madde-içi	35720,2	36795,17	-17611,10	687,198	120	0,00
test9	maddeler-arası	14751,08	15180,38	-7246,54			
	madde-içi	14793,3	15621,94	-7147,65	197,782	120	0,00
test10	maddeler-arası	13499,69	13908,93	-6618,85			
	madde-içi	13502,73	14293,1	-6498,37	240,957	122	0,00

## EK 4. ORJİNALLİK RAPORU

Turnitin Document Viewer - Google Chrome  
https://api.turnitin.com/dv?u=1041173245&s=3&lang=en\_us&o=561244737

tez tez - DUE 21-Aug-2015 Roadmap Paper 1 of 1

Originality GradeMark PeerMark

MADDE DÜZEYİNDE BOYUTLULUK MODELLERİNİN BİLGİSAYAR ORTAMINDA  
BY BURHANETTİN ÖZDEMİR

turnitin 5% SIMILAR OUT OF 0

**MADDE DÜZEYİNDE BOYUTLULUK MODELLERİNİN  
BİLGİSAYAR ORTAMINDA BİREYSELLEŞTİRİLMİŞ TEST  
YÖNTEMLERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ**

**EXAMINING THE EFFECTS OF ITEM LEVEL  
DIMENSIONALITY MODELS ON MULTIDIMENSIONAL  
COMPUTERIZED ADAPTIVE TESTING METHODS**

Burhanettin ÖZDEMİR

**Match Overview**

1	www.pegem.net Internet source	1%
2	Submitted to TechNo... Student paper	1%
3	Özkan, Yeşim Özer. "A... Publication	<1%
4	Submitted to Middle E... Student paper	<1%
5	www.iacat.org Internet source	<1%
6	www.researchgate.net Internet source	<1%
7	www.jret.org Internet source	<1%
8	Submitted to Bilkent U... Student paper	<1%

PAGE: 1 OF 148 Text-Only Report

## ÖZGEÇMİŞ

### Kişisel Bilgiler

<b>Adı Soyadı</b>	Burhanettin Özdemir
<b>Doğum Yeri</b>	Köprüköy/ERZURUM
<b>Doğum Tarihi</b>	30.11.1987

### Eğitim Durumu

<b>Lise</b>	Erzurum Merkez Anadolu Lisesi	2005
<b>Lisans</b>	Atatürk Üniversitesi	2010
<b>Bütünleşik Doktora</b>	Hacettepe Üniversitesi	2015
<b>Yabancı Dil</b>	İngilizce: Okuma (Çok iyi), Yazma (Çok İyi), Konuşma (Çok iyi) Çince: Okuma (Orta), Yazma (Orta), Konuşma (İyi)	

### İş Deneyimi

<b>Çalıştığı Kurumlar</b>	Siirt Üniversitesi	2010-2011
---------------------------	--------------------	-----------

### Akademik Çalışmalar

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

<p>Ozdemir, B. (2014). <i>Partitioning Variance into Constituents in Regression Models: Commonality Analysis. Chapter 25. Ed. R.E Millsap, D. M. Bolt, L. A. van der Ark and W-C Wang. Quantitative Psychology Research, NC: Springer</i></p> <p>Ozdemir B &amp; Gelbal S., (2014). "Investigating Factors That Affect Turkish Students' Academic Success with Canonical Commonality Analysis According to PISA 2009 Results". Education and Science Vol 39 (2014) No 175 41-57</p> <p>Ozdemir, B. (2015). <i>A comparison of IRT-based methods for examining differential item functioning in TIMSS 2011 mathematics subtest</i>, Procedia - Social and Behavioral Sciences, Elsevier inc</p> <p>Ozdemir B., Gelbal S., (2013) "Determining Relative Importance of Factor Affecting Students Academic Success by Rank-order Scaling Methods" Proceeding International Conference on New Horizons in Education,INTE 2013, Rome, June 24<sup>th</sup>-27<sup>th</sup> 2013</p>
---

### Seminer ve Çalıştaylar

<p>Ozdemir, B. "Testlet Effect on Different IRT-Based DIF Methods: Bayesian Testlet Response Theory Approach". Psychometric Society, the 80<sup>th</sup> Annual Meeting of Psychometric Society, IMPS 2015, Beijing, China, July 12<sup>th</sup>-16<sup>th</sup> 2015</p> <p>Ozdemir, B. "Equating TIMSS Mathematic Subtests with Nonlinear Equating Methods Using NEAT Design: Circle-Arc Equating Approaches". International Congress</p>
---

on Education For the Future: Issues and Challenges, ICEFIC 2015 . Ankara, Turkey, May 13<sup>th</sup>- 15<sup>th</sup> 2015

Ozdemir, B. “*Measuring Language Ability of Students with CAT: An Application of Post-Hoc Simulation to English Proficiency Test*”. The 2014 Computerized Adaptive Testing Summit. IACAT, Princeton, New Jersey, October 8<sup>th</sup>- 10<sup>th</sup> 2014

Ozdemir, B.. “*Examining Testlet Effects In English Proficiency Test: A Bayesian Testlet Response Theory Approach*”. Psychometric Society, the 79<sup>th</sup> Annual Meeting of Psychometric Society, IMPS 2014, Madison Wisconsin, July 20<sup>nd</sup>-26<sup>th</sup> 2014

Ozdemir, B.. “*A Comparison of Different Linear And Nonlinear Observed Score Equating Methods Using Random Group Design*”. IV. National Congress on Measurement and Statistics in Education and Psychology. EPOD, Ankara,Turkey, June 9<sup>th</sup> – June 13<sup>th</sup> 2014

Ozdemir B., “*Partitioning Variance into Constituents in Regression Models: Commonality Analysis*” Psychometric Society, the 78<sup>th</sup> Annual Meeting of Psychometric Society, IMPS 2013, Arnhem, July 22<sup>nd</sup>-26<sup>th</sup> 2013

Ozdemir B., Gelbal S., “*Determining Relative Importance of Factor Affecting Students Academic Success by Rank-order Scaling Methods*” International Conference on New Horizons in Education,INTE 2013, Rome, June 24<sup>th</sup>-27<sup>th</sup> 2013

Ozdemir B., Yagci E., “*Teachers Opinion on New Mathematics Curriculum and Integers Subject Covered by Program: A Case Study* ” II. National Congress on Curriculum And Instruction, Bolu, September 27<sup>th</sup>-29<sup>th</sup> 2012

## Sertifikalar

2004-2005 Eğitim ve Öğretim yılında Amerika Eğitim Bakanlığı ve Türkiye Amerikan Başkonsolosluğu'nun birlikte düzenlediği ‘**American Culture and Languages**’ Programını başarıyla tamamlanarak İngilizce Sertifikası alındı.

2008-2009 Eğitim ve Öğretim yılında **PAX programı (Programing of Academic Exchange)** aracılığı ile 1 yıl Tayvan'dan Uluslararası Chung Hsing Üniversitesi'nde yüksek lisans dersleri alındı. Bu program kapsamında ayrıca 1 yıl Çince eğitimi alıp her iki dönemde de Çince sertifikası alındı.

2011 Erzurum Üniversiteler Kış Olimpiyatı(2011 Universiade Winter Olympics)'nda Çin-Tayvan Ateşesi olarak görev aldım ve olimpiyat sertifikası alındı.

2011 yılında YÖK'ün belirlemiş olduğu Dil Programı çerçevesinde Malta'da 3 aylık İngilizce Dil Eğitimi Programına (English Language Programme) katıldım. Program Sonunda Advanced English sertifikası alındı.

## İletişim

<b>e-Posta Adresi</b>	b.ozdemir@hacettepe.edu.tr
	b.ozdemir025@gmail.com

<b>Jüri Tarihi</b>	22.06.2015
--------------------	------------