

**GERÇEK VE ÜRETİLMİŞ VERİ SETLERİNDE ÇOK
BOYUTLU BİREYSELLEŞTİRİLMİŞ BİLGİSAYARLI TEST
UYGULAMA SONUÇLARININ İNCELENMESİ**

**EXAMINING THE RESULTS OF MULTIDIMENSIONAL
COMPUTERIZED ADAPTIVE TESTING APPLICATIONS IN
REAL AND GENERATED DATA SETS**

Murat Doğan ŞAHİN

Hacettepe Üniversitesi

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Doktora Tezi olarak hazırlanmıştır.

2017

KABUL ve ONAY

Eğitim Bilimleri Enstitüsü Müdürlüğü'ne,

Murat Doğan ŞAHİN'in hazırladığı "Gerçek ve Üretilmiş Veri Setlerinde Çok Boyutlu Bireyselleştirilmiş Bilgisayarlı Test Uygulama Sonuçlarının İncelenmesi" başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiştir.

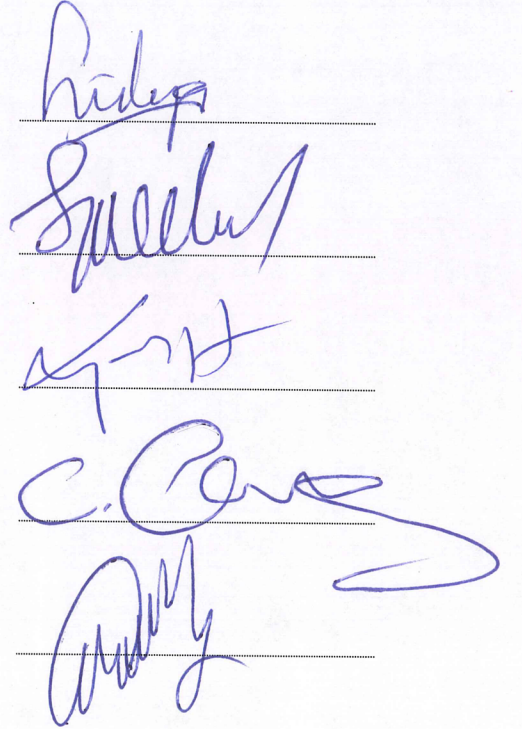
Başkan Prof. Dr. Hülya KELECİOĞLU

Üye (Danışman) Prof. Dr. Selahattin GELBAL

Üye Prof. Dr. Nuri DOĞAN

Üye Doç. Dr. Cem Oktay GÜZELLER

Üye Yard. Doç. Dr. Murat AKYILDIZ



ONAY

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 14 / 04 / 2017 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca / / tarihinde kabul edilmiştir.

Prof. Dr. Ali Ekber ŞAHİN
Eğitim Bilimleri Enstitüsü Müdürü

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

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

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

☐ Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirse bile, teziniz arama motorlarının önbelleklerinde kalmaya devam edebilecektir)

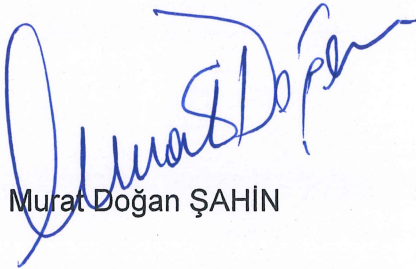
☐ Tezimin/Raporumun tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir).

☒ Tezimin/Raporumun ...14.04.2018... tarihine kadar erişime açılmasını istemiyorum ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.

☐ Serbest Seçenek/Yazarın Seçimi:

.....



Murat Doğan ŞAHİN

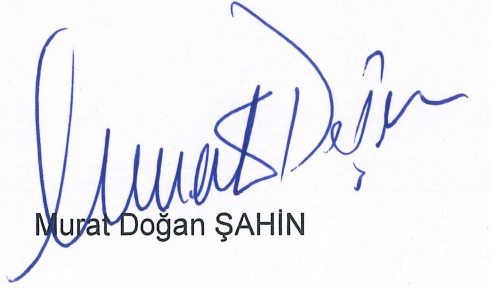
29/05/2017

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.



Murat Doğan ŞAHİN

TEŞEKKÜR

Verdiği destek ve yüreklendirici tavrıyla doktora eğitim ve tez yazım sürecini büyük bir motivasyonla yürütmemi sağlayan değerli danışmanım Prof. Dr. Selahattin GELBAL'a teşekkürü bir borç bilirim. Gerek danışmanlığımı yürüttüğü tezimin fikir sürecinde, gerekse sonrasında ve son olarak tez jürisinde verdiği değerli katkılar ve cesaretlendirici tavrı için Doç. Dr. Cem Oktay GÜZELLER'e teşekkür ederim. Tez jürisinde yer alan değerli hocalarım Prof. Dr. Hülya KELECİOĞLU ve Prof. Dr. Nuri DOĞAN'a, gerek tezimin gelişimine sundukları katkılar ve gerekse ders dönemindeki verimli süreç için, içten teşekkürlerimi sunarım.

Beraber çalışmaya başladığımız günden itibaren akademik birikimi, ölçme ve değerlendirme alanına bakış açısı ve vizyonu ile kişisel ve akademik gelişime büyük katkı sağlayan, yardıma ihtiyacım olduğu her anda sonsuz destek veren değerli hocam-abim Yard. Doç. Dr. Murat AKYILDIZ'a sonsuz teşekkürlerimi sunarım.

Lisansüstü öğrenimim boyunca gösterdiği destek ve yakınlık için sevgili hocam Doç. Dr. Duygu ANIL'a, doktora ders dönemini verimli geçirmemizde büyük katkısı olan Doç. Dr. Burcu ATAR'a içten teşekkürlerimi sunarım.

Veri toplama ve uygulama sürecimi kendi projeleri gibi sahiplenen, ihtiyacım olan her türlü desteği büyük bir özveri ve samimiyetle sunan Anadolu Üniversitesi YDYO Müdürü Prof. Dr. Belgin AYDIN'a, Müdür Yardımcıları Okt. Meral Melek Ünver ve Okt. Bülent ALAN'a teşekkürü bir borç bilirim.

Gözümü açtığım andan beri her zaman yanımda olan Babama; bugüne ulaşmamda en büyük pay sahibi, çok özlediğim rahmetli Anneme şükranlarımı sunarım.

Hayatıma girdiği andan itibaren her konudaki en büyük destekçim; sevgisi ve inancıyla tez sürecimi her daim kolaylaştıran sevgili eşim ve en iyi dostum Özlem'e minnetimi ve en derin sevgimi sunarım.

Gökyüzündeki Anneme...

“..söyle, kim önleyebilir kavuşmamızı?”

“I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you can not measure it, when you can not express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of *Science*, whatever the matter may be.”

- *William Thompson (Lord Kelvin)*

GERÇEK ve ÜRETİLMİŞ VERİ SETLERİNDE ÇOK BOYUTLU BİREYSELLEŞTİRİLMİŞ BİLGİSAYARLI TEST UYGULAMA SONUÇLARININ İNCELENMESİ

Murat Doğan ŞAHİN

ÖZ

Bu çalışmada; Anadolu Üniversite YDYO'da öğrenim gören hazırlık sınıfı öğrencilerinin eğitimlerini bitirmelerinin ardından, lisans eğitimleri öncesi vermeleri gereken Yeterlilik Tespit Sınavının (YTS) verileri kullanılarak gramer ve kelime bilgisi becerileriyle birlikte bir de genel yetenek ölçeği çok boyutlu bir bireyselleştirilmiş bilgisayarlı test (BBT) geliştirilmesi amaçlanmıştır. Buna yönelik olarak öncelikle, Anadolu Üniversitesi YDYO'da geçmiş yıllardaki YTS'lerde uygulanan ve her biri 50'şer sorudan oluşan dört farklı madde seti ile bu maddelere ait veriler kullanılarak bir madde havuzu oluşturulmuştur. MTK'ya dayalı olarak yapılan analizlerde her bir veri seti için sırasıyla tek boyutlu model, iki boyutlu model ve üç boyutlu bifaktör model uygulanmış ve tüm madde setleri için en uygun modelin üç boyutlu bifaktör model olduğu sonucuna varılmıştır. Madde havuzunun oluşturulmasının ardından, gerçek-zamanlı çok boyutlu BBT uygulamasında kullanılacak algoritmayı belirlemek amacıyla, eldeki eksik veri matrisi de göz önüne alınarak bir *hibrit* simülasyon uygulaması yapılmıştır. Bu simülasyon uygulamasında farklı yetenek kestirim yöntemlerinin (EAP ve MAP), farklı madde seçim yöntemlerinin (D-rule, KL, W-rule, T-rule, ağırlıklandırılmış W-rule ve ağırlıklandırılmış T-rule) ve farklı sonlandırma kurallarının (standart hata, θ değişimi ve sabit madde) kullanıldığı toplam 36 farklı koşul yaratılmış; bu koşullarda her bir boyut için gerçek ve kestirilen θ değerleri arasındaki korelasyon ile birlikte yanlışlık, RMSD ve standart hata değerleri elde edilmiştir. Korelasyon değerleri ve hata istatistikleri ile birlikte, BBT uygulamalarında ölçme kesinliğinin önemli bir göstergesi olduğundan, değişken madde sayısına dayalı sonlandırma kuralının kullanıldığı koşullarda yanıtlanan madde sayısı da raporlanmıştır. Elde edilen sonuçlara göre; üç farklı sonlandırma kuralı için de D-rule madde seçim yöntemi ve MAP yetenek kestirim yönteminin en iyi istatistikleri verdiği bulgulanmıştır. Bir sonraki adımda her bir sonlandırma kuralı için en iyi koşullar karşılaştırılarak gerçek-zamanlı çok boyutlu

BBT uygulaması için D-rule madde seçim yöntemi ve MAP yetenek kestirim yöntemiyle birlikte standart hataya dayalı sonlandırma kuralının kullanılmasına karar verilmiştir. Söz konusu algoritma kullanılarak gerçek-zamanlı çok boyutlu BBT uygulamasının yapıldığı 99 öğrenciden 32'sinin kağıt-kalem testiyle elde edilen toplam puanları ile gerçek-zamanlı uygulama sonucu elde edilen genel yetenek düzeyleri arasındaki korelasyon elde edilmiştir. Bunun yanında, uygulamanın yapıldığı toplam 99 kişinin sonuçlarına dayanarak havuzda yer alan maddelerin kullanım sıklıkları ve her bir bireyin yanıtladığı toplam madde sayısı da raporlanmıştır. Gerçek-zamanlı uygulama sonucu elde edilen sonuçlara göre; 200 maddeden oluşan havuzun %30'unun kullanıldığı görülmüştür. Testin bitirilmesiyle yanıtlanan ortalama madde sayısının 12.3 olduğu, ayrıca katılımcıların yaklaşık %75'inin 10 ile 12 arasında madde yanıtladıkları sonucuna ulaşılmıştır

Anahtar sözcükler: Çok Boyutlu Bireyselleştirilmiş Bilgisayarlı Test, Madde Tepki Kuramı, Bifaktör Model, Hibrit Simülasyon, Gerçek-Zamanlı Uygulama

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

EXAMINATION OF MULTIDIMENSIONAL COMPUTERIZED ADAPTIVE TESTING APPLICATION RESULTS IN REAL AND GENERATED DATA SETS

Murat Doğan ŞAHİN

ABSTRACT

This study aims to develop a multi-dimensional computerized adaptive test (MCAT) that measures grammar and vocabulary knowledge as well as general ability of the students attending Anadolu University School of Foreign Languages by using the data from Proficiency Exams administered for these student at the end of preparatory school education. To achieve this purpose, an item pool was created by using four different data sets each of which consists of 50 questions used in the previous proficiency exams administered at Anadolu University School of Foreign Languages. According to IRT-based analyses applied to the data, it was found that three-dimensional bifactor model was the most appropriate model for all item sets. Following the development of the item pool, a hybrid simulation was applied –by taking the missing matrix into consideration – in order to determine the algorithm to be used in real-time MCAT application. In this application, a total of 36 conditions were created in which different ability estimation methods (EAP and MAP), different item selection methods (D-rule, KL, W-rule, T-rule, weighted W-rule and weighted T-rule) and different termination rules (standard error, θ convergence, fixed number of items) were used. In these conditions, the correlation between estimated θ value and real value for each dimension; bias, RSMD and standard error values were obtained. Being an important indicator of measurement accuracy in CAT applications in addition to correlation values and error statistic, the number of items replied was also reported under the variable-length simulation conditions. According to the results obtained, it was found that D-rule item selection method and MAP ability prediction method give the best statistics for each different termination rule. Following the comparisons of best conditions for each termination rule, it was decided that the termination rules based on standard error as well as D-rule item selection method and MAP ability prediction method should be used for real-time MCAT application. By using this algorithm, a correlation was obtained between general ability levels obtained from the real-time application and the scores 32

students out of 99 received from paper-pencil test, who were applied real-time application. In addition, the total number of items administered to each individual and how frequently the items in item pool are used were reported based on the results obtained from the application carried out for 99 students. According to the results obtained from real-time application, it was found that 30% of the item pool, which includes a total of 200 items were used. It was also observed that the average number of items replied was 12.3 and nearly 75% of the participants replied 10 to 12 items.

Keywords: Multi-dimensional Computerized Adaptive Testing, Item Response Theory, Bifactor Model, Hybrid Simulation, Real-time Application

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

İÇİNDEKİLER

KABUL ve ONAY.....	ii
ETİK BEYANNAMESİ	iv
TEŞEKKÜR.....	v
ÖZ	vii
ABSTRACT	ix
İÇİNDEKİLER.....	xi
TABLolar DİZİNİ	xiv
ŞEKİLLER DİZİNİ.....	xv
SİMGELER VE KISALTMALAR DİZİNİ	xvi
1. GİRİŞ	1
1.1. Tek Boyutlu ve Çok Boyutlu Bireyselleştirilmiş Bilgisayarlı Testlerin Tarihi.....	5
1.2. Problem Durumu.....	8
1.3. Araştırmanın Amacı ve Önemi.....	9
1.4. Problem Cümlesi	11
1.4.1. Alt Problemler.....	11
1.5. Sayıtlar.....	13
1.6. Sınırlılıklar.....	13
1.7. Tanımlar.....	13
1.8. Araştırmanın Kuramsal Temeli	14
1.8.1 Madde Tepki Kuramı	14
1.8.1.1. Tek Boyutlu Madde Tepki Kuramı.....	14
1.8.1.2. Çok Boyutlu Madde Tepki Kuramı	18
1.8.2. Bireyselleştirilmiş Bilgisayarlı Testleri Geliştirme ve Uygulama Süreci	22
1.8.2.1. Madde Havuzunun Oluşturulması.....	24
1.8.2.2. Teste Başlama Kuralı	26
1.8.2.3. Yetenek Kestirim Yöntemi	26
1.8.2.3.1. Maksimum Olabilirlik Yöntemi (ML).....	27
1.8.2.3.2. Bayes Kestirim Yöntemleri	28
1.8.2.4. Madde Seçim Yöntemi.....	28
1.8.2.5. Sonlandırma Kuralı	30
1.8.2.6. İçerik Ağırlıklandırması ve Madde Teşhir Oranının Kontrol Edilmesi	31
1.8.3. Simülasyon Süreci.....	32
1.8.4. BBT Uygulamalarının Avantajları ve Sınırlılıkları.....	33
1.8.5. Simülasyon ve Gerçek-Zamanlı BBT Uygulamaları için Kullanılan Yazılımlar	34
2. İLGİLİ ARAŞTIRMALAR.....	36
2.1. Çok Boyutlu BBT Uygulamaları ile İlgili Yurtdışında Yapılan Çalışmalar	36
2.2. BBT Uygulamaları ile İlgili Yurtiçinde Yapılan Çalışmalar	43

3. YÖNTEM	52
3.1. Araştırmanın Türü	52
3.2. Verilerin Elde Edilmesi	52
3.3. Verilerin Analizi	53
3.3.1. Madde Havuzunun Oluşturulması	54
3.3.1.1. Model – Veri Uyumunun Test Edilmesi	54
3.3.1.2. Madde Havuzunda Yer Alacak Maddelerin Kalibrasyonu	57
3.3.2. Simülasyon Süreci	60
3.3.3. Gerçek-Zamanlı Çok Boyutlu BBT (real-time MCAT) Uygulama Süreci	65
3.3.3.1. mirtCAT Paketinin Özelleştirilerek Yazılımın Oluşturulması	65
3.3.3.2. Gerçek-Zamanlı Uygulamanın Yapılması	70
4. BULGULAR	71
4.1. Alt Problem 1: Standart hataya dayalı sonlandırma kuralının kullanıldığı çok boyutlu BBT uygulamalarında farklı yetenek kestirim yöntemleri (EAP ve MAP) ve madde seçim kurallarına (D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule) göre gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanıtlanan ortalama madde sayısı, yanlılık (bias) ve RMSD değerleri nasıldır?	71
4.2. Alt Problem 2: θ değişimine dayalı sonlandırma kuralının kullanıldığı çokboyutlu BBT uygulamalarında farklı yetenek kestirim yöntemleri (EAP ve MAP) ve madde seçim kurallarına (D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule) göre gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanıtlanan ortalama madde sayısı, yanlılık (bias) ve RMSD değerleri nasıldır?	75
4.3. Alt Problem 3: Sabit madde sayısına dayalı sonlandırma kuralının kullanıldığı çok boyutlu BBT uygulamalarında farklı yetenek kestirim yöntemleri (EAP ve MAP) ve madde seçim kurallarına (D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule) göre gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanıtlanan ortalama madde sayısı, yanlılık (bias) ve RMSD değerleri nasıldır?	78
4.4. Alt Problem 4: Simülasyon sonuçlarına göre her bir sonlandırma kuralı için en uygun çok boyutlu BBT koşulu nedir?	81
4.5. Alt Problem 5: Gerçek-zamanlı çok boyutlu BBT uygulaması sonuçları nasıldır?	101
5. TARTIŞMA	105
5.1 Simülasyon Uygulaması Sonucu Elde Edilen Bulgulara İlişkin Yorumlar ...	106
5.2 Gerçek-Zamanlı Çok Boyutlu Uygulaması Sonucu Elde Edilen Bulgulara İlişkin Yorumlar	108
6. SONUÇ ve ÖNERİLER	110
6.1. Sonuçlar	110
6.2. Öneriler	114
6.2.1. Gelecekte Yapılacak Çok Boyutlu BBT Araştırmalarına Yönelik Öneriler	114

6.2.2 Gerçek-Zamanlı Çok Boyutlu BBT Uygulamalarına Yönelik Öneriler ..	116
KAYNAKÇA.....	118
EK 1. ETİK KOMİSYON İZİN MUAFİYETİ FORMU	126
EK 2. ORJİNALLİK RAPORU	127
EK 3. ÇOK BOYUTLU BBT UYGULAMASI İÇİN KULLANILAN HAVUZDA YER ALAN MADDELERE AİT PARAMETRELER.....	129
EK 4. SİMÜLASYONLAR SONUCU HER BİR KOŞULDA TÜM BOYUTLAR İÇİN ELDE EDİLEN GRAFİKLER.....	131
EK 5. GERÇEK-ZAMANLI ÇOK BOYUTLU BBT UYGULAMASINDA KULLANILAN R KODLARI	158
EK 6. FARKLI SONLANDIRMA KURALLARININ KULLANILDIĞI KOŞULLAR İÇİN ELDE EDİLEN İSTATİSTİKLERE AİT GRAFİKLER.....	159
ÖZGEÇMİŞ	165

TABLolar DİZİNİ

Tablo 3.1: Madde Setlerinin Farklı Faktör Analitik Yöntemlere Göre Kalibre Edilmesi Sonucu Elde Edilen Uyum İstatistikleri	56
Tablo 3.2: Madde Setlerinin Madde Parametrelerinin Değişmezliğine İlişkin Korelasyon Değerleri.....	58
Tablo 3.3: Madde Setlerinin Yetenek Parametrelerinin Değişmezliğine İlişkin Korelasyon Değerleri.....	59
Tablo 3.4: Eksik Birey-Yanıt Matrisinin Gösterimi.....	61
Tablo 3.5: Simülasyon Çalışmasındaki 36 Koşulu Oluşturan BBT Bileşenleri.....	65
Tablo 4.1: Standart Hataya Dayalı Durdurma Kuralının Kullanıldığı Koşullarda Elde Edilen Korelasyon, Yanlılık, RMSD, Standart Hata ve Yanıtlanan Madde Sayısı İstatistikleri.....	72
Tablo 4.2: θ Değişimine Dayalı Durdurma Kuralının Kullanıldığı Koşullarda Elde Edilen Korelasyon, Yanlılık, RMSD, Standart Hata ve Yanıtlanan Madde Sayısı İstatistikleri.....	76
Tablo 4.3: Sabit Madde Sayısına Dayalı Durdurma Kuralının Kullanıldığı Koşullarda Elde Edilen Korelasyon, Yanlılık, RMSD, Standart Hata ve Yanıtlanan Madde Sayısı İstatistikleri.....	79
Tablo 4.4: Her Bir Sonlandırma Kuralı için En İyi İstatistikleri Veren Koşullara Ait İstatistikler	96
Tablo 4.5: Katılımcıların Yanıtladıkları Madde Sayılarının Frekans Dağılımları....	102

ŞEKİLLER DİZİNİ

Şekil 1.1: BBT Uygulamalarının Akış Şeması.....	4
Şekil 1.2: 2PL Modele Göre Kalibre Edilmiş Bir Madde İçin Elde Edilen Madde Karakteristik Eğrisi.....	15
Şekil 1.3: 3PL Modele Göre Kalibre Edilmiş Bir Madde İçin Elde Edilen Madde Karakteristik Eğrisi	16
Şekil 1.4: Tek boyutlu MTK'ya Göre Kalibre Edilmiş Bir Maddeye Ait Grafikler	17
Şekil 1.5: 2PL Çok Boyutlu MTK'ya Göre Kalibre Edilmiş Bir Maddeye Ait Grafikler	22
Şekil 3.1: Araştırma Kapsamında Kullanılan Faktör Analitik Modellerin Hipotetik Gösterimi.....	54
Şekil 3.2: Gerçek-Zamanlı Çok boyutlu BBT Uygulamasının Karşılama Ekranı	66
Şekil 3.3: Bireylerin Demografik Bilgilerini Girdikleri Ekran Görüntüsü	66
Şekil 3.4: Maddenin Yanıtlanmadan Önceki Ekran Görüntüsü	67
Şekil 3.5: Maddeye Yanıt Verildikten Sonraki Ekran Görüntüsü	67
Şekil 3.6: Testin Sonlanması Ardından Görülen Sonuç Ekranı	68
Şekil 3.7: Bireyin Veri Tabanına Yansıyan Ayrıntılı Sınav Sonucu	69
Şekil 4.1: Standart Hataya Dayalı Sonlandırma Kuralı için θ Değerleri Arasındaki Korelasyon Grafikleri	83
Şekil 4.2: Standart Hataya Dayalı Sonlandırma Kuralı için Yanıtlanan Madde Sayısı Grafiği.....	86
Şekil 4.3: θ Değişimine Dayalı Sonlandırma Kuralı için θ Değerleri Arasındaki Korelasyon Grafikleri	88
Şekil 4.4: : θ Değişimine Dayalı Sonlandırma Kuralı için Yanıtlanan Madde Sayısı Grafiği.....	91
Şekil 4.5: Sabit Madde Sayısına Dayalı Sonlandırma Kuralı için θ Değerleri Arasındaki Korelasyon Grafikleri	93
Şekil 4.6: θ_g için Elde Edilen Standart Hata – θ Grafikleri	98
Şekil 4.7: θ_g için Elde Edilen Yanıtlanan Madde Sayısı – θ Grafikleri.....	99
Şekil 4.8: Standart Hata Durdurma Kuralı, MAP ve D-rule'nin Kullanıldığı Çok Boyutlu BBT Koşulu için Yanıtlanan Madde Sayılarının Frekans Değerleri .	100
Şekil 4.9: θ Değişimine Dayalı Durdurma Kuralı, MAP ve D-rule'nin Kullanıldığı Koşul için Yanıtlanan Madde Sayılarının Frekans Değerleri	101
Şekil 4.10: Madde Havuzunda Gerçek-Zamanlı Uygulamada Yer Alan Maddelerin Kullanım Miktarları	103

SİMGELER VE KISALTMALAR DİZİNİ

KTK: Klasik Test Kuramı

MTK: Madde Tepki Kuramı

BBT: Bireyselleştirilmiş Bilgisayarlı Test

YTS: Yeterlilik Tespit Sınavı

YDS: Yabancı Dil Sınavı

CAT: Computerized Adaptive Testing



1. GİRİŞ

Ülkemizde; gerek ilköğretim ve ortaöğretim, gerekse yükseköğretim seviyesinde ulusal çapta bir çok sınav uygulanmaktadır. Ortaokul son sınıfta öğrenim gören öğrenciler, her dönemde bir kez olmak üzere toplam iki kez girdikleri TEOG (Temel Eğitimden Ortaöğretime Geçiş) sınavı puan ortalamaları ile, ortaöğretimde öğrenim görecekları lise için tercih yapmaktadırlar. Lise eğitimlerini tamamlayan öğrenciler ise eğitim görecekları yükseköğretim kurumunun belirlenmesi amacıyla gerçekleştirilen YGS (Yükseköğretime Geçiş Sınavı) ve LYS (Lisans Yerleştirme Sınavı)'ye girmek durumundadırlar. Lisans eğitimlerini tamamlayan bireylerden; kamuda görev yapmak isteyenler için KPSS (Kamu Personeli Seçme Sınavı), lisansüstü eğitime devam etmek ya da akademisyen olmak isteyenler içinse ALES (Akademik Lisansüstü Eğitim Sınavı) ile YDS (Yabancı Dil Sınavı) ya da YÖKDİL (Yükseköğretim Kurumu Dil Sınavı) uygulanmaktadır. Sözü geçen tüm ulusal sınavlar, öğrencilerin öğrenim görecekları bir üst eğitim kurumunu belirleme ya da istihdam etme amaçlı olduklarından, yüksek önem arz etmektedirler. Bu nedenle; söz konusu sınavların hatasız, güvenli ve amacına hizmet eden sınavlar olmaları son derece önemlidir. Yüksek önem arz eden söz konusu ulusal sınavların tamamı, Klasik Test Kuramı'nı (KTK) temel alır. Sadece başarı testleri değil, psikolojik yapıların ölçülmesi amacıyla geliştirilen testlerin büyük çoğunluğu da KTK'ya dayanmaktadır.

Geçtiğimiz yüzyılda KTK'nın yaygın olarak tercih edilmesinde, kuramsal varsayımlarının kolaylığının yanında (Hambleton & Swaminathan, 1985) II. Dünya Savaşı sırasında askerlerden hızlı ve düşük maliyetli bilgi edinme ihtiyacı da etkili olmuştur (Weiss, 2004). KTK'da, sabit sayıdaki madde setinin belirli bir gruba uygulanmasıyla elde edilen sonuçlar madde analizi için kullanılarak madde özellikleri (istatistikleri) elde edilir. Bu madde özelliklerden ilki madde güçlüğü, ikincisi ise madde ayırt ediciliğidir. Madde güçlüğü, maddeye verilen doğru yanıt sayısının maddeyi yanıtlayan birey sayısına bölünmesiyle elde edilen bir yüzdelik değerkendir, madde ayırt ediciliği ise maddeye verilen yanıtın, toplam ölçek puanıyla korelasyonundan elde edilir (Baykul, 1999). Ancak, KTK'nın uygulanması ve anlaşılmasının kolaylığı yanında bir çok sınırlılığı da vardır. Bunlardan ilki, madde istatistikleriyle ilgilidir. Belirli bir gruba uygulanan ölçeğin, aynı evrene ait başka bir

gruba uygulanmasıyla aynı maddeler için farklı istatistikler elde edilebilir. Örneğin; farklı gruplarda maddelerin yanıtlama oranları değişeceğinden aynı madde için farklı güçlük indeksleri elde edilebilirken, maddenin uygulandığı grubun homojenleşmesi, maddenin ayırt ediciliğini artıracaktır. Bir başka ifadeyle KTK'da madde özellikleri grup bağımlıdır. Bunun yanında KTK'nın kullanıldığı bir test geliştirme sürecinde, yüksek ayırt edicilik ve orta güçlükteki maddeler seçilir (Gelbal, 2013). Bu durum, iç tutarlılık anlamındaki güvenirlik değerini artırırken, test bilgi fonksiyonunun ölçeğin orta noktasında maksimum değeri aldığı sivri bir dağılımın elde edilmesine neden olur. Böylece, ortalama yetenek düzeyindeki bireyler için hatası daha düşük ölçme sonuçları elde edilirken, yetenek dağılımının iki ucuna gittikçe standart hata değeri artarak uçlarda yer alan bireylerin kestirimlerine büyük miktarda hata karışmasına neden olur (Hambleton & Swaminathan, 1985; Weiss, 2004; 2011). Kuramsal alt yapısıyla ilgili dezavantajlarının yanında, psikometri literatürüne hakim olmayan bireylerce de anlaşılabilir olması ve uygulamanın kolaylığı nedeniyle, özellikle ulusal sınavlarda KTK uygulamaları halen geniş yer tutmaktadır. 1950'li yıllardan itibaren KTK'nın söz konusu kuramsal sınırlılıklarına çözüm getirme iddiasıyla ortaya çıkan Madde Tepki Kuramı (MTK) uygulamaları ise, günümüzde bazı uluslararası sınavlarda (TOEFL, GRE vb.) uygulanmaktadır. MTK'ya dayalı uygulamalar, özellikle bilgisayar temelli sınavların bireye uygulanmasında büyük avantajlar sağlamaktadır.

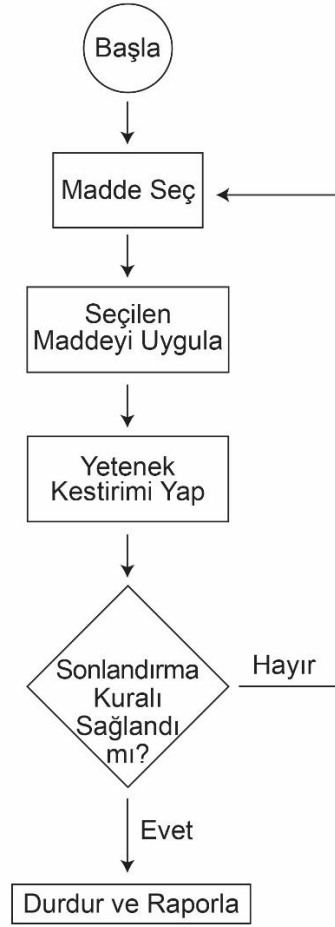
MTK, test tarafından ölçülmek istenilen özellik ile bireyin bu maddeyi doğru yanıtlama olasılığı arasındaki ilişkiyi matematiksel bir formülle ifade etmektedir. Bu kurama göre, geliştirilen testlerden elde edilen yetenek parametreleri bireylere uygulanan testlerden bağımsız olarak elde edilebilir. Bu özellik, test puanlarının eşitlenmesi durumunda, bireylerin yeteneklerinin gruptan bağımsız olarak karşılaştırılmasını sağlamaktadır (Kelecioğlu, 2001). MTK'ya göre madde ve yetenek parametreleri gruptan bağımsız olduğu gibi, standart hata da her bir bireyin kestirilen yetenek düzeyi için elde edilebilmektedir. Ancak MTK'ya göre ölçekleme yapılabilmesi için, d eğişmezlik özelliğinin yanı sıra tek boyutluluk ve buna bağlı olarak ortaya çıkan yerel bağımsızlığın da sağlanması gerekmektedir. Bu varsayımların karşılanması durumunda ise, hangi MTK modelinin kullanılacağına karar verilir. Bunun için, madde setinin en iyi uyum sağladığı model seçilir (1PL, 2PL ya da 3PL). Her ne kadar MTK tek boyutluluk varsayımı üzerine kurulmuş olsa

da ölçeklerin tek bir baskın gizil değişkeni ölçtüğünü kabul etmek pratikte psikolojik yapıların çok boyutlu doğasına aykırı bir durumdur (Reise, Morizot & Hays, 2007). Öyle ki, psikolojik yapıları ölçmek amacıyla geliştirilen birçok ölçek, tek boyutlu bir yapıdan ziyade çok boyutlu gizil değişkenleri ölçer (Seo & Weiss, 2015). Bu nedenle, tek boyutlu MTK'nın genişletilmesiyle, çok boyutlu MTK ortaya çıkmıştır (Bock & Aitkin, 1981).

MTK ile ilgili çalışmalar 1930'lu yıllara dayanmasına rağmen uzun bir süre kuramla ilgili yapılan çalışma sayısının oldukça az olduğu görülmektedir. Her ne kadar 1950'li yıllarda Lord tarafından bazı çalışmalar (1952, 1953a, 1953b) yapılmış olsa da gerektirdiği matematiksel alt yapı nedeniyle 1960'ların sonuna kadar kuramın gelişimi durağandır. Bu tarihten itibaren özellikle KTK'nın doğasından kaynaklanan zayıflıkları nedeniyle MTK'ya ağırlık verilmiştir. 1970'lerde MTK üzerine yapılan bilimsel çalışmaların baskın bir hal aldığı görülmektedir (Hambleton & Swaminathan, 1985). Bu tarihten itibaren, MTK'nın kuramsal gelişimine katkı sunan çalışmaların yanında bireylere uygulanan testlerin KTK'ya ve MTK'ya göre ölçeklenmesinden elde edilen yetenek kestirimlerinin farklılaşmasına yönelik çalışmalar da yapılmıştır. Bu çalışmalardan elde edilen sonuçlar, gerek tek boyutlu gerekse çok boyutlu MTK modellerinin kullanılmasıyla elde edilen yetenek kestirimleri ile KTK'nın kullanılmasıyla elde edilen yetenek kestirimleri arasında yüksek korelasyon olduğunu göstermektedir (Gelbal, 1994; Fan, 1998; Progar & Sočan, 2008; Zaman & diğ., 2008; Çelen & Aybek, 2013; Akyıldız ve Şahin, 2015). Bu durum, kağıt-kalem testlerinin MTK'ya göre ölçeklenmesinin sahip olduğu karmaşık matematiksel altyapı nedeniyle ne derece gerekli olduğu sorusunu akla getirmektedir. Bazı psikometristler (Weiss, 1985; Wainer ve ark., 2000; Ware ve ark., 2003), MTK'nın varoluş amacının CAT (Computerized Adaptive Testing)/Bireyselleştirilmiş Bilgisayarlı Test (BBT) uygulamaları olduğu görüşündedirler (akt; Kezer, 2011).

Bireyselleştirilmiş Bilgisayarlı Test algoritması; önceden kalibre edilmiş bir madde havuzunu kullanarak, yanıtladığı her madde sonrasında bireyin kestirilen geçici yetenek düzeyi baz alınıp, havuzdan bireyin yetenek düzeyine en uygun maddenin seçilerek uygulanması ve yine önceden belirlenen bir kurala göre testin sonlandırılmasına dayanır (Frey, 2009; Thompson & Weiss, 2011; Bulut & Kan,

2012). Aşağıda, BBT uygulamasının gerçekleşme biçimine dair bir şema yer almaktadır.



Şekil 1.1: BBT Uygulamalarının Akış Şeması

BBT uygulamalarında; bireyin yanıtlaması için havuzdan seçilen her madde bireyin yetenek düzeyine uygun olacağından birey, testle ölçülmek istenen yetenek düzeyinin çok üstünde ve çok altında olan maddelere maruz kalmayacak ve bu sayede ölçme sonuçlarına karışan tesadüfi hata miktarı da azalacaktır. Bireyin sadece kendi yetenek düzeyine uygun maddeleri yanıtlamasıyla; grubun tamamına hitap eden standart bir testle aynı ölçme kesinliği, çok daha az madde uygulanarak elde edilmiş olur (Segall, 2005; Weiss, 2011). BBT uygulamalarında bireye düzeyine uygun maddenin yöneltilebilmesi, MTK'da bireyin yetenek düzeyi ile maddelerin güçlüğü'nün aynı ölçek üzerine yer almasına dayanır (Reckase, 2009). Yapılan çalışmalar, BBT uygulamalarının kağıt-kalem formatındaki geleneksel testlerle en az aynı düzeyde ölçme kesinliğini, ortalama %50 oranında daha az maddeyle sağladığını göstermektedir. (Segall, 1996; Luecht, 1996; Eggen, 2007; Weiss & Gibbons, 2007; Gibbons ve diğ., 2008; Weiss, 1982, 1985, 2011). BBT üzerine

yapılan bilimsel alıřmalar bilgisayar programlarının geliřmesiyle son 20 yılda hız kazansa ve popülerlięi her geen gün artsa da bireyselleřtirilmiř testlerin tarihi ok daha eskilere dayanmaktadır.

1.1. Tek Boyutlu ve ok Boyutlu Bireyselleřtirilmiř Bilgisayarlı Testlerin Tarihi

Bireyselleřtirilmiř testlerin ilk rneęi, Binet & Simon (1905) tarafından geliřtirilen “Binet Zeka Testi”dir. Binet, 5-15 yař aralıęında tam ve ara yařlardan oluřan toplam 21 zihinsel yař aralıęı belirlemiřtir. Her bir zihinsel yař dzeyinde 10 madde olmak zere, madde havuzu toplam 210 maddeden oluřmaktadır. Bir maddenin yer alacaęı zihinsel yař kategorisi belirlenirken; gerekte ilgili yařta olan bireylere sz konusu madde uygulanarak, en az %50’sinin yanıtlaması durumunda maddenin ilgili zihinsel yař soru havuzunda bulunmasına karar verilir. Test, uygulama konusunda eęitim almıř bir uzman tarafından uygulanır. Testi alan bireyler hakkındaki nsel (a priori) bilgiye gre bireyin ilk adımda yanıtlayacaęı 10 madde belirlenir. Bireyin bu maddeleri yanıtlama oranına gre bir sonraki adımda bir st ya da alttaki zihinsel yař soru grubunu yanıtlaması saęlanır. Eęer birey ilk adımdaki 10 maddenin bir kısmını doęru yanıtlamıřsa bir sonraki adımda, bir alttaki zihinsel yařa ait 10 maddeyi yanıtlar. Bu durum, bireyin maddelerinin tamamını yanıtlayabildięi zihinsel yař bulunana kadar devam eder ve bu řekilde alt dzey (basal level) belirlenmiř olur. Eęer birey ilk basamakta maddelerin tamamını doęru yanıtlamıřsa, bundan sonra st dzeyin belirlenmesi ařamasına geilir. Bunun iin bireye, bařlangıta tamamını doęru yanıtladıęı zihinsel yařın bir zerindeki zihinsel yařa ait madde grubu verilir. Eęer birey, maddelerin en az birini doęru yanıtlarsa, bir st zihinsel yařa ait madde grubuna geer. Bu durum, bireyin maddelerinin hi birini yanıtlayamadıęı zihinsel yař belirleninceye kadar devam eder ve bu řekilde birey iin st dzey (ceiling level) belirlenir. Zeka (IQ) puanının belirlenmesi iinse, bireyin doęru yanıtlarının ait olduęu zihinsel yař grubuna gre aęırlıklandırılarak ortalaması alınır. Bu deęer, bireyin gerek yařına blndkten sonra son olarak yz ile arpılarak bireyin IQ dzeyi belirlenir. Bu řekilde bireyin soru havuzundaki 210 maddenin tamamını yanıtlamasına gerek kalmadan zeka dzeyi belirlenmiř olur. Bununla beraber birey, her bir zihinsel yařa ait 10 maddenin tamamını yanıtlamak durumunda olduęundan, bu testin kısmi olarak bireyselleřtirilmiř olduęu sylenebilir. Uygulanması iin eęitim almıř bir uzmana ihtiya olması, yanıtlanan

her bir madde grubundan sonraki maddelerin belirlenerek hazırlanmasının manuel yapılmasının ciddi bir zaman kaybına neden olması gibi nedenlerle bu yöntem verimli bulunmamış ve süreklilik göstermemiştir. Sonraki dönemde; özellikle dünya savaşlarının etkisiyle hızlı, ucuz ve kolay uygulanabilir yöntemler psikometri dünyasında ağırlık kazanmıştır (Weiss, 1985, 2004, 2011).

Her ne kadar KTK'ya dayalı BBT çalışmaları yapılmış olsa da (Frick, 1992; Rudner, 2002, Rudner & Guo, 2011), modern BBT çalışmaları MTK üzerine kurulmuştur. Dünya savaşları sonrasında geleneksel kağıt-kalem testlerinin yaygınlaşması, 1950'li yıllarda gelişmeye başlayan MTK'nın karmaşık yapısı ve MTK temelli BBT algoritmalarının o dönemki bilgisayar teknolojisi için oldukça karmaşık olması vb. nedenlerle modern BBT çalışmaları 1970'li yılların ortasına kadar oldukça yavaş ilerlemiştir.

60'lı yılların ortasından itibaren bilgisayar teknolojisinin anlık yetenek kestirimindeki yavaşlığı nedeniyle erken dönem çalışmaları geleneksel yöntemler etrafında şekillenmiştir. Bunlar arasında İki Basamaklı Testler (Cronbach & Gleser, 1965), Yetenek Parametrelerinin Sonsal Dağılımına Dayalı Madde Seçimi (Owen, 1969), Yaklaşık (up-and-down) Madde Seçimi (Lord, 1970), Robbins-Monro Algoritması (Lord, 1971a), Esnek-Basamak Testler (Lord, 1971b), "Stradaptive" (Larkin & Weiss, 1975) yer alır. Bunun yanında; Birleşik Devletler Deniz Kuvvetlerinin 1960'larda başlayan desteği sonrasında BBT araştırmalarına, Savunma Bakanlığının Amerikan Ordusuna asker alımı için gerçekleştirdiği Silahlı Kuvvetler Mesleki Yetenek Testinin (Armed Services Vocational Aptitude Battery-ASVAB) BBT haline getirilmesi için yapılan çalışmalarla devam etmiştir. 1980'lerle beraber BBT çalışmaları hız kazanmış, bazı düşük önem arz eden sınavlar BBT olarak (Sands, 1985; Ward, 1988; Lunz & Bergstom, 1991) uygulanmıştır. Ancak, 1992 yılında Amerikan Silahlı Kuvvetler Mesleki Yetenek Testinin BBT olarak uygulanması (Moreno & Segall, 1992) ve 1993 yılında Amerikan Devlet Hemşirelik Ulusal Heyetinin lisanslama sınavında BBT uygulamalarına başlaması ile kağıt-kalem testlerinden BBT'ye geçiş için asıl büyük adımlar atılmıştır. Bu sınavları, aynı yıl içinde Educational Testing Service (ETS) tarafından gerçekleştirilen Yüksek Lisans Kayıt Sınavının (Graduate Record Examination-GRE) BBT olarak uygulanması izlemiştir. BBT'lerin yüksek önem arz eden geniş ölçekli sınavlarda kullanılmaya başlamasıyla madde teşhir oranı (item exposure rate), içerik

ağırlıklandırılması (content balancing) vb. bazı operasyonel konular üzerine bilimsel çalışmalar yapılması gerekliliği ortaya çıkmıştır (Mills & Stocking, 1996; Sands, Waters & McBride, 1997; van der Linden & Glas, 2002; Thompson & Weiss, 2011).

BBT uygulamaları üzerine yapılan çalışmaların büyük bölümü, tek boyutlu MTK'ya göre geliştirilmiştir (Hambleton & Swaminathan, 1985; Fischer & Molenaar, 1995). Bununla birlikte bilgisayar teknolojisindeki gelişmeler, çok boyutlu BBT çalışmalarına olan ilgiyi artırmaktadır (Reckase, 2009). Çok boyutlu BBT çalışmaları, Bloxom & Vale'nin (1987) çalışmasıyla başlamıştır. Bu çalışmada araştırmacılar, Owen'in (1975) ardışık Bayes güncelleme yöntemini çok değişkenli hale getirerek yeni ve etkili bir yetenek kestirim yöntemi ortaya koymuşlardır. Bu çalışmayı, iki boyutlu normal-ogive model için iteratif bir Maximum Likelihood (ML) yetenek kestirim yöntemi öneren Tam (1992) izlemiştir. Ancak bu iki çalışma yetenek kestirim düzeyinde kalmış, çok boyutlu BBT uygulamaları için gerçek-zamanlı kestirimlerde kullanılabilecek bir madde seçim yöntemi önerilmemiştir. Bu nedenle bu çalışmalar, çok boyutlu yöntemlerin tek boyutlu BBT uygulamaları karşısındaki avantajlarını ortaya koymaktan uzak kalmışlardır.

Çok boyutlu BBT çalışmaları bağlamında Segall (1996) tarafından yapılan araştırma bir dönüm noktasıdır. Bu çalışmaya kadar, çok boyutlu bir BBT uygulamasının geliştirilebilecek olmasına rağmen bu çalışmayı yapmaya motive edecek geçerli bir neden olmadığı düşünülmektedir (Wainer ve diğerleri, 1990). Segall'e göre; çok boyutlu doğasına rağmen baskın tek boyutluluk baz alınarak geliştirilen BBT uygulamaları, farklı alt boyutlarda benzer sayıda madde olmasını sağlamak için içerik ağırlıklandırmasını kullanırlar. Fakat, burada bazı boyutların doğasından dolayı güçlüğü daha yüksek maddeleri içermesi, ölçme işleminin veriminin düşmesine neden olur. Bunun yerine çok boyutluluk esas alınarak yapılacak BBT uygulamalarının; birbiriyle pozitif yönde ilişkili boyutlardaki maddelerden birine verilen yanıtla, diğer boyutla ilgili de bilgiyi ve buna bağlı olarak ölçme kesinliğini artırması beklenir.

Segall (1996) bu çalışmasında; çok boyutlu BBT uygulamaları için yetenek kestirim yöntemlerinin yanında Bayes madde seçim yöntemleri de geliştirerek çok boyutlu BBT uygulamalarının tek boyutlu uygulamalarla aynı ölçme kesinliğini üçte bir oranında daha az maddeyle yapabildiğini ortaya koymuştur. Aynı yıl içinde Luecht tarafından yapılan çalışma da benzer sonuçları ortaya koyar niteliktedir. Çok boyutlu

BBT uygulamaları üzerinde kritik önem taşıyan bu çalışmaları, van der Linden'in (1999) yaptığı çalışma izlemiştir. Bu çalışma, farklı yeteneklerin lineer kombinasyonlarına ait hata varyanslarının bir madde seçim algoritması olarak kullanılabilirliği üzerinedir. Segall (2000), çalışmasında yine çok boyutlu BBT uygulamalarının verimi üzerinde durmuş, bu uygulamaların tek boyutlu uygulamalara göre daha efektif madde seçimi sağladığı ve yetenek kestirimindeki etkililiğini artırdığını raporlamıştır. Li & Schafer (2005), madde teşhir oranını kullanarak tek ve çok boyutlu BBT uygulamaları karşılaştırmış, çok boyutlu uygulamaların yetenek dağılımının iki ucunda ölçme kesinliğini artırdığı sonucuna varmışlardır. 2009 yılında Mulder ile van der Linden, farklı madde seçim yöntemlerinin etkililiğini araştırmışlardır.

2010'lu yıllarda ise çok boyutlu BBT uygulamaları üzerine yapılan çalışmalar büyük bir ivme kazanmıştır. Özellikle açık kaynak kodlu bir istatistik programı olan R'nin kazandığı popülerite ile birlikte, R'de tanımlı bir çok BBT paketi ortaya çıkmıştır. Bu paketler arasında gerçek-zamanlı çok boyutlu bir uygulamaya imkan veren yegane paket ise mirtCAT (Chalmers, 2016) paketidir.

1.2. Problem Durumu

Alan yazında; çok boyutlu BBT uygulamalarının tek boyutlu BBT'lere kıyasla ölçme kesinliğini artırdığına yönelik çalışmalar sonrası, çok boyutlu BBT uygulamalarının etkililiğini artırmaya yönelik çalışmalar ağırlık kazanmıştır. Geçtiğimiz on yılda, çok boyutlu BBT uygulamaları için farklı madde seçim, sonlandırma, içerik ağırlıklandırılması vb. yöntemlerinin geliştirilmesi üzerine bir çok çalışma yapılmıştır (Choi, Grady ve Dodd, 2010; Yao, 2012, 2013, 2014; Wang, Chang & Boughton, 2012; Yao, Pommerich & Segall, 2014; Su, 2016). Söz konusu çalışmaların ağırlıklı olarak madde içi ya da maddeler arası boyutluluğu baz alan çalışmalar oldukları görülmektedir. Bu çalışmaların yanında, çok boyutluluğu göz ardı etmeden maddelerin ya da yapıyı oluşturan boyutların (alt faktörlerin) altında yatan ortak varyans kaynağını da ele alarak genel bir yetenek kestirimi yapan çok boyutlu BBT çalışmalarının sayısının ise oldukça az olduğu görülmektedir (Weiss & Gibbons, 2007; Seo, 2012; Huang, Chen & Wang, 2012; Seo & Weiss, 2015). Bu araştırmalar arasında yapılan tek gerçek-zamanlı uygulamanın (Weiss & Gibbons, 2007), aslında her bir boyutun ayrı ayrı kestirildiği bir uygulamaya dayandığı, başka bir ifadeyle tek boyutlu olduğu bilinmektedir. Diğer çalışmalar ise çok boyutluluğa

dayanmakla birlikte, tamamı *monte carlo* simülasyon çalışmalarıdır. Buradan yola çıkarak, söz konusu çok boyutlu yapılar baz alınarak gerçek verilere dayalı bir *post-hoc* simülasyon ya da gerçek-zamanlı BBT uygulamasının henüz yapılmamış olduğu görülmektedir. Bunun yanında, ülkemizde yapılan sınırlı sayıdaki BBT çalışmasının neredeyse tamamının tek boyutluluk varsayımı ile yapılmış çalışmalar olduğu, çok boyutlu BBT uygulamaları üzerine yapılan sınırlı sayıdaki çalışmanın ise gerçek veri setleri üzerinde yürütülen *post-hoc* simülasyon çalışmaları olduğu bilinmektedir.

Türkiye’de, bireylerin İngilizce seviyesinin belirlenmesi üzerine bir çok sınav yapılmaktadır. Bunlar arasında; ÖSYM tarafında yürütülen YDS (Yabancı Dil Sınavı) ile Ankara Üniversitesi ve Anadolu Üniversitesi ortak organizasyonu ile ilk uygulaması Mart 2017’de yapılan YÖKDİL (Yükseköğretim Kurumları Dil Sınavı) ön plana çıkmaktadır. Söz konusu sınavların her birine, sayıları yaklaşık 100-250 bin arasında değişen öğrenci girmektedir. Bunlardan YDS için kağıt-kalem formatındaki sınavın online olarak da uygulanabilme imkanı bulunmakta olup, bireyselleştirilmiş bir sınav uygulamasına henüz geçilememiştir. Bireylerin akademik yeterlilikleri ya da kamuda görev yapan personelin İngilizce yeterliliklerini belirlemeye yönelik olarak uygulanan bu sınavların yanında diğer önemli bir İngilizce sınavı ise, üniversitede öğrenim görme hakkı kazanan öğrencilerin İngilizce yeterliliklerini ortaya koyarak lisans eğitimlerine başlayabilmek için üniversitelerce düzenlenen yeterlilik sınavlarıdır. Söz konusu sınavlar da, ülkemizdeki diğer İngilizce sınavları gibi kağıt-kalem formatında ya da bu formatın online olarak uygulanması biçimindedir. Bu bilgilerden yola çıkarak, ülkemizde farklı amaçlarla İngilizce yeterliliğini belirlemeye yönelik olarak uygulanan tüm bu sınavların KTK’yı temel alan geleneksel sınavlar olduğu görülmektedir. İngilizce yeterliliğinin tespitine yönelik olarak uygulanan tüm bu sınavların çok boyutlu BBT olarak uygulanabilmesi için, gerçek-zamanlı çok boyutlu BBT çalışmalarının yapılması ve bu çalışma sonuçlarının tüm paydaşlarca tartışılmasına ihtiyaç duyulmaktadır.

1.3. Araştırmanın Amacı ve Önemi

Bu çalışmanın amacı, Anadolu Üniversitesi YDYO tarafından hazırlık eğitimi sonunda gerçekleştirilen İngilizce yeterlilik sınavı verilerini kullanarak, gerçek-zamanlı çok boyutlu bir BBT uygulaması yapmaktır. Bu amaçla; öncelikle gerçek veri setleri kullanılarak farklı yetenek kestirim yöntemi, madde seçim yöntemi ve

sonlandırma kuralları ile oluşturulan 36 farklı koşul arasından gerçek-zamanlı uygulama için en uygun olanı belirlenmiştir. Söz konusu koşul kullanılarak bireylere gerçek-zamanlı uygulama yapılması sonucunda elde edilen yetenek kestirimleri ile bu bireylere kağıt-kalem testi uygulanmasıyla elde edilen toplam puanları arasındaki korelasyon hesaplanmıştır.

Ülkemizdeki artan genç nüfus gerek eğitime, gerekse sınırlı sayıdaki istihdama olan talebi her geçen gün artırmaktadır. Bu amaçla gerçekleştirilen ölçme ve değerlendirme faaliyetlerinin tamamı kağıt-kalem testinin kullanıldığı ve gün geçtikçe gücünü artıran geleneksel sınavlardır. Bu durum da her geçen gün ölçme sonuçlarına karışan hatanın artması sorununu beraberinde getirmektedir. Zira kağıt-kalem testleriyle puan dağılımının ortalamasından uzak kalan bireylerin yetenek kestirimine daha fazla hata karışmasının engellenemediği gibi, birbirinden çok farklı yetenek düzeyinde olan bireylere aynı madde setinin kullanılması da bireylerin yetenek düzeylerine ait olmayan çok fazla soruyla muhatap olmalarına neden olarak tesadüfi hata miktarını artırmaktadır. Bu nedenle ülkemizde; çok daha az maddeyle (dolayısıyla çok daha kısa zamanda) KTK'ya dayalı geleneksel sınavlardan daha güvenilir sonuçların elde edilebildiği, farklı sene/dönemlerde sınavı alan öğrencilerin aynı ölçekte değerlendirilebileceği BBT uygulamalarının yaygınlaşmasına ihtiyaç vardır. Ayrıca BBT uygulamaları için geniş bir madde havuzu gerekli olup, bireyin bu sorulardan hangisini yanıtlayacağına sınav esnasındaki performansına göre karar verildiğinden; sınav öncesinde bireyin yanıtlayacağı sorulara ulaşmasının mümkün olmadığı, dolayısıyla bu sınavların geleneksel sınavlara göre çok daha güvenli olduğu da unutulmamalıdır. Getirdiği bu avantajların yanında BBT uygulamalarında gerek bireylere farklı sorular sorulmasının, gerekse yanıtlanacak madde sayısının farklı olmasının sınavı alan bireyler ve kamuoyunda tartışmalara neden olabileceği düşünülebilir. Bununla beraber bireylere yöneltilen soruların ve soru sayılarının farklı olmasının nedeni MTK'nın doğasıdır. Söz konusu farklılıkların bireylerin yetenek kestirimine herhangi bir etkisi bulunmamaktadır.

BBT uygulamalarına duyulan ihtiyaç alanlarının birisinin de üniversitelerin İngilizce hazırlık sınıflarının sonunda yapılan yeterlilik sınavları olduğu söylenebilir. Bu sınavlarla kestirilen İngilizce becerisi, farklı alt boyutlardan oluşan çok boyutlu bir yapıdır. Dolayısıyla İngilizce becerisinin ölçülmesinde kullanılacak bireyselleştirilmiş testlerin, bu beceri kestirilken çok boyutlu karmaşık yapıyı da göz ardı etmemeleri

beklenir. Bu durum göz önüne alındığında bu çalışma ile gerçek-zamanlı çok boyutlu bir BBT'nin uygulanabilirliğinin kanıtlanmasının, farklı alanlarda öğrencilerin yeteneklerinin kestirilmesi amacıyla kullanılan karmaşık yapıların ölçülmesinde de çok boyutlu BBT uygulamalarının önünü açacağı düşünülmektedir.

Çok boyutlu BBT uygulamalarında temel olarak maddeler arası ya da madde içi boyutluluk modellerini kullanan çok boyutlu MTK modelleri kullanılmaktadır. Maddeler arası boyutluluk modeli ile her bir maddenin yalnızca bir boyutu ölçtüğü kabul edilir. Ancak bu durum, psikolojik yapıların doğası göz önüne alındığında gerçekçi değildir (Seo, 2012). Madde içi boyutluluk modelinde ise her madde tüm boyutlara yük verir. Ancak bu modellerde de boyutların tanımlanabilirlik problemi söz konusudur (Li & Schafer, 2005). Bu araştırma kapsamında kullanılan bifaktör model ise, genel bir faktör/yetenek (general trait) içerdiği öngörülen ilişkili yapılar için bir çözüm sunmaktadır (Gustafson & Balke, 1993). Bu anlamda yapılan çalışmanın, çok boyutluluğu göz ardı etmeden genel bir yetenek kestirimi yapacak olmasının İngilizce yeterliliğinin doğasına da uygun olduğu söylenebilir. Bugüne kadar bifaktör modelle yapılan çok boyutlu BBT uygulamalarının üretilmiş veri setleri üzerinde yapılan *monte carlo* simülasyon çalışmaları olduğu göz önüne alındığında; gerek simülasyon çalışmasının gerçek veri kullanılarak yapılması, gerekse gerçek-zamanlı bir uygulama yapılmış olması nedeniyle bu çalışmanın literatüre de önemli bir katkı sunacağı düşünülmektedir.

1.4. Problem Cümlesi

Çok boyutlu bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı madde seçim kuralları, yetenek kestirim yöntemleri ve sonlandırma kuralları kullanılarak elde edilen gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanıtlanan ortalama madde sayısı, yanlışlık (bias) ve RMSD değerleri farklılık göstermekte midir? Gerçek-zamanlı uygulama sonuçlarına ait özellikler kağıt-kalem testleri ile ne düzeyde bir benzerlik göstermektedir?

1.4.1. Alt Problemler

1. Standart hataya dayalı sonlandırma kuralının kullanıldığı çok boyutlu BBT uygulamalarında farklı yetenek kestirim yöntemleri (EAP ve MAP) ve madde seçim kurallarına (D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule) göre gerçek ve kestirilen yetenek düzeyleri arasındaki

korelasyon, yanıtlanan ortalama madde sayısı, yanlılık (bias) ve RMSD değerleri nasıldır?

2. θ değişimine dayalı sonlandırma kuralının kullanıldığı çok boyutlu BBT uygulamalarında farklı yetenek kestirim yöntemleri (EAP ve MAP) ve madde seçim kurallarına (D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule) göre gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanıtlanan ortalama madde sayısı, yanlılık (bias) ve RMSD değerleri nasıldır?
3. Sabit madde sayısına dayalı sonlandırma kuralının kullanıldığı çok boyutlu BBT uygulamalarında farklı yetenek kestirim yöntemleri (EAP ve MAP) ve madde seçim kurallarına (D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule) göre gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanıtlanan ortalama madde sayısı, yanlılık (bias) ve RMSD değerleri nasıldır?
4. Simülasyon sonuçlarına göre her bir sonlandırma kuralı için en uygun çok boyutlu BBT koşulu nedir?
 - a. Standart hataya dayalı sonlandırma kuralı için en uygun yetenek kestirim ve madde seçim yöntemi hangisidir?
 - b. θ değişimine dayalı sonlandırma kuralı için en uygun yetenek kestirim ve madde seçim yöntemi hangisidir?
 - c. Sabit madde sayısına dayalı sonlandırma kuralı için en uygun yetenek kestirim ve madde seçim yöntemi hangisidir?
 - d. Simülasyon sonuçlarına göre gerçek-zamanlı çok boyutlu BBT uygulaması için en uygun koşul hangisidir?
5. Gerçek-zamanlı çok boyutlu BBT uygulaması sonuçları nasıldır?
 - a. Gerçek-zamanlı çok boyutlu BBT uygulaması sonucunda bireylerin yanıtladıkları madde sayılarının dağılımı nasıldır?
 - b. Gerçek-zamanlı çok boyutlu BBT uygulaması sonucunda madde havuzunda yer alan maddelerin kullanım oranları nasıldır?
 - c. Bireylerin, gerçek-zamanlı çok boyutlu BBT uygulamasıyla elde edilen genel yetenek yetenek düzeyleri ile kağıt-kalem testi sonucu elde edilen toplam puanları arasında anlamlı bir ilişki var mıdır?

1.5. Sayıtlılar

Bu çalışmada, veri setlerine en iyi uyum gösteren model bifaktör modeldir. Bifaktör modelde, grup faktörleri birbirinden ve genel yetenekten bağımsızdırlar. Bu nedenle, yeteneklerin dağılımıyla ilgili önsel dağılımın bilinmesine gerek yoktur.

1.6. Sınırlılıklar

Çalışma kapsamında, gerçek veri setleri üzerinde yapılan simülasyon çalışması ve gerçek-zamanlı çok boyutlu BBT uygulaması için R’de tanımlı mirtCAT (Chalmers, 2016) paketi kullanılmıştır. Bu nedenle uygulama sırasında kullanılan çok boyutlu BBT koşulları, paketin kullanıma sunduğu koşullarla sınırlıdır.

1.7. Tanımlar

Alt problemlerin çözümlenmesinde kullanılacak olan değerlerin tanımları aşağıda görülmektedir:

RMSD (Root Mean Squared Difference): Kestirimin doğruluğuna (hatasızlığına) dair bir değerdir. Gerçek ve kestirilen θ değerlerinin farklarının karelerinin ortalamasının kare köküdür. Bu değer küçülmesi, daha hatasız kestirim yapıldığı anlamına gelmektedir.

Bifaktör Model: Çok boyutlu yapılarda; ölçme aracıyla ölçülmesi amaçlanan boyutların (faktörlerin) yanında, tüm maddelerin altında yatan ortak varyans kaynağına dayanarak bir de genel yetenek/faktör kestiriminin yapıldığı doğrulayıcı faktör analitik modeldir. Her bir madde, ait olduğu boyutun yanında bir de genel faktöre yük verir.

Yanlılık (bias): Gerçek ve kestirilen yetenek ölçüleri arasındaki farkın ortalamasını veren bir değerdir. Bu değer, 0’a ne kadar yakın olursa kestirimin yanlılığının da o derece düşük olduğu sonucuna varılır.

Ortalama Madde Sayısı: Her bir koşul için, yetenek kestiriminin sonlandırılması sonrası bireylerin yanıtladığı madde sayısını (testin uzunluğunu) belirtir.

Madde Teşhir Oranı: Havuzda yer alan bir maddenin; kullanım sıklığına ilişkin, her bir uygulama sonrası güncellenen değerdir.

1.8. Araştırmanın Kuramsal Temeli

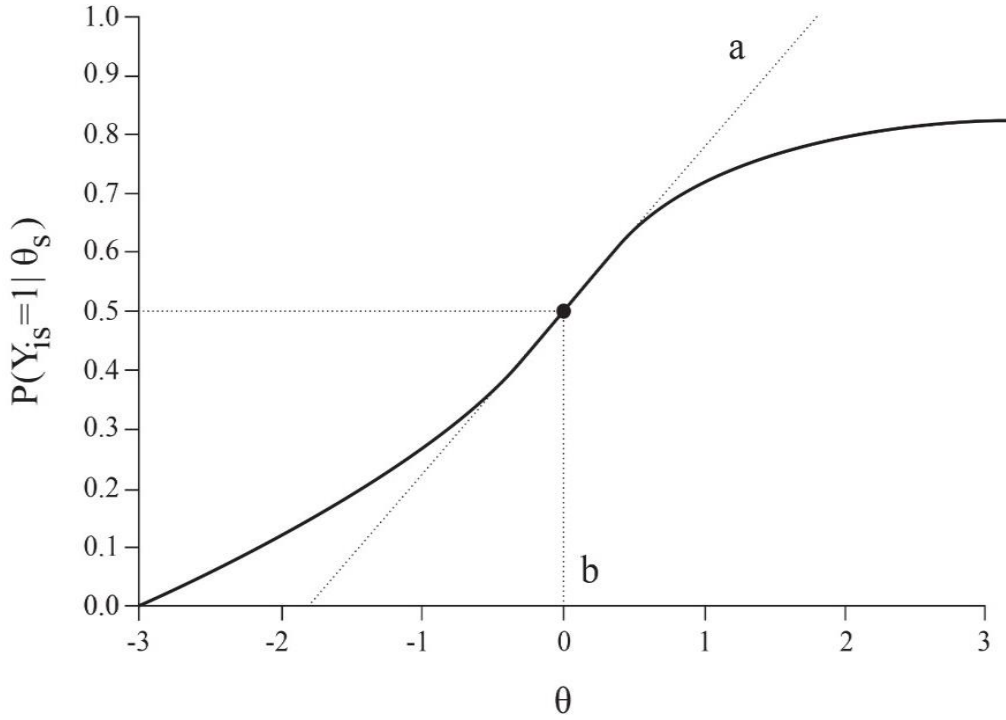
Araştırmanın bu bölümü iki kısımdan oluşmaktadır. İlk bölümde, çok boyutlu BBT uygulamalarının temelini oluşturan çok boyutlu MTK hakkında bilgi verilmiştir. Çok boyutlu MTK'ya geçmeden önce, tek boyutlu MTK modelleri tanıtılmıştır. İkinci bölümde ise, tek ve çok boyutlu BBT'lerin geliştirilme ve uygulama süreci üzerinde durulmuştur.

1.8.1 Madde Tepki Kuramı

1.8.1.1. Tek Boyutlu Madde Tepki Kuramı

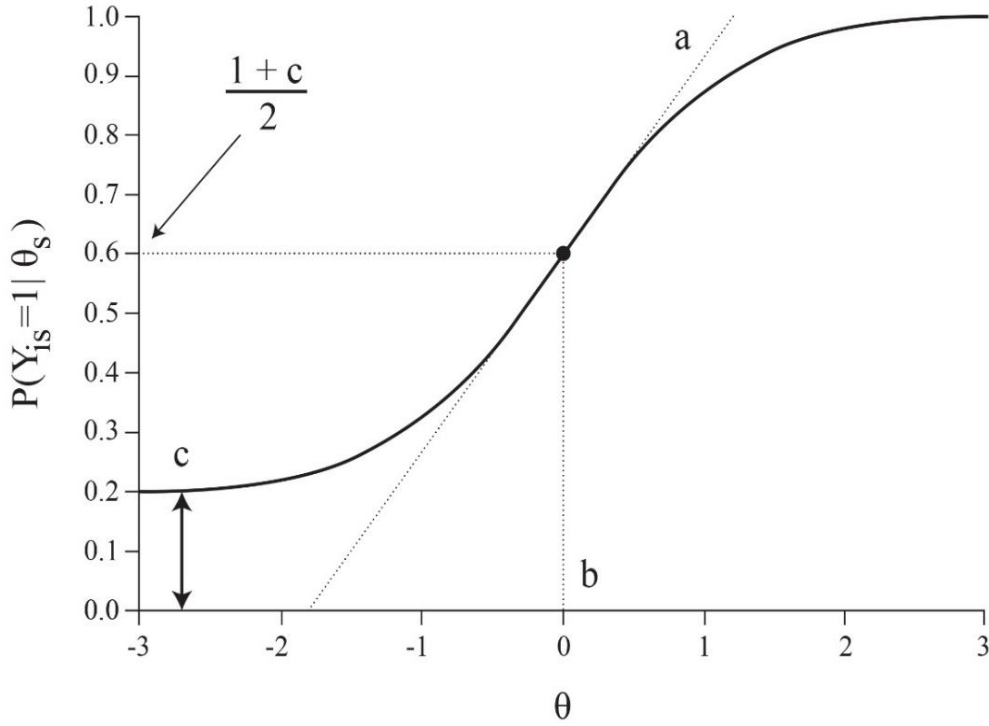
BBT uygulamalarında maddenin bireyin yetenek düzeyine göre belirlenmesini sağlamada, MTK'da maddenin güçlüğü ile bireyin yetenek düzeyinin aynı ölçek üzerinde yer alması kritik önem taşır. MTK, bireyin yetenek düzeyi ile maddeyi doğru yanıtlama ihtimali arasındaki ilişkiyi matematiksel bir formülle ifade eder. MTK uygulamaları başlangıçta tek boyutluluk üzerine kurulmuş, erken dönem çalışmaları tek boyutlu MTK modelleri üzerine yapılmıştır. Bu dönemde yapılan çalışmalarda veri setinde baskın bir boyutun var oluşu, tek boyutluluk varsayımının sağlanması için yeterli görülmüştür. Tek boyutlu MTK modellerinde maddelerin, iki (dichotomous) ya da çok kategorili (polytomous) olmak üzere iki gruba ayrıldıkları söylenebilir. Bu çalışma kapsamında kullanılan veriler çoktan seçmeli test maddelerine ait olduğundan, iki kategorilidirler. Bu nedenle, bu aşamadan sonra iki kategorili MTK modelleri üzerinde durulacaktır.

Her ne kadar MTK'ya dayalı çalışmalar 1930'lu yıllara kadar uzansa da, MTK'nın çoğunlukla Lord ve Novick (1968) tarafından kaleme alınan "*Statistical Theories of Mental Test Scores*" kitabıyla başladığı düşünülür. MTK'da iki temel varsayım yer alır. Bunlardan ilki; özel bir şekle sahip olan Madde Karakteristik Eğrisi (MKE), ikincisi ise Yerel Bağımsızlık Varsayımdır (Embretson & Reise, 2000). MKE, bireyin bir maddeyi doğru yanıtlama ihtimalini yeteneğin monoton-artan bir fonksiyonu olarak ortaya koyan "S" şeklinde bir grafikdir. Bu grafik, bireyin yeteneğindeki değişimin maddeyi yanıtlama olasılığına etkisini ortaya koyar. Aşağıda, 2PL modele uygun bir MKE örneği görülmektedir.



Şekil 1.2: 2PL Modele Göre Kalibre Edilmiş Bir Madde İçin Elde Edilen Madde Karakteristik Eğrisi

Yukarıdaki MKE, 2PL modele göre kalibre edilmiş bir maddeye aittir. Maddeyi doğru yanıtlama olasılığının .50 olduğu noktadan yetenek düzeyini temsil eden yatay eksene paralel bir izafi doğru çizildiğinde, bu doğrunun MKE'yi kestiği noktadan (MKE'nin büküm noktası) yatay eksene bir izdüşüm alınır. Bu izdüşümde karşılaşılan değer, maddenin güçlük parametresi (b) olarak tanımlanır. Daha sade bir ifadeyle madde güçlüğü, yetenek ölçeğinde MKE'nin en dik eğime sahip olduğu noktadır. Buna dayanarak, yukarıda MKE'si verilen maddenin güçlük değerinin "0" olduğu görülecektir. Maddenin güçlüğü arttıkça MKE'nin en dik eğime sahip olduğu nokta sağa, azaldıkça ise sola doğru hareket eder. Maddenin ayırt ediciliği ise, MKE'nin büküm noktasının yatay ekseni kestiği noktanın eğimidir. Bu nokta, maddenin doğru yanıtlanma ihtimalinin .50 olduğu noktadır. Maddenin ayırt ediciliği arttıkça eğim artarken, azaldıkça eğim de azalır. Madde ayırt edicilik parametresi, a ile sembolize edilir. Yukarıdaki grafiğe ait maddenin ayırt ediciliği 1'dir. Aşağıda, 3PL modele ait bir MKE yer almaktadır.



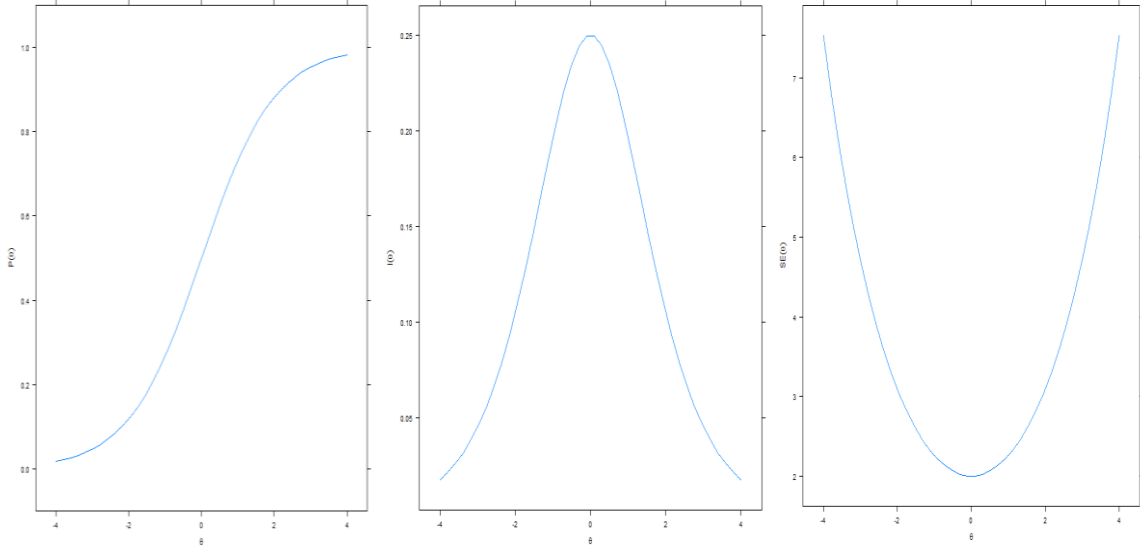
Şekil 1.3: 3PL Modele Göre Kalibre Edilmiş Bir Madde İçin Elde Edilen Madde Karakteristik Eğrisi

Yukarıda 3PL modele göre kalibre edilmiş bir maddeye ait MKE yer almaktadır. Burada, 2PL'den farklı olarak MKE'nin başladığı alt nokta, bir başka ifadeyle $P(\theta)$ dikey eksenini kestiği nokta maddenin şans parametresini verir. Bu parametre, madde ile kestirilen yetenek düzeyi en düşük olan bireylerin maddeyi yanıtlama olasılığını verir ve c ile sembolize edilir. Şans parametresine ait değerin artması, maddenin rastgele yanıtlandığında verilen cevabın doğru olma ihtimalinin artması anlamına gelir (Hambleton & Swaminathan, 1985; Embretson & Reise, 2000). MKE'si verilen bu maddenin a ve b parametreleri, bir önceki maddede olduğu gibi sırasıyla 1 ve 0'dır. Ancak 2PL modelden farklı olarak maddenin, maddeyle ölçülmek istenen yeteneğe sahip olmayan bir birey tarafında yanıtlanma ihtimali .2 olduğundan, yetenek skalasının tam ortasında ($\theta=0$) yer alan bir birey tarafından yanıtlanma ihtimalinin .6 olduğu görülmektedir.

MTK'nın ikinci varsayımı olan Yerel Bağımsızlık; madde ve yetenek parametreleri kontrol altına alındığında, bir maddeyi doğru yanıtlama olasılığının başka bir maddeye verilen yanıtın doğru olma olasılığı ile ilişkili olmaması durumudur. Bir başka ifadeyle, model parametreleri kontrol altına alındığında maddeler birbirinden bağımsız/ilişkisiz olmalıdır. Buna göre; madde parametreleri gruptan, bireylerin

kestirilen yetenekleri ise maddelerden bağımsızdır (Hambleton & Swaminathan, 1985; Embretson & Reise, 2000; Reckase, 2009). KTK'da, kestirilen madde parametreleri, ancak evreni temsil eden bir grupta benzer değerler verir. Oysa MTK'da, aynı evrene ait herhangi bir grubun verileri kullanılarak kestirilen madde parametreleri ile evrenden elde edilen parametreler arasında yüksek korelasyon olması beklenir. Yetenek parametrelerinin değişmezliği ise madde seti (boyutluluk durumu göz önüne alınarak) iki gruba ayrılıp, tüm bireylerin yetenekleri bu iki madde grubuna göre ayrı ayrı kestirildiğinde elde edilen yetenek kestirimleri arasında yüksek bir korelasyon olması durumunda sağlanmış olur.

MTK için diğer bir önemli öge, Madde Bilgi Fonksiyonu (MBF)'dur (Item Information Function). MBF; maddenin yetenek skalasının hangi değerinde, hangi miktarda bilgi verdiğini ortaya koyar. Her madde için bilgi fonksiyonuna ait grafik, kullanılan MTK modeline (1PL, 2PL ya da 3PL) göre farklılık gösterecektir. Aşağıda, iki parametrelili modele uygun kalibre edilmiş tek boyutlu bir maddenin; MKE'sine bağlı olarak ortaya çıkan madde bilgi ve standart hata grafikleri yer almaktadır.



Şekil 1.4: Tek boyutlu MTK'ya Göre Kalibre Edilmiş Bir Maddeye Ait Grafikler

Şekil 1.3'te; ayırt ediciliği $a=1$, güçlüğü $b=0$ olan 2PL bir maddeye ait grafikler yer almaktadır. Grafiklerden ilki MKE'dir. İkinci grafik MBF'ye, son grafik ise standart hataya aittir. Madde, oldukça geniş bir yetenek aralığında bilgi vermektedir. Ancak bununla birlikte, yetenek skalasının iki ucunda maddenin yeterli bilgi vermediği görülmektedir. Buna bağlı olarak da, uç yetenek değerlerinde standart hata miktarı artmaktadır. Maddenin ayırt edicilik değeri arttıkça MKE daha da dikleşecek, buna

bağlı olarak da daha dar bir aralıkta bilgi verecektir. Söz konusu durum göz önüne alındığında, BBT uygulamaları kapsamında bir madde havuzunda her yetenek düzeyinde yüksek bilgi veren maddelerin yer almasının gerekliliği daha anlaşılır olmaktadır. Öyle ki, bireye uyarlanmış bir testin etkili olabilmesi için bireyin yetenek düzeyinde yüksek bilgi veren, dolayısıyla daha az hatalı kestirimler yapan çok sayıda maddeye ihtiyaç vardır.

Tek boyutlu MTK modellerinde önemli bir husus da hangi modelin veriye uyum sağladığının belirlenmesidir. Model-veri uyumunun, eldeki veri setinin madde karakteristik eğrisi ile uyum göstermesi olduğu söylenebilir (De Ayala, 2009). Tek boyutluluğun sağlandığından emin olunmasının ardından, veri seti sırasıyla 1PL, 2PL ve 3PL modele göre kalibre edilir ve elde edilen ki-kare fark istatistiğine bağlı olarak hangi modelin veriye daha iyi uyum sağladığına karar verilir (Emretson & Reise, 2000). Matematiksel olarak çok daha kolay olması, tek boyutlu MTK modellerinde lojistik modeli bir çok model arasında daha popüler hale getirmiştir (Emretson & Reise, 2000). Aşağıda, 3PL lojistik modele ait formül yer almaktadır (Birnbaum, 1968).

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

Burada yer alan değerlerden şans parametresini temsil eden c değerinin “0” olması durumunda model 2PL haline gelecektir. Tüm ayırt edicilik parametrelerinin (a) eşit olması durumunda ise 1PL model elde edilmiş olur.

1.8.1.2. Çok Boyutlu Madde Tepki Kuramı

Erken dönem MTK çalışmalarında, her ne kadar madde setinde yer alan baskın bir boyut tek boyutluluğun göstergesi olarak kabul edilse de, psikolojik yapılar çoğunlukla çok boyutludur. Bu nedenle, 70’li yılların sonuyla birlikte çok boyutlu MTK çalışmaları hız kazanmıştır. Erken dönem çok boyutlu MTK çalışmaları Reckase (1972) ile başlamıştır. Bu çalışmayı Mulaik (1972), Sympson (1978) ve Whitely (1980) tarafından yapılan çalışmalar izlemiştir (Reckase, 2009). Çok boyutlu MTK’ya ivme kazandıran asıl çalışma ise; Bock & Aitken (1981) tarafından yapılan, MTK ve faktör analizini bir araya getiren çalışmadır (Reckase, 2009; Seo & Weiss, 2015; Chalmers, 2016).

Çok boyutlu MTK, belirli matematiksel formülleri kullanarak bireylerin çok boyutlu uzaydaki konumları ile test maddelerine verdikleri yanıtların doğru olma olasılığı arasındaki ilişkiyi gösterir. Çok boyutlu MTK, araştırmacının geçmişine göre çok değişkenliği istatistiksel analizin, özellikle faktör analizi ya da yapısal eşitlik modellemesinin özel bir durumu ya da tek boyutlu MTK'nın genişletilmiş hali olarak düşünülebilir (Reckase, 2009). Çok boyutlu MTK'nın iki kategorili maddeler için temel formülü şu şekildedir:

$$P(U = u \mid \theta) = f(u, \theta, \gamma)$$

Burada;

$U \rightarrow$ Bireyin test maddesi üzerindeki skoru

$u \rightarrow$ Yanıtlama ihtimali bulunan skorlar arasından atanmış olanı

$\theta \rightarrow$ Bireyin çok boyutlu uzaydaki konumunu gösteren parametrelerin vektörü

$\gamma \rightarrow$ Test maddelerinin özelliklerini açıklayan parametrelerin vektörü

Çok boyutlu modeller, çoğunlukla iki grupta incelenirler. Bunlardan ilki, telafi edici (compensatory) çok boyutlu MTK modelidir. Bu model, θ koordinatlarının lineer kombinasyonuna dayanır. θ değerlerinin farklı kombinasyonlarının kullanılması durumunda dahi toplamaları aynı değeri verecektir. Bu durum, bir θ koordinatının düşük olması durumunda diğerinin bunu telafi edecek derecede yüksek olmasına dayanır. Daha sade bir anlatımla; telafi edici çok boyutlu MTK modellerinde bireyin testle kestirilen boyutlardan birindeki düşük yeteneği, diğer boyuttaki yüksek yeteneği ile telafi edebilir. Diğer çok boyutlu MTK modeli ise, telafisel olmayan ya da kısmi telafisel (non-compensatory – partially compensatory) modeldir. Bu model, bilişsel görevleri ayırır ve her birinde tek boyutluluğu göz önüne alarak uygulama yapar. Bir maddeyi doğru yanıtlama ihtimali, her bir parça için var olan ihtimallerin bir ürünüdür. Bir başka ifadeyle bireyin bir boyuttaki yeterliliği, diğer boyuttaki eksikliğini telafi etmez (Embretson & Reise, 2000; Reckase, 2009).

Embretson & Reise (2000), çok boyutlu MTK modellerini doğrulayıcı ve açıklayıcı olmak üzere iki başlıkta incelemişlerdir. Buna göre açıklayıcı modeller, 1988'de Bock, Gibbons ve Muraki tarafından geliştirilen tam bilgi madde faktör analizi (Full Information Item Factor Analysis-FIIFA) olarak da bilinir. Doğrulayıcı modeller ise, maddeleri önsel (a priori) bilgiye dayalı hipotetik boyutlarla eşler. Çok boyutlu MTK'yı doğrulayıcı ve açıklayıcı olmak üzere iki bölümde inceleyen bu yaklaşım,

çok boyutlu MTK modellerindeki madde düzeyinde boyutluluk modelleriyle örtüşmektedir.

Çok boyutlu MTK modellerinde madde düzeyinde boyutluluk ikiye ayrılır. Bunlardan ilki, madde içi boyutluluktur (within-item dimensionality). FIIFA ile çok boyutlu MTK kalibrasyonu yapılması durumunda maddeler birden fazla boyuta yük verecekler, dolayısıyla ayırt edicilik parametrelerinin sayısı da yük verdikleri boyut kadar olacaktır. Madde düzeyinde boyutluluk modellerinden ikinci ise maddeler arası boyutluluktur (between-item dimensionality). Bu modelde, çok boyutlu yapıya dair önsel bilgi doğrultusunda her maddenin ait olduğu boyut önceden belirlenir. Buna göre; madde sadece belirlenen boyuta yük verecek, o boyutta ayırt edicilik parametresine sahip olacaktır. Bu modellere, çoklu-tek boyutlu (multi-unidimensional) modeller de denir (Wang & Chen, 2004; Sheng & Wikle, 2007).

Söz konusu iki modelin de kendine göre sınırlılıkları vardır. Maddeler arası boyutluluk modeli, her maddenin yalnızca bir faktöre yük vermesi nedeniyle, psikolojik yapıların doğası gereği çok gerçekçi bulunmamaktadır. Bununla birlikte, FIIFA sonucu elde edilen madde içi boyutluluk modelinde ise, boyutların tanımlanabilir (interpretable) olmadığı düşünülmektedir (Li & Schafer, 2005). Bu duruma alternatif olarak, geçmişte Holzinger ve Swineford (1937)'ye dayanmakla birlikte 90'lı yıllara kadar pek önemsenmeyen bifaktör modellerin kullanılması düşünülebilir. Bifaktör modellerde, tüm maddelerin altında yatan ortak bir varyans kaynağı, dolayısıyla bir "genel faktör" (general trait) bulunur. Bununla birlikte her bir madde genel yetenek faktörü dışında, önsel olarak ilişkili olduğu düşünülen başka bir faktöre daha yük verir (Seo & Weiss, 2015).

Bifaktör modellerin MTK'ya uyarlanması, Gibbons & Hedeker (1992) tarafından yapılan ve FIIFA'ya yenilikçi bir doğrulayıcı yaklaşım getiren çalışmayla başlar. Bu çalışmanın Gibbons ve diğ. (2007) çok kategorili maddeler üzerinde uyarlanmasıyla, tam bilgi madde bifaktör analizi (Full Information Item Bifactor Factor Analysis- FIIBFA) ortaya çıkmıştır. FIIBFA, çok boyutlu bir veri setinden madde parametresi elde etmek için MTK'ya dayalı ilk doğrulayıcı model olup, önsel bir teori ile gizil ve gözlenen değişkenler arasındaki ilişkiyi inceler (Seo, 2012).

Gibbons & Hedeker, 1992 yılındaki çalışmalarında tek boyutlu MTK ve çoklu-tek boyutlu MTK modellerini bir araya getirerek FIIBFA modelini ortaya koymuşlardır.

Buna dayalı olarak, FIIBFA modelde bireyin maddeyi doğru yanıtlama ihtimalini ortaya koyan formül ortaya çıkmıştır:

$$P(u_{ij} = 1 | \theta_{j1}, \theta_{jk}, \lambda_{i1}, \lambda_{ik}, \tau_i) = \Phi_i(\theta_1, \theta_k) = \frac{1}{\sqrt{2\pi\sigma_i}} \int_{\tau_i}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{X_i - \lambda_{i1}\theta_{j1} - \lambda_{ik}\theta_{jk}}{\sigma_i}\right)^2\right) dX_i,$$

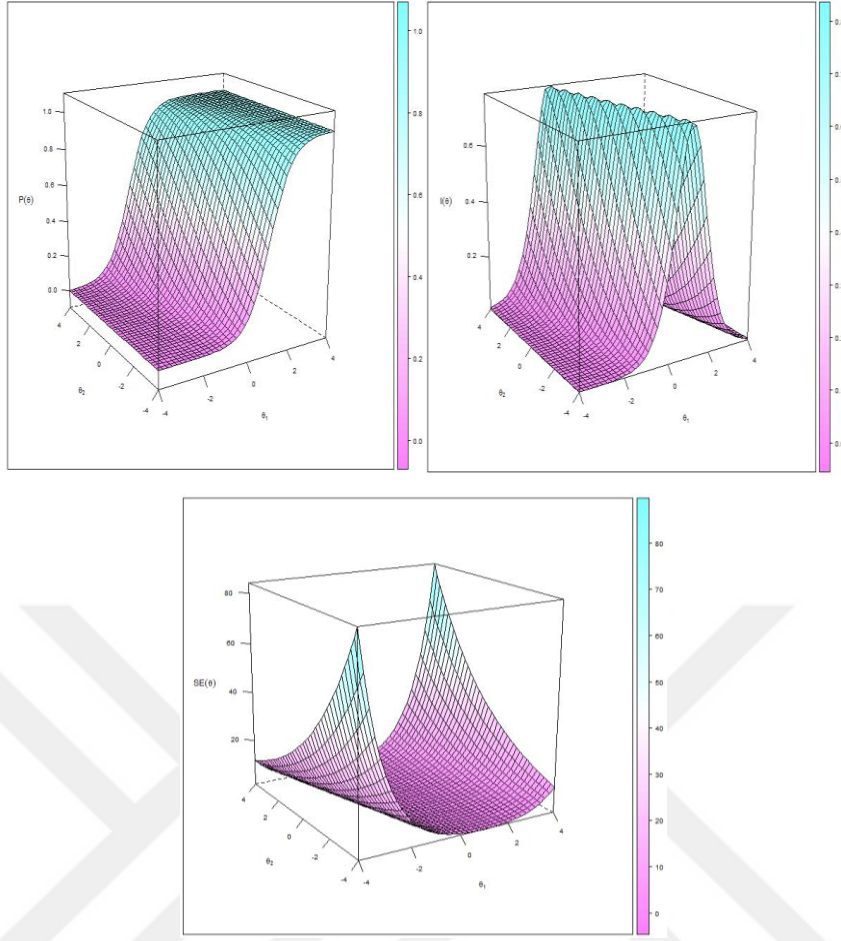
Bu formülde, $\sigma_i = \sqrt{1 - \lambda_{i1}^2 - \lambda_{ik}^2}$ dir. Burada, gizil değişken olan X_i , normal dağılıma sahiptir.

3PL bifaktör modelde, maddenin doğru yanıtlanma olasılığı aşağıdaki formülle ifade edilir:

$$P(u_{ij} = 1) = c_i + \frac{1 - c_i}{1 + \exp[-1.702 \times \mathbf{a}'_i(\boldsymbol{\theta}_j - b_i \mathbf{1})]},$$

Burada a parametresi i maddesinin ayırt ediciliğini, θ_j ise j bireyinin genel ve grup faktörlerine ait gizil değişkenlerin vektörünü ifade eder. b parametresi i maddesine ait güçlük, c ise şans parametresidir. Şans parametresinin “0” a eşitlenmesiyle, model 2PL haline gelir. Bu çalışmada, 2PL bifaktör model kullanılmıştır.

Çok boyutlu MTK modellerinde üzerinde durulması gereken bir başka husus da tek boyutlu modellerdeki MKE’nin kullanılış biçimidir. Tek boyutlu modellerdeki MKE, çok boyutlu modellerde bir yüzey halini alır. Aşağıda; boyutlara ait ayırt edicilik değerleri $a_1=2$ ve $a_2=.5$, güçlüğü ise $b=0$ olan çok boyutlu bir maddeye ait Madde Karakteristik Yüzeyiyle (MKY) birlikte, MBF ve bunlara ait hata yüzeyi grafiklerine ait şekiller yer almaktadır.



Şekil 1.5: 2PL Çok Boyutlu MTK'ya Göre Kalibre Edilmiş Bir Maddeye Ait Grafikler

Şekil 1.4'te yer alan grafiklerden ilki, çok boyutlu maddeye ait karakteristik yüzeydir. İkinci maddeye ait bilgi fonksiyonuyla elde edilen yüzey grafiği olup, buna bağlı olarak elde edilen standart hata yüzey grafiğinin, maddenin düşük bilgi verdiği alanlarda oldukça yüksek olduğu görülmektedir.

1.8.2. Bireyselleştirilmiş Bilgisayarlı Testleri Geliştirme ve Uygulama Süreci

Tanılama ya da değer biçme amacıyla uygulanan testlerin bireye uyarlanması/bireyselleştirilmesi, Binet (1905)'ten beri psikometristlerin üzerinde çalıştığı bir konudur. Başlangıçta, testin bireye özgü hale getirilmesi için bireyin yetenek düzeyinin bilinmesinin gerektiği, bununla birlikte yetenek düzeyinin bilinmesi durumunda ise bireyin bir teste tabi tutulmasına gerek olmadığı düşünülmüştür. Bu duruma literatürde “test deseni ikilemi” adı verilir. Bu ikileme çözüm olarak ise testin başlangıçta bireye uyarlanması yerine, her bir soruya verdiği yanıtlara göre bireyselleştirilmesi gündeme gelmiştir (Kalender, 2009). Günümüzde kullanılan BBT algoritması; kalibre edilmiş bir madde havuzu kullanılarak önceden

belirlenmiş kurallar çerçevesinde testin başlatılması, yanıtlanan her maddenin ardından bireyin yetenek kestiriminin yapılarak bireye soru havuzundan bu yetenek kestirimine uygun yeni bir madde sorulması ve testin yine önceden belirlenen kriter sağlandığında sonlandırılmasına dayanır.

Thompson & Weiss (2011)'e göre BBT uygulamalarının en önemli avantajı, bireyin yetenek düzeyi ve madde güçlüğü aynı ölçek üzerinde yerleştirilerek, bireyin test ile ölçülen yetenek düzeyine uygun maddelerin seçilmesini sağlamasıdır. Bu şekilde bireyler sadece kendi düzeylerine uygun maddeleri yanıtlamak durumunda olacaklarından, geleneksel kağıt-kalem testlerinde yanıtladıkları madde sayısından çok daha az madde ile testi sonlandıracaklardır. Testin bu şekilde bireye uyarlanmış hale getirilerek bireyin yetenek düzeyinin çok üzerinde veya çok altında maddeleri yanıtlama zorunluluğundan kurtulması, ölçme sonuçlarına karışan hata miktarını azaltarak ölçme kesinliğini artıracaktır. Bir başka ifadeyle BBT uygulamalarıyla geleneksel testlerle aynı ölçme kesinliğine çok daha az madde kullanılarak ulaşılabilecektir (Gibbons ve diğ., 2008; Weiss, 2011). Segall (2005)'e de göre BBT uygulamalarının ölçme etkililiğini (verimini) artırması, ölçme kesinliği ve testin uzunluğuna bağlıdır. Weiss (2011)'e göre ise; ölçme kesinliğinin artması, direkt olarak yanıtlanan madde sayısının azalmasına bağlıdır.

BBT uygulamalarının gerçekleştirilmesi karmaşık bir süreçtir. Öyle ki, testin kestireceği düşük hata değerlerinin yanında geçerliğin de yüksek olmasının sağlanması için tüm bileşenler üzerinde önceden bir çerçeve belirlenerek çalışılmalıdır. Thompson & Weiss (2011), BBT geliştirilmesinde aşağıdaki aşamaların uygulanmasını önermişlerdir:

1. Fizibilite, uygulanabilirlik ve planlama amacıyla *monte carlo* simülasyon çalışmalarının yapılması
2. Yeni madde yazım süreci ya da var olan maddelerin uygulanabilir hale getirilmesi
3. Maddelerin uygulanması ve sonrasında uygun modele göre kalibre edilmesi
4. *Post-hoc* ve *hibrit* simülasyon çalışmalarıyla gerçek-zamanlı BBT bileşenleri için uygun koşulların belirlenmesi
5. Belirlenen koşullar için hazırlanan yazılım kullanılarak gerçek-zamanlı uygulamanın yapılması

Gerçek-zamanlı bir BBT uygulamasının bileşenleri ise şunlardır (Weiss & Kingsbury, 1984):

- Madde havuzuna uygun tepki modeli,
- Kalibre edilmiş bir madde havuzu,
- Teste başlama kuralı,
- Madde seçim yöntemi,
- Yetenek kestirim yöntemi,
- Sonlandırma kuralı

Thompson & Weiss (2011) tarafında yapılan çalışmada ise yukarıdaki maddelerden ilkinde yer verilmemiş, BBT uygulamalarının kalan beş bileşenden oluştuğu belirtilmiştir. BBT uygulamalarına ait bu bileşenler üzerinde daha ayrıntılı olarak durmak gerekir.

1.8.2.1. Madde Havuzunun Oluşturulması

Verimli ve etkili bir BBT uygulaması için, yetenek ranjının her düzeyinde yüksek bilgi veren geniş bir madde havuzuna ihtiyaç duyulur. Söz konusu madde havuzu, her bir maddenin 500 – 1000 kişiye uygulanmasıyla (Yoes, 1995) kalibre edilmiş maddelerden oluşur. Bu havuzda yer alacak maddeler, BBT uygulamasının amacına uygun olarak kağıt-kalem formatında uygulanmalı ve bu doğrultuda MTK varsayımlarına göre kalibre edilmelidir.

Geniş bir madde havuzu elde etmek için bu havuzda yer alması planlanan maddelerin tamamını tek seferde ya da aynı gruba uygulamak mümkün olmaz. Bu durumda, aynı gruba farklı seferlerde ya da MTK'nın değişmezlik varsayımlarından yararlanarak aynı evrenden farklı gruplara bu maddeler uygulanır (Kolen & Brennan, 2004). Farklı zamanlarda ya da farklı gruplara yapılacak bu uygulamalarda testlerin eşitlenebilmesi için “anchor” maddeler kullanılması yararlı olacaktır (Nydick & Weiss, 2009).

Bazı araştırmacılar (Patsula & Steffan, 1997; Luecht, 1998) madde havuzunda, testin bireye uyarlanmasıyla yanıtlayacağı ortalama madde sayısının 6 – 12 katı arasında madde olması gerektiği düşüncesindedirler. Stocking (1994) ise, havuzda yer alması gereken soru sayısının kağıt-kalem şeklinde uygulanan sınavdaki soru sayısının en az 6 katı olması gerektiğini belirtmiştir. Bununla birlikte, madde

havuzunun büyüklüğü ve havuzda yer alacak madde sayısı BBT uygulamasının amacıyla da ilgilidir. Öyle ki, BBT ile bir kesme puanı belirlenerek bireylerin başarılarına göre gruplanmaları bekleniyorsa (geçti/kaldı vb.), söz konusu kesme puanına yakın çok fazla madde olması beklenirken, yetenek dağılımının uç kısımlarında fazla sayıda madde olmasına gerek duyulmayacaktır. Eğer BBT uygulaması sonucunda bireye bir not verilecekse, yetenek skalasının her bölümünde yeterli sayıda madde olması beklenir (Thompson & Weiss, 2011).

Maddelerin uygulanmasının ardından, madde setiyle ölçülen özellik hakkındaki önsel bilgi de göz önüne alınarak faktör analizi yapılır. Burada, araştırmacı ölçülen özelliğin kuramsal olarak tek boyutlu olduğunu düşünüyor ya da BBT algoritmasını tek boyutlu çerçevede yürütmek istiyorsa, tek boyutluluğu doğrulama yoluna gidecektir. Eğer yapının çok boyutlu olduğu düşünülüyor ise; öncelikle maddelerin tek boyutluğu göz önüne alınarak, sonrasında ölçülen yapıya uygun olacağı düşünülen çok boyutlu yapı ya da yapılar göz önüne alınarak MTK'ya dayalı faktör analizi yapılmalı ve elde edilen istatistiklere göre madde setine uygun boyutluluk durumuna karar verilmelidir. Bir sonraki adımda ise, belirlenen faktör analitik modele göre uygun MTK modeline (1PL, 2PL ya da 3PL) karar verilir.

BBT uygulamalarında üzerinde durulan önemli bir husus da madde teşhir oranıdır. Gerçek-zamanlı uygulamalarda farklı yetenek düzeylerinde yüksek bilgi veren madde sayısının az olması, yetenek düzeyleri benzer bireylerin karşısına sürekli aynı maddelerin çıkmasına ve dolayısıyla sınav güvenliğinin tehlikeye girmesine neden olur. Böyle bir durumda madde teşhir oranının kontrol altına alınmasına yönelik yöntemler yüksek madde bilgi fonksiyonuna sahip maddelerin sonraki uygulamalarda ortaya çıkmamasına, dolayısıyla havuzda bireyin kestirilen geçici yetenek parametresine uygun madde olmayacağından (ya da uygun olan maddenin düşük bilgi vereceğinden) bu düzeye en yakın maddeyi yanıtlamak durumunda kalmasına neden olacaktır. Böyle bir durum, bireylerin yanıtlaması gereken madde sayısının artmasına ve dolayısıyla BBT uygulamasının veriminin düşmesine neden olabilir. Bu nedenle her bir yetenek düzeyinde yüksek bilgi veren madde sayısının çok olması, madde teşhir oranını kontrol altına alındığı durumlarda ölçme kesinliğinin değişmemesini sağlayacak olması nedeniyle son derece önemlidir.

1.8.2.2. Teste Başlama Kuralı

Bireyselleştirilmiş testlerde bireyin teste hangi madde ile başlayacağına karar vermek; testin daha kısa sürede ve daha düşük hatayla tamamlanmasını, dolayısıyla ölçme kesinliğinin artmasını sağlayacaktır. Bireyle ilgili önsel (a priori) bir bilgi olmadığında testin yetenek ranjının ortasındaki bir maddeyle ($\theta = 0$) başlaması, bu yöntemin basitliği ve anlaşılabilirliği nedeniyle en çok kullanılan yöntemdir. Ancak tüm bireylerin aynı noktadan teste başlaması, bazı maddelerin kullanım sıklığını artıracığından bunun yerine yetenek ranjının orta noktasına .5'lik eşit uzaklıktaki aralıkta güçlük düzeyine sahip maddelerle başlanması uygun olacaktır (Thompson & Weiss, 2011; Weiss, 2011).

Bireyin teste hangi madde ile başlayacağı; bireyin test ile ölçülen yetenek düzeyi ile ilgili önsel bir bilgi olduğunda, bu bilgi kullanılarak da belirlenebilir. Örneğin, İngilizce hazırlık sınıfı öğrencilerinin yeterliliğinin ölçüldüğü bir dönem sonu sınavının BBT olarak uygulandığını düşünelim. Bu durumda, bireyin dönem içinde girdiği sınav, quiz vb. ölçme araçlarıyla kestirilen puanlarının ortalaması, bireyselleştirilmiş testteki başlama noktası olarak belirlenebilir. Bunun yanında, ölçülen özellikle ilgili bireye dair bir önsel bilgiye sahip olunmadığı, ancak ölçülen bu yetenekle yüksek korelasyon gösteren bireye dair başka bir bilgiye sahip olduğu durumda, bu bilgi de bireyin teste başlayacağı noktayı belirlemede kullanılabilir. Söz konusu duruma örnek olacak çalışmalarında Castro, Suarez & Chirinos (2010), bireylerin motivasyon ve sosyo-ekonomik düzeylerinin bu amaçla kullanılabileceğini, bu yöntemin tüm bireylerde olmasa da genel olarak verimliliği artırmada ve madde teşhir oranını azaltmada etkili olabileceğini belirtmişlerdir (akt; Thompson & Weiss, 2011). Bununla beraber bazı araştırmacılar yanıtlanan madde sayısının çok az olmadığı sürece başlama noktasının farklılaşmasının önemli bir farklılığa neden olmayacağını düşünürken bu görüşün aksine Wainer & Kiely (1987) ise, bireyin yetenek düzeyiyle farklılık gösteren başlama noktalarının kaygı ve moral bozukluğunu artırdığı görüşündedirler (akt; Hambleton, Zaal & Pieters, 1991).

1.8.2.3. Yetenek Kestirim Yöntemi

Testi alan bireyin başlama kuralına göre ilk maddeyi yanıtlamasının ardından, birey için geçici yetenek kestirimi yapılır. Yanıtlanan her bir maddenin ardından bireyin yanıtlaması gereken yeni maddenin belirlenmesi için bu yetenek kestiriminin

yapılması gerekir. Yetenek kestirimi için BBT uygulamalarında en sık kullanılan yöntemler aşağıda geniş şekilde açıklanmıştır.

1.8.2.3.1. Maksimum Olabilirlik Yöntemi (ML)

Diğer yetenek kestirim yöntemleriyle karşılaştırıldığında “en az yanlışlık gösteren yöntem” (Lord, 1968) olmasının yanında aynı zamanda genel olarak diğer yöntemlere göre daha fazla tercih edilir (de Ayala, 2009). ML yöntemiyle, madde düzeyinde kullanılabilir durumda olan tüm bilgi hesaba katılır. Bu yolla birey zor bir maddeyi yanıtladığında yetenek kestiriminde elde edilen değişim, kolay bir maddeyi yanıtladığında elde edilen değişime göre daha yüksek olacaktır. Kolay bir madde yanlış yapıldığında ise yetenek kestirimindeki düşüş, zor bir maddenin yanlış yapıldığı duruma göre daha yüksek olacaktır (Weiss, 2011). Bu yöntem, yanıt örüntüsünün olabilirlik fonksiyonunun optimize edilmesine dayanır. Söz konusu fonksiyon aşağıda görülmektedir:

$$l(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{j=1}^J \prod_{k=0}^{K_j-1} P_j(y = k|\boldsymbol{\theta}, \boldsymbol{\psi}_j)^{\chi_{jk}}$$

Bu denklemde χ_{jk} , 0-1 şeklinde kodlanan ikili bir değişkeni temsil eder. Bu değişken, uygun kategorilerle ilgili olan olasılık koşullarını seçmek için kullanılır. ML yöntemi ile ilgili önemli bir problem, madde karakteristik eğrilerinin “0” değerini içermesidir. Bunu önlemek için, logL (log-likelihood) fonksiyonu kullanılır:

$$LL(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{j=1}^J \sum_{k=0}^{K_j-1} \chi_{jk} \cdot \log [P_j(y = k|\boldsymbol{\theta}, \boldsymbol{\psi}_j)] .$$

Yukarıdaki denklem, madde karakteristik eğrisinin doğal logaritması alınarak elde edilir. Bununla birlikte, ML yönteminin önemli bir sınırlılığı daha vardır. Eğer yanıt vektöründe bir değişkenlik yoksa, başka bir ifadeyle bireyin verdiği yanıtların tamamı doğru ya da tamamı yanlışsa, yetenek kestirimi yapılamaz. ML yöntemindeki bu önemli sınırlılık, Bayes yöntemlerin kullanılmasıyla aşılabılır (Embretson & Reise, 2000; Frey, 2009; Weiss, 2011; Chalmers, 2016).

1.8.2.3.2. Bayes Kestirim Yöntemleri

ML yönteminde sonlu yetenek kestiriminin elde edilememesi durumuna çözüm olması amacıyla, çok boyutlu BBT uygulamaları için Segall (1996) tarafından Bayes yöntemlerin kullanılması önerilmiştir. Bayes yöntemlerde bir önsel dağılımın kullanılması gerekmektedir. Bu önsel dağılımla ilk maddeye verilen yanıt birleştirilerek bir sonsal dağılım elde edilmiş olur. Bu sonsal dağılım ise, yeni maddenin seçiminde önsel dağılım olarak kullanılacaktır. Söz konusu önsel dağılımda ortalama kullanılması durumunda EAP, mod kullanılması durumunda ise MAP yöntemi tercih edilmiş olur (Reckase, 2009). Bayes metotlarda yararlanılan sonsal fonksiyon şöyledir:

$$\pi(\theta|y, \psi, \eta) = \frac{l(y|\theta, \psi)\phi(\theta|\eta)}{\int l(y|\theta, \psi)\phi(\theta|\eta)}$$

Bu fonksiyon üzerinde, integral kestirimlerin bulunması ya da maksimize edilmesine dayalı olarak sırasıyla EAP ve MAP kestirimleri elde edilmiş olur (Chalmers, 2016). EAP ile yapılacak hesaplamalar, olabilirlik fonksiyonunun birinci ve ikinci türevini almak zorunda olmaması nedeniyle nispeten daha kolaydır. Ayrıca EAP'ın iteratif bir yöntem olmaması, geçici yetenek kestirimini kolaylaştıran başka bir etkidir (Bock & Aitkin, 1981). Bayes yöntemler, bireyin maddelerin tamamına doğru ya da yanlış yanıt vermesi durumunda dahi düşük standart hata değerleri elde edilmesini sağlar (Hambleton & Swaminathan, 1985).

1.8.2.4. Madde Seçim Yöntemi

Yetenek kestiriminin yapılmasının ardından önceden belirlenen sonlandırma kriterinin sağlanmamış olması durumunda, bireyin geçici yetenek düzeyine en uygun maddenin soru havuzundan çekilerek bireye sunulması gerekir. Madde seçim yöntemi olarak bilinen bu süreç, diğer tüm süreçler gibi belirli bir kurallar silsilesine göre gerçekleştirilir. Madde seçim yöntemi, BBT uygulamalarının özüdür. (Frey, 2009). Çok boyutlu BBT uygulamalarında diğer BBT bileşenleri tek boyutlu uygulamalarla paralel olsa da madde seçim yöntemleri farklılık gösterir. Bununla birlikte burada da yöntemler tek boyutlu madde seçim yöntemlerinde olduğu gibi çoğunlukla “maksimum bilgi”ye dayanır (Babcock & Weiss, 2012). Literatürde, çok boyutlu BBT uygulamalarında kullanılabilecek çok fazla madde seçim yöntemi bulunmamaktadır. Çok boyutlu madde seçim yöntemleri temel olarak geçici çok

boyutlu yetenek vektörünün beklenen varyansında bir küçülmeyi amaçlar. Çok boyutlu MTK çerçevesinde Fisher'in bilgi matrisi, önceden uygulanmış maddelere verilen yanıtlarla bireyin yetenek düzeyi ile ilgili alınan bilginin miktarı olarak görülebilir. Matematiksel olarak, log-olabilirlik L'nin ikinci türevinin negatif halidir (Frey, 2009):

$$I(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = -E\left(\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta}^2}\right)$$

Bu yöntem, henüz uygulanmayan maddelerin katkı sağladığı beklenen bilginin büyüklüğünü tahmin etmek için kullanılabilirliği nedeniyle çok boyutlu BBT uygulamaları için oldukça kullanışlıdır. Çok boyutlu BBT uygulamalarında kullanılan yöntemlerin çoğu, bu formüle dayalı olarak geliştirilmiştir. Bu formül, çok değişkenli normal sonsal dağılım söz konusu olduğunda şu hali alır:

$$\mathcal{F}_{J+m}(\hat{\boldsymbol{\theta}}) = \mathcal{F}_J(\hat{\boldsymbol{\theta}}) + \mathcal{F}_m(\hat{\boldsymbol{\theta}}) + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$$

$\mathcal{F}_{J+m}(\boldsymbol{\theta})$ 'daki ortak madde bilgisini ölçmenin bir yolu, en büyük matris determinantını sağlayan maddeyi seçmektir. Söz konusu madde, hangi maddenin F'nin hacminde en büyük artışı yarattığını bulur. Bu madde seçim yöntemine “D-rule” adı verilir. Formülü ise şu şekildedir:

$$\text{D-rule} = \max \left(|\mathcal{F}_{J+1}(\hat{\boldsymbol{\theta}})|, |\mathcal{F}_{J+2}(\hat{\boldsymbol{\theta}})|, \dots, |\mathcal{F}_{J+M}(\hat{\boldsymbol{\theta}})| \right)$$

D-rule madde seçim yönteminin yanında yine kullanışlı başka bir madde seçim yöntemi de “T-rule”dir. Bu yöntem, gizil değişkenler hakkında ortalama ağırlıklandırılmamış bilgiyi artıran maddeleri seçer. Formülü ise şu şekildedir:

$$\text{T-rule} = \max \left(\text{Tr}(\mathcal{F}_{J+1}(\hat{\boldsymbol{\theta}})), \text{Tr}(\mathcal{F}_{J+2}(\hat{\boldsymbol{\theta}})), \dots, \text{Tr}(\mathcal{F}_{J+M}(\hat{\boldsymbol{\theta}})) \right)$$

Bu madde seçim yönteminde olduğu gibi “W-rule” yöntemi de bazı gizil değişkenlerin diğerlerinden daha önemli görülmesi durumunda ağırlıklandırma yapabilecek bir yöntemdir (Chalmers, 2016).

Bu madde seçim yöntemlerinin yanında, Kullback–Leibler (KL) madde seçim yöntemi de çok boyutlu BBT uygulamalarında kullanılabilir. KL yöntemi ile parametre aralığı üzerinden iki olabilirlik arasındaki mesafe hesaplanabilir (Veldkamp &

Linden, 2002). Bu yöntem, az sayıda madde uygulanması durumunda bile θ değerlerindeki belirsizliği hesaba katabilir (Chang & Ying, 1996). Formülü ise şu şekildedir:

$$KL(\theta||\theta_0) = E_{\theta_0} (LL(y|\theta_0) - LL(y|\theta)) .$$

1.8.2.5. Sonlandırma Kuralı

BBT uygulamalarının temel bileşenlerinden sonuncusu, bireyin testinin nasıl sonlanacağına karar verilmesini sağlayan kuralın belirlenmesidir. Testin sonlandırılmasında kullanılan yöntemler ilki, sabit madde sayısına dayalı sonlandırma kuralıdır. Bu kurala göre; testin öncesinde belirlenen madde sayısı, testi yanıtlayan herkes için sabittir. Testi yanıtlayanlar açısından çok belirgin ve basit olması, kağıt-kalem testleriyle olan benzerliği nedeniyle en popüler sonlandırma kuralı olarak görülmüştür (Gushta, 2003; akt. Babcock & Weiss, 2012). Simülasyon çalışmalarında da çoğunlukla bu kural kullanılır. Bununla birlikte, bireyselleştirilmiş testlerin doğası gereği bireylerin yanıtlaması gereken soru sayısı farklılaşacağından; özellikle yetenek skalasının iki ucundaki bireyler için bu yöntemle farklı, kimi zaman çok yüksek standart hata değerlerinin elde edilmesi söz konusu olabilir. (Finkelman, Nering & Roussos, 2009; Frey, 2009).

Değişken madde sayısının söz konusu olduğu BBT uygulamalarında en çok kullanılan durdurma yöntemi ise standart hataya dayalı sonlandırma kuralıdır (Weiss & Kingsbury, 1984). Psikometrik bilgi, standart hata ile direkt ilişkili olduğundan yola çıkılarak (Samejina, 1977) kullanılan bu yöntem; bireyin yanıtladığı madde sonrasında hesaplanan standart hata değerinin, BBT uygulaması başlamadan önce belirlenen bir standart hata değerinin altına düşmesi sonrası testin durdurulmasına dayanır (Babcock & Weiss, 2012). Çok boyutlu testlerde tüm boyutlar için bu kuralın kullanılmasının önerilmiş olmasının (Wang & Wang, 2001) yanında, bifaktör modellerde sadece genel yetenek düzeyi için belirli bir standart hata değeri belirlenmesi söz konusu olabilir.

Değişken madde sayısına dayalı bir diğer sonlandırma kuralı ise, θ değişimine bağlıdır. BBT uygulamalarında bireyin yanıtladığı her madde sonrası kestirilen θ değerindeki değişim giderek azalır. Bu kural kullanıldığında, söz konusu değişimin önceden belirlenen bir değer altına düşmesiyle test sonlandırılır (Weiss & Kingsbury, 1984; Weiss, 2011).

BBT uygulamalarında en sık kullanılan yukarıdaki yöntemlerin yanında, bazı farklı sonlandırma kuralları da bulunmaktadır. Bunlardan biri; madde havuzunda, anlık θ kestirimi sonucu elde edilen değer düzeyinde bilgi verecek madde kalmadığında kullanılan maksimum bilgiye dayalı sonlandırma kuralıdır (Gialluca & Weiss, 1979; Maurelli & Weiss, 1981). Bir başka ifadeyle, soru havuzunda testin devam etmesini gerektirecek düzeyde bilgi verecek madde kalmamıştır. Choi, Grady ve Dodd (2010) ise, bu yöntemi standart hataya uyarlamıştır. Bireyin yanıtlaması için havuzdan seçilecek yeni maddeyle önceden belirlenen bir standart hata değer düşüşü sağlanamıyorsa test sonlandırılır. Hart ve diğ. (2006) tarafından önerilen hibrit sonlandırma kuralı ise, standart hata ve θ değişimine dayalı durdurma kurallarının beraber kullanılmasına dayanır.

Değişken madde sayısına dayalı sonlandırma kurallarında bireyin yetenek düzeyinde madde havuzunda yeterli madde bulunmaması durumunda, bireyin yanıtlaması gereken soru sayısı artacak, test uzayacaktır. Böyle durumlar göz önüne alınarak, yanıtlanacak soru sayısı için maksimum bir değerin belirlenmesi etkili olacaktır (Frey, 2009). Zaman zaman ise, yanıtlanacak bazı maddelerin bireyin anlık yetenek kestirimiyle örtüşmesi ve bu seviyede yüksek bilgi veren maddeler olmaları, testin çok az maddeyle sonlanmasına neden olabilir. Böyle durumlar için de yanıtlanacak minimum soru sayısının BBT uygulaması öncesi belirlenmesinin, testin kapsam geçerliğine yönelik eleştirileri azaltacağı düşünülebilir.

BBT uygulamalarında olmazsa-olmaz bu beş bileşenin yanında, uygulamaların gelişmesiyle üzerinde yapılan çalışmaların arttığı farklı bileşenler de vardır. Bunlar arasında, özellikle testin geçerliği için önceden belirlenen bir konu-boyut ağırlıklandırılmasının yapılması ya da çeşitli nedenlerle kullanım sıklıkları yüksek olan maddelerin teşhir oranlarının belirlenmesi ve kontrol altına alınmasına yönelik yöntemler yer almaktadır.

1.8.2.6. İçerik Ağırlıklandırması ve Madde Teşhir Oranının Kontrol Edilmesi

BBT uygulamalarının yaygınlık kazanması ve yüksek önem arz eden gerçek uygulamalarda kullanılmaya başlamasıyla beraber, uygulamaların geçerliği ve madde havuzunun güvenliği açısından başka yöntemler de gündeme gelmiştir. İçerik ağırlıklandırılması; maddelerin önceden belirli kıstaslara göre (kazanım, ünite, boyut, MTK modeli vb.) gruplanarak, farklı gruptaki maddelerin testte yer alma

oranının bu doğrultuda uygulamadan önce belirlenmesine dayanır. Bu amaçla kullanılabilecek bir yöntem, Kingsbury & Zara (1991) tarafından geliştirilmiştir. Çok boyutlu uygulamalarda genel bir yeteneğin kestirilmesi amaçlandığında bifaktör modelin kullanılması bir içerik ağırlıklandırılmasını sağlayacaktır (Segall, 1996).

BBT uygulamalarında bir diğer sorun ise madde teşhir oranıdır. Gerçek uygulamalarda kimi maddeler, çeşitli nedenlerle (başlama noktası, maddelerin yüksek bilgi vermesi vb.) havuzda yer alan diğer maddelere göre çok daha fazla kullanılır. Yapılan araştırmalar, BBT uygulamalarının %50'sinde madde havuzunun ancak %14'lük kısmının kullanılabildiğini göstermiştir (Wainer, 2000). Bu ve benzeri durumlar, testi henüz almamış kimselerin testi alan bireylerden elde ettikleri bilgilerle gerçek yetenek düzeylerinden daha yüksek kestirimler elde etmelerine ve dolayısıyla sınav güvenliğinin tehlikeye girmesine neden olabilir. Bu nedenle, madde teşhir oranının kontrol altına alınması için yapılmış birçok çalışma mevcuttur.

Stocking & Lewis (1998), teşhir oranının kontrolü üzerine yaptıkları çalışmalarında Sympson & Hetter (1985) metodunu genişlettikleri bir model sunmuşlardır. Çok boyutlu yapılar içinse Lee, Ip & Fuh (2002), Chang & Ying (1999) tarafından geliştirilen tabakalaşma metodunu geliştirerek koşulsuz madde teşhir oranı kontrol metodunu ortaya koymuşlardır. Bu ve benzeri yöntemlerin gerçek zamanlı uygulamalarda meydana getirdiği önemli bir problemten söz edilebilir. Buna göre; madde teşhir oranını kontrol altına alan yöntemlerin kullanılması madde seçim yöntemlerinin etkililiğini azaltacak, bu durum da özellikle standart hataya dayalı durdurma yönteminin kullanıldığı durumlarda yanıtlanan madde sayısının giderek artmasına neden olacaktır (Chalmers, 2016). Bu çalışma kapsamında kullanılan madde havuzunun yeterince büyük olmadığı, bu nedenle madde teşhir oranının kontrol edilmesine yönelik yöntemlerin kullanılmamasının uygun olacağı düşünülmüştür. Ayrıca çalışma kapsamında bifaktör MTK modeli kullanıldığından, başka bir içerik ağırlıklandırma yönteminin kullanılmasına gerek kalmamıştır.

1.8.3. Simülasyon Süreci

Madde havuzu henüz oluşturulmadan önce yapılacak fizibilite çalışmaları için *monte carlo* simülasyon süreci gerçekleştirilmelidir. Bu simülasyon, hipotetik bir boyutluluk durumuna göre madde ve yetenek parametrelerinin istenilen dağılıma göre üretilmesiyle gerçekleştirilir. Gerçek-zamanlı tek ve çok boyutlu BBT algoritmasının

belirlenmesi sürecinde, bireylerin madde havuzunu oluşturan maddelere kağıt-kalem testiyle verdikleri yanıtları kullanan *post-hoc* simülasyonun kullanılması daha güvenilir sonuçlar elde edilmesini sağlar. *Monte carlo* simülasyon tekniğinin aksine *post-hoc* simülasyonlar, gerçek madde parametreleri ve bu maddelere kağıt-kalem testinde verilen yanıtları kullanırlar (Weiss, 2004).

Post-hoc simülasyonlarda bireyin, madde havuzu oluşturulurken kağıt-kalem şeklinde uygulanan maddelere verdikleri yanıtlar kullanılır. Ancak çoğu zaman, madde havuzunda yer alacak maddelerin tamamının aynı zamanda ya da aynı gruba uygulanması mümkün olmaz. Bu durum, *post-hoc* simülasyon için kullanılacak madde yanıt-birey matrisinin eksik olmasına neden olacaktır. Bu tür durumlarda ise eksik yanıtların, bireylerin yanıtladıkları maddelerden elde edilen yetenek kestirimleri ve yanıtlamadıkları maddelerin madde parametrelerine dayanılarak üretilmesi ve böylece *post-hoc* simülasyonda kullanılacak tam veri matrisinin elde edilmesi gerekir. *Post-hoc* ve *monte carlo* simülasyonlarının beraber kullanılmasına dayanan bu tür simülasyonlara ise, *hibrit* simülasyon adı verilir (Nydick & Weiss, 2009).

Bu çalışmada; bireylerin yanıtlamadıkları maddeler için, kağıt-kalem formatında yanıtladıkları maddelerden elde edilen yetenek kestirimlerine ve söz konusu maddelerin parametrelerine dayanılarak yanıt (response) üretilerek *post-hoc* ve *monte carlo* simülasyonlarına dayanan *hibrit* bir simülasyon uygulaması yapılmıştır.

1.8.4. BBT Uygulamalarının Avantajları ve Sınırlılıkları

BBT uygulamaları, birçok avantajı beraberinde getirmektedir. Bireylerin sadece kendi yetenek düzeylerine uygun maddeleri yanıtlamaları; testin çok daha az maddeyle, daha kısa zamanda ve daha az hatayla kestirilmesini sağlar. Bireyler, uygulayıcıların uygun görmeleri durumunda sınava tekrar tekrar girebilir ve sınav sonucunu uygulama sonlandığında elde edebilirler. Bu durum, yüksek önem arz eden sınavları yılda yalnızca bir kez alabilen ya da bu sınavların uygulanmalarının zorunlu olduğu tarihte kendilerini henüz hazır hissetmeyen bireylerin üzerindeki sınav baskısını azaltacaktır. Bunların yanında, BBT uygulamaları sınav organizasyonu yapan kurumlar için de büyük avantajlar sağlar. Örneğin, büyük ölçekli sınavları Türkiye'nin her tarafında aynı anda yapmak için gerekli organizasyonu (soruların her sınav öncesi tekrar hazırlanması, kitapçıkların basımı ve sınav merkezlerine ulaştırılması vb.) sağlayan kurumlar BBT uygulamalarına

geçemeleri durumunda bu yükten kurtulmuş olacaklardır. Getirdiği bunca avantajın yanında, BBT uygulamaları bazı sınırlılıklar da taşımaktadır. Bunlar aşağıda birkaç maddeyle özetlenmiştir:

1. BBT uygulamalarının, özellikle bilgisayar kullanımına yönelik önyargı ya da endişe taşıyan bireylerin başarılarını olumsuz etkilediğine yönelik çalışmalar da mevcuttur (Colwell, 2013; Fritzt & Marszalek, 2010; Legg & Buhr, 1992).
2. BBT uygulamalarında her bir maddeye verilen yanıt sonrası yetenek kestirimi yapılarak bu yetenek düzeyine uygun yeni soru seçildiğinden, bireyin herhangi bir soruyu atlama, boş bırakma ve sonra geri dönerek yanıtlama şansı yoktur.
3. BBT uygulamalarında önceden kalibre edilmiş geniş bir madde havuzuna ihtiyaç vardır. Bu nedenle, her tür sınav için uygun olmayabilir.
4. BBT uygulamalarının temel dayanağı her birey için kendi düzeyine uygun maddelerin seçilmesi olduğundan bu durum, bireylerde eşitsizlik algısının oluşmasına neden olabilir.

1.8.5. Simülasyon ve Gerçek-Zamanlı BBT Uygulamaları için Kullanılan Yazılımlar

BBT uygulamalarında simülasyon amaçlı kullanılan, ticari bir amaç gütmeyen yazılımlar arasında; Choi (2009) tarafından çok kategorili BBT uygulamaları için geliştirilen Firestar, Han (2012) tarafından geliştirilen SimulCAT, Yao (2011) tarafından geliştirilen simuMCAT yer almaktadır. Bunların yanında R’de tanımlı; sadece tek boyutlu BBT simülasyonlarının yapılabildiği paketler (catIrt – Nydick, 2014; catR – Magis, Raiche & Barrada, 2012) ile çok boyutlu simülasyonlar için de kullanılabilen (MAT – Choi & King, 2014) paketler yer almaktadır.

Gerçek-zamanlı uygulama imkanı sunan açık kaynaklı R temelli BBT uygulamalarından ilki Concerto’dur. Cambridge Üniversitesi Psikometri Merkezinin desteğiyle Kosinski & Rust (2011) tarafından geliştirilen Concerto, geri planda catR paketini kullanarak tek boyutlu BBT uygulamalarının yapılmasına imkan sağlar. Bununla birlikte tam olarak R temelli bir uygulama olmaması, R’nin sadece arka planda hesaplama amacıyla kullanılması Concerto’nun önemli bir sınırlılığıdır. Bu özellikleri nedeniyle Concerto kullanımı için R’nin dışında bir çok bilgisayar temelli beceriye sahip olmak gerekir. Tüm bunların yanında yalnızca tek boyutlu uygulamalara izin vermesi, Concerto için başka bir sınırlılıktır.

Chalmers (2016) tarafından geliştirilen mirtCAT paketi; simülasyon çalışmalarına, tek ve çok boyutlu bireyselleştirilmiş uygulamalara gerçek-zamanlı uygulamalara imkan sağlayan bir R paketidir. Paket, yazar tarafından 2012'de geliştirilen mirt paketi üzerine kurulmuştur. mirtCAT, yine R'de tanımlı shiny (2014) paketinde kullanılabilir durumdaki HTML üreten araçlarla tek ve çok boyutlu gerçek-zamanlı BBT uygulamaları yapma imkanı sunar. Bu pakete <https://CRAN.Rproject.org/package=mirtCAT> adresinden erişilebilir. Bunların yanında ayrıca mirCAT paketi ile; iki kategorili testlerin yanında çok kategorili uygulamalar yapılabildiği gibi, herhangi farklı MTK modellerinin de aynı anda kullanılabilmesi mümkündür (Chalmers, 2016).



2. İLGİLİ ARAŞTIRMALAR

Bu bölümde, BBT uygulamaları ile ilgili yapılan araştırmalar, yurtiçi ve yurtdışında yapılan çalışmalar olmak üzere iki bölümde incelenmiştir. Yurtdışında yapılan çalışmalardan sadece çok boyutlu BBT uygulamaları incelenirken, yurtiçinde tek ve çok boyutlu BBT uygulamalarına yönelik araştırmalar bulunmaktadır.

2.1. Çok Boyutlu BBT Uygulamaları ile İlgili Yurtdışında Yapılan Çalışmalar

Bloxom & Vale, 1987 yılındaki çalışmalarıyla tek boyutlu BBT uygulamalarını çok boyutlu BBT'ye uyarlamaya çalışmışlardır. Bu yönüyle araştırma, çok boyutlu BBT literatürü açısından milat özelliğini taşır. Araştırmacılar, Owen (1975)'in ardışık güncelleme sistemini çok değişkenli uyarlamasını bir yetenek kestirim yöntemi olarak ortaya koymuşlardır. Bunu yapmalarındaki amaç, iteratif yöntemlerin sayısal algortimalarla ilgili yoğun hesaplamalarından kaçınmaktır. Araştırmacılar çok boyutlu bireyselleştirilmiş çalışmalar çerçevesinde bir yetenek kestirim yöntemi sunarken, madde seçimiyle ilgili herhangi bir noktaya değinmemişlerdir.

Tam (1992), araştırmasında iki boyutlu normal ogive modele dayalı iteratif bir ML yetenek kestirim yöntemi geliştirmiştir. Bu yöntemi geliştirirken ölçme kesinliği, testten elde edilen bilgi ve hesaplama zamanı da dikkate alınmıştır. Bloxom & Vale (1987) gibi Tam da çok boyutlu bireyselleştirilmiş çerçevede herhangi bir yeni madde seçim yöntemi önermemiştir. Araştırma kapsamında kullandığı madde seçim yöntemi, madde güçlüğü ile bireyin yetenek düzeyinin eşleştiği kusursuz bir madde havuzu gerektirmektedir.

Bloxom ve Vale (1987) ile Tam (1992)'in yaptığı çalışmalar her ne kadar çok boyutlu BBT'nin ilk çalışmaları olsa da bu çalışmalar sadece theta kestirimine odaklanmış ve theta'nın çok değişkenli önsel dağılımını göz önüne alan çok boyutlu bireyselleştirilmiş madde seçim algoritması eksikliğini giderememişlerdir. Bu nedenle söz konusu çalışmalar, çok boyutlu BBT uygulamalarının tek boyutlu BBT uygulamaları karşısındaki performansını gösterme fırsatı sunmamıştır. Çok boyutlu BBT çalışmalarındaki bu probleme çözüm getiren ise 1996 yılındaki çalışmasıyla Segall olmuştur.

Segall (1996)'in çok boyutlu BBT üzerine yaptığı çalışma, çok boyutlu BBT literatürü açısından kritik önem taşımaktadır. Araştırmacının kendi fadesiyle "Bloxom ve Vale

(1987) ile Tam (1992)'in alışmaların genişletilmiş hali" olan bu alışmada; yeteneğın ortak (bileşik) dağılımını içeren bir madde seçim algoritması (Bayes madde seçim algoritması), ML ve Bayes yetenek kestirim yöntemlerinin ok boyutlu yapılar çerçevesinde incelenmesi ve ok boyutlu BBT ile tek boyutlu BBT uygulamalarının performanslarının karşılaştırılmasının yapılması amacıyla Silahlı Hizmet Mesleki Beceri Testi (ASVAB) verileri kullanılmıştır. 9 boyuttan oluşan söz konusu test verileri üzerinde iki simölasyon alışması gerçekleştirilmiştir. Simölasyon alışmalarının amacı tek ve ok boyutlu BBT uygulamalarının güvenilirliklerini karşılaştırmak ve ok deėişkenli madde seçim algoritmasının etkililiğini test etmektir. Tüm koşullarda kullanılan 1145 maddenin parametreleri 3PL modele göre elde edilmiştir. Gerçekleştirilen simölasyon alışmalarının ilki, ok boyutluluk çerçevesinde Bayes madde seçim yetenek kestiriminin kullanıldığı, test uzunluğunun ise 9 ila 120 madde arasında deėiştii toplam 15 koşul için gerçekleştirilmiştir. İkinci simölasyon alışması ise tek boyutlu çerçevede olup maksimum bilgi madde seçim yönteminin kullanıldığı, toplam 2000 birey için theta üretildiėi bir alışmadır. İlk simölasyonda olduėu gibi burada da sonlandırma sabit madde sayısına dayalı olarak yapılmıştır. Simölasyonlar sonucu elde edile sonuçlara göre, madde sayısı arttıka alt testlerin güvenilirliklerinin de arttığı görölmektedir. ok boyutlu BBT'lerin 80 soruyla ulaştıkları güvenirliliėe tek boyutlu BBT'lerin 120 soru ulaştii görölmüştür. Eşit sayıda maddenin kullanıldığı durumlarda ise ok boyutlu BBT ile elde edilen güvenilirlik deėerlerinin daha büyük olduėu bulgulanmıştır. Sonuç olarak aynı güvenilirlik deėerlerine ok boyutlu uygulamaların tek boyutlulara göre yaklaşık 1/3 oranında daha az maddeyle ulaştii sonucuna varılmıştır.

Luecht'in 1996 yılındaki alışması, Segall'in alışmasıyla aynı yılda yapılmış olup, ok boyutlu BBT uygulamalarının tek boyutlu uygulamalar karşısındaki performansını gösteren bir başka önemli alışma olma özelliğini taşır. alışma, medikal lisanslama sınavında tek boyutlu ve ok boyutlu BBT performanslarını karşılaştırmak üzere 8 alt kategori için tam içerik aėırlıklandırması kullanarak gerçekleştirilmiştir. Önceden uygulanmış 2458 maddenin kullanıldığı simölasyon alışması öncesinde yapılan kaėıt-kalem sınavı, iki gün süren 600 soruluk bir sınavdır. Analizlere göre, BBT'nin tek bir içerik için yetenek kestirmek ya da bireylerin geme-kalma durumu hakkında karar vermek için kullanıldığı durumlarda

çok boyutlu BBT'lerin bir çok dezavantajın yanında sınırlı sayıda avantaj sunuyor olması nedeniyle tek boyutlu BBT'lerin daha uygun olduğu sonucuna ulaşılmıştır. Eğer amaç birden fazla yetenek kestirimi yapmak ya da birey için birden fazla profil verilmesi ise çok boyutlu testlerin kullanılmasının uygun olduğu belirtilmiştir. Ayrıca çok boyutlu BBT uygulamaları ile tek boyutlu BBT uygulamalarına göre %25-%80 arasında daha az madde kullanılarak aynı ölçme kesinliği elde edilmiştir.

van der Linden, 1996'da yine kendisinin yürüttüğü çalışmayla aşamalılık arz ettiğini belirttiği 1999 yılındaki çalışmasında, farklı yeteneklerin lineer kombinasyonlarına ait hata varyanslarının bir madde seçim algortiması olarak kullanılabilirliğini tartışmıştır. Teorik alt yapıyı sunmasının ardından yaptığı simülasyon çalışmasında, iki boyutlu faktör yapısına sahip 500 maddeden oluşan bir madde havuzu kullanmıştır. Oluşturduğu algoritma doğrultusunda önce yetenek bileşeni tanımlanır. Sonraki adımda ise madde havuzundan maksimum bilgi içeren maddeler seçilir. Araştırmacı, çalışmanın kritik kısmının ise test uzunluğu olduğunu belirtmiştir. Buna göre oluşturulan madde seçim algortimasının ancak 50 maddelik testlerde etkili olduğunu belirtilmiştir. Daha kısa testlerde ise ML ile kestirilen iki boyuta ait theta değerlerinin yanlış ve yetersiz olduğu sonucuna ulaşmıştır. Segall (1996) yılındaki çalışmasına benzer olarak Bayes yöntemlerin ML'ye göre daha bilgilendirici olduğu üzerinde durulmuştur.

Veldkamp & van der Linden (2002), çalışmalarında gölge testlerle çok boyutlu BBT uygulamalarında KL madde seçim yönteminin kullanılabilirliğini test etmişlerdir. Segall (1996)'in ortaya koyduğu madde seçim yöntemleriyle KL'yi karşılaştırdıkları bu çalışmada 2PL iki boyutlu 176 maddelik bir madde havuzu kullanarak simülasyon çalışması gerçekleştirmişler, madde havuzunu da 6 içerik ve 3 yetenek grubuna ayırmışlardır. Sabit madde sayısı $k = 50$, son yetenek kestirimi olarak ise EAP kullandıkları bu araştırmada herhangi bir madde teşhir oranı kontrol algoritması kullanılmamıştır. Bu nedenle, testlerin %80'inden fazlasında madde havuzunun %20'si kullanılmıştır. Havuzun %65'inin ise ancak testlerin %10'unda kullanıldığı görülmüştür. KL madde seçim yönteminin Fisher'in bilgi yönteminden daha iyi sonuçlar verdiği, KL madde seçim yönteminin gölge testlerde kullanılabilir olduğu bulgusuna ulaşılmıştır.

Wang & Chen (2004) araştırmalarında, Segall (1996)'in yaptığı çalışmadaki çok boyutluluğa atıfta bulunarak kullanılan çoklu-tekboyutluluğun (maddeler-arası

boyutluluk modeli) getirdiği avantajlar yanında dezavantajların da üzerinde durmuşlar, madde-içi boyutluluğun da hesaba katıldığı bir çalışma yapılmamış olmasından yola çıkarak her iki boyutluluk durumu için de Rasch modeli kullandıkları bir araştırma yürütmüşlerdir. Bu çalışmada; çok boyutlu, tek boyutlu ve random uygulamaları, madde-içi ve maddeler-arası boyutluluk modellerini, iki ve çok kategorili ölçekleri, boyutlar arası korelasyonların orta ve yüksek olduğu durumları ve uygulanan toplam madde sayısını manipüle ederek bir simülasyon çalışması yürütmüşlerdir. Bağımlı değişken ölçme verimliliği (etkililiği) olup, bunun göstergesi olarak da aynı güvenilirliğe erişmek için kullanılan madde sayısı belirlenmiştir. Maddeler-arası boyutluluk için 5 tek boyutlu test, madde-içi boyutluluk içinse 6 boyutlu testler kullanılmıştır.

Bifaktör modellerin BBT uygulamaları üzerine ilk çalışma ise Weiss ve Gibbons (2007)'a aittir. Bu çalışmada, Cassona ve diğerleri (1998) tarafından geliştirilen Kaygı Durumu Ölçekleri (The Mood-Anxiety Spectrum Scales) kullanılmıştır. Toplam 615 madde kullanılarak genel ve dört ayrı grup faktörü için olmak üzere toplam beş boyut için kestirim yapılmıştır. Maddelerin 800 kişiye uygulanmasıyla yapılan kalibrasyonun ardından yaklaşık beş ay sonra aynı gruptan 156 kişiye gerçek zamanlı BBT uygulaması yapılmıştır. BBT uygulaması için başlangıç noktası olarak $\theta = 0$, madde seçimi için Fisher'in maksimum bilgi yöntemi, yetenek kestirimi için MAP ve sonlandırma kuralı olarak da standart hataya dayalı sonlandırma kullanılmıştır. Algoritma olarak genel θ için yapılan ilk kestirim sonucunda uygulanan maddelerden ilk grup faktör ile ilgili olanlar belirlenerek bu maddelere dayalı yeni bir kestirim yapılmış ve bu kestirim söz konusu grup faktörü için başlangıç noktası olarak kabul edilmiştir. Bir anlamda bu uygulama, tüm boyutlar için ayrı ayrı yapılmıştır. Bireylerin kağıt-kalem sınavında verdiği yanıtlara dayalı post-hoc simülasyon sonucu kestirilen θ değerleri ile gerçek zamanlı BBT uygulaması sonucu elde edilen θ değerleri arasında .75 ile .85 arasında değişen korelasyon değerleri elde edilmiştir. Bunun yanında BBT algoritmasıyla tüm faktörlerde %80-%85 daha az madde kullanıldığı ve uygulama zamanının %82 oranında azaldığı sonucuna ulaşılmıştır.

2009 yılındaki çalışmalarında Mulder & van der Linden, farklı koşullarda A, D, E ve Ds-optimality madde seçim yöntemlerinin etkililiğini araştırmışlardır. Bu amaçla; çok boyutlu bir yapının var olduğu durumlarda tüm boyutlara ait yetenek kestiriminin

maksatlı (intentional) olduğu, kiminin maksatlı kiminin ise “nuisance” faktör¹ olduğu ve testle kestirilen tüm yeteneklerin maksatlıyken bunların lineer kombinasyonunun önemli olduğu durum olmak üzere toplam üç farklı durum için simülasyon çalışması gerçekleştirilmiştir. İki boyutlu hipotetik bir yapıya ait 200 maddelik bir havuz için güçlük ve ayırt edicilik parametresi üretilmiştir. Yetenek kestirimi olarak ise ML kullanılmıştır. Değerlendirme, yanlılık (bias) ve ortalama hata değerlerine bakılarak yapılmıştır. Araştırma sonuçlarına göre, testle kestirilen tüm yeteneklerin maksatlı olduğu koşulda A ve D-optimality yöntemleri en iyi sonuçları verdiği, E-optimality yönteminin ise beklendiği üzere kötü sonuçlar verdiği görülmüştür. Bir boyutun maksatlı diğer boyutun ise “nuisance” olduğu ikinci koşulda; ilk theta için A ve D-optimality yöntemleriyle elde edilen ortalama hatalar, tüm boyutların maksatlı olduğu durumda theta-1 için elde edilen değerlere göre daha düşüktür. Tüm boyutların lineer kombinasyonun elde edildiği üçüncü ve son koşulda ise ağırlıklandırılmış C-optimality yöntemi en iyi sonucu vermiş, ancak Ds-optimality ile elde edilen değerler de kimi koşullarda ağırlıklandırılmış C-optimality ile elde edilenlere çok yakın değerler vermiştir.

Wang, Chan ve Boughton (2012) çalışmalarında çok boyutlu BBT uygulamalarında kullanılabilecek D, E, T ve K-kuralı adını verdikleri dört yeni sonlandırma kuralını tanıtmışlardır. Araştırmacıların amacı, söz konusu sonlandırma kurallarının performanslarını farklı koşullar altında karşılaştırmak değil, her birinin önceden belirlenen kesme noktasına göre testi sonlandırıp-sonlandıramayacağını belirlemektir. Bu amaçla üç farklı simülasyon çalışması yürütülmüştür. İlkinde, hipotetik iki boyutlu yapıdaki 450 maddeye ait 2PL model için parametre üretilerek D-optimality, minimum varyans ve KL uzaklığı madde seçim yöntemleri ve farklı test uzunluklarının (5, 10, 15 ve 25 madde) kullanıldığı durumlarda söz konusu dört sonlandırma kuralının çalışma durumunu incelenmiştir. İkinci simülasyon çalışması ise gerçek veriye dayanmaktadır. 2 boyutlu, 3PL doğrulayıcı modele göre 200 madde kalibre edilmiş, ilk simülasyonda farklılaşma olmadığı için sadece D-optimality madde seçim yöntemi söz konusu dört sonlandırma kuralıyla

¹ Alanyazında; bifaktör modellerde, grup faktörlerinin “nuisance” faktör olarak da tanımlandığı görülmektedir. Kelime anlamı olarak “sıkıntı, dert, baş belası” gibi anlamlar taşıyor olması nedeniyle metinde İngilizce terimin kullanması tercih edilmiştir. Söz konusu terimin, grup faktörlerinin kestirimine önem verilmediği durumlarda kullanıldığı göz önüne alınmalıdır.

kullanılmıştır. Sonuçta, D ve T kuralı ile aynı test uzunluğu yakalanırken, E sonlandırma kuralıyla nispeten daha uzun testler uygulanmıştır. En kısa testler ise K kuralının uygulandığı koşula aittir. Üçüncü ve son simülasyon “özel durumlar”a aittir. Hipotetik üç boyutlu 480 maddelik havuz için parametre üretilmiş, boyutlardan ilk ikisinin maksatlı sonuncusunun ise “nuisance” faktör olduğu durumda dört sonlandırma kuralının beklenen biçimde çalışması durumu incelenmiştir. Tüm analizlerden sonra, tüm koşullarda KL sonlandırma kuralının önsel dağılımın yoğunluğundan etkilendiği, aynı zamanda diğer yöntemlere göre çok dramatik farklılıklar yaratmasa da madde havuzunu daha etkin kullandığı görülmüştür. Son olarak D, E ve T kurallarının dağılımın iki ucundaki yetenek düzeylerindeki bireyler için orta düzeydekilere göre yaklaşık iki kat fazla madde ile testi sonlandığı sonucuna ulaşılmıştır.

Seo (2012) yaptığı doktora tez çalışması; o dönemde bifaktör modele dayalı BBT simülasyon çalışması yürütülebilecek herhangi bir program ya da R’de tanımlı bir paket olmadığından, R’de araştırmacı tarafından yazılmış kodlarla yürütülmüştür. Söz konusu monte carlo simülasyon çalışması, iki kısımdan oluşmaktadır. İlk bölümde, Weiss & Gibbons (2007) tarafından geliştirilen bifaktör BBT algoritması ile araştırmacı tarafından geliştirilen çok boyutlu bifaktör BBT algoritmasını karşılaştırmıştır. Weiss & Gibbons tarafından geliştirilen söz konusu algoritma, öncelikle yanıtlanan tüm maddelere dayanarak genel yetenek için kestirim yaparken, grup faktörler için yetenek kestirimini genel faktörün ardından sırayla yapmaktadır. Bir başka ifadeyle bu yöntemde, genel yetenek ve grup faktörlerine ait kestirimler simultane olarak yapılmamaktadır. Bu bölümdeki; üç yetenek kestirim, iki ve dört grup faktörü olmak üzere iki farklı faktöriyel yapı ve her faktör için sabit sayıda 20 maddenin olduğu koşullar kullanılmıştır. Madde havuzları için, 400’er maddenin parametreleri 2PL modele göre üretilmiştir. Araştırma kapsamında 1000 kişi için theta üretilmiştir. Son olarak bu bölümde, grup faktörlerine ait ayırt edcilik parametreleri düşük, orta ve yüksek olarak üretilmiştir. Bağımlı değişkenler ise, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon, RMSE ve standart hata değerleridir. Elde edilen sonuçlara göre, araştırmacı tarafından geliştirilen çok boyutlu bifaktör algoritma, her boyutu ayrı ayrı ölçen modele göre çok daha yüksek theta korelasyon ve düşük hata değerleri vermiştir. Grup faktörlerine ait ayırt edcilik değerlerinin düşük olduğu durumlarda bu boyutlar için elde edilen korelasyon .70

civarı iken, yüksek ayırt edicilik değerleri ile korelasyonlar .80'in üzerine çıkmıştır. Araştırmanın ikinci bölümünde ise; madde sayısı, birey sayısı, yetenek kestirim yöntemleri ilk çalışmayla aynıdır. Farklı olarak, standart hataya dayalı sonlandırma kullanılmıştır. elde edilen sonuçlara göre, standart hata değeri azaldıkça genel ve grup faktörleri için elde edilen korelasyon değerlerinin de arttığı bulgulanmıştır.

Huang, Chen & Wang (2012)'ın, çok boyutlu BBT çalışmaları, yüksek dereceden MTK modellerinin BBT uygulamalarında kullanımı üzerinedir. Yazarlar, testlet tepki modelini yüksek derece MTK modellerinin özel bir hali olarak görmekle birlikte, bu özel hale ait madde seçim yöntemlerinin yüksek derece MTK modellerinde kullanılamayacağını düşünmüşler ve Segall (1999, 2004)'in madde seçim yöntemlerinden yola çıkarak bunları yüksek derece MTK modellerine uyarlamışlardır. Çalışma kapsamında, üç tane birinci dereceden ve bir tane ikinci dereceden gizil değişken kullanarak oluşturulan yüksek derece MTK modeliyle iki farklı simülasyon uygulaması yapılmıştır. İlkinde; test uzunluğu (30 ve 60 madde), madde havuzu (200 ve 400 madde), MTK modeli (2PL ve 3PL) ve birinci dereceden gizil değişkenlerin faktör yüklerinin farklı olduğu durumlar manipüle edilerek, yazarlar tarafından yüksek derece çok boyutlu BBT algoritması için adapte edilen Fisher'in maksimum bilgi yönteminin farklı durumlardaki etkililiği incelenmiştir. İkinci çalışmada ise maksimum bilgiyle birlikte yenilikçi (progressive) ve oranlı (proportional) madde seçim yöntemlerinin hangisinin madde teşhir oranı ve içerik ağırlıklandırılmasıyla en iyi sonucu verdiği sınanmıştır. İlk simülasyon sonuçlarına göre test uzunluğu, madde havuzu ve birinci dereceden gizil değişkenlere ait faktör yükleri arttıkça ölçme keskinliğinin de arttığı, 2PL modelin 3PL modele göre daha iyi sonuçlar verdiği görülmüştür. İkinci simülasyon sonuçlarına göre ise yenilikçi (progressive) madde seçim yönteminin madde teşhir oranı, içerik ağırlıklandırması, test güvenliği ve madde havuzunun kullanımı açısından en iyi sonuçları verdiği görülmüştür.

Seo ve Weiss (2015), çalışmalarında çok boyutlu BBT uygulamalarının tarihsel gelişmelerinden bahsederek maddelar-arası boyutluluğun psikolojik yapıların doğasına aykırılığı, madde-içi boyutluluğa ise faktörlerin tanımlanabilirliği ile ilgili getirilen eleştiriler üzerinde durmuşlardır. Araştırmacılar, benzeri durumlarda testin genel bir yeteneği ölçerken bir yandan da yapının çok boyutluluğunu göz ardı etmeyen yöntemlerin uygulanması gerektiğini belirtmişlerdir. Böyle durumlarda akla

ilk olarak ikinci derece/yüksek derece modeller gelmektedir. Ancak, Huang ve diğerlerinin (2012) yaptığı çalışmaya atıfta bulunularak, bu modellerin MTK uyarlamalarında ikinci dereceden gizil değişken için parametre kestiriminin getirdiği problemler üzerinde durulmuş ve bu durumda bifaktör modellerin getirdiği çözüm tartışılmıştır. Simülasyonlar için tüm veri setlerinde 2PL model 3PL'ye göre iyi sonuçlar verdiğinden bu model tercih edilmiştir. Bifaktör modele dayalı simülasyon sonuçlarına göre, eğer genel yetenekle ilgili daha dakik kestirimler yapılmak isteniyorsa Ds-rule madde seçim yöntemi ile EAP yetenek kestiriminin, grup faktörlerin önemsendiği durumda ise D-rule madde seçim yöntemi ve yetenek kestirim yöntemi olarak da MAP'ın kullanılmasının uygun olduğu sonucuna varılmıştır. İleride yapılacak çok boyutlu bifaktör BBT çalışmalarında ise KL madde seçim yönteminin uygunluğunun test edilmesini ve bunun yanında bifaktör modellerde uygulanabilirliğinin tartışılması önerilmiştir.

2.2. BBT Uygulamaları ile İlgili Yurtiçinde Yapılan Çalışmalar

Ülkemizde, çok boyutlu BBT uygulamalarına yönelik yürütülen çalışmaların sayısı oldukça azdır. Bu nedenle bu bölümde, boyutluluk ayrımı yapılmaksızın literatürde yer alan çalışmalar raporlanmıştır.

Türkiye'de BBT ile ilgili yapılan çalışmalar Kaptan (1993)'in araştırmasıyla başlamıştır. Araştırma kapsamında sayısal yeteneği kestirmek amacıyla geliştirilen BBT ile kağıt-kalem uygulamaları; uygulama zamanı, kullanılan madde sayısı ve elde edilen yetenek kestirimleri açısından karşılaştırılmıştır. Bunun yanında bireylerin bu iki testten elde edilen yetenek kestirimleri ile dış ölçüt olarak belirlenen ÖYS matematik testinden elde edilen yetenek kestirimleri karşılaştırılmıştır. Çalışma kapsamında sayısal yeteneğin ölçülmesi amacıyla araştırmacı tarafından 150 soruluk bir havuz oluşturulmuştur. Bu maddeler 404 kişiye uygulanarak elde edilen veriler Rasch modele göre kalibre edilmiş ve 105 maddenin havuzda kalmasına karar verilmiştir. Yine bu maddeler arasından 50 soruluk bir test hazırlanarak kağıt-kalem sınavında kullanılmıştır. Madde seçim algoritması olarak her bir yanlış cevabın ardından daha kolay, doğru cevabın sonrasında ise daha zor bir madde yöneltilmiştir. BBT uygulaması için gerekli program araştırmacı tarafından geliştirilmiştir. 40 kişiye yapılan bu uygulama sonucunda BBT ile yapılan uygulamaların kağıt-kalem testine göre daha kısa sürede ve daha az maddeyle gerçekleştirildiği, iki testle yapılan yetenek kestirimleri arasında anlamlı bir fark

olmadığı görülmüştür. BBT ve kağıt-kalem uygulamalarının dış ölçüt olan ÖYS matematik testiyle korelasyonları ise anlamlı bulunmuş olup, bu korelasyon değerleri arasındaki ise anlamlı bir fark olmadığı sonucuna ulaşılmıştır.

İşeri (2002), öğrencilerin matematik başarısının BBT ile kestirilmesine yönelik yaptığı çalışmada farklı BBT stratejilerinin karşılaştırılmasının üzerinde durmuştur. İki bölümden oluşan çalışmanın ilk kısmında gerçekleştirilen simülasyon çalışmasıyla farklı yetenek kestirim, madde seçim yöntemi ve sonlandırma kurallarının hangisinin eldeki havuza en iyi uyumu gösterdiği belirlenmiştir. Araştırmanın ikinci kısmında ise, simülasyon çalışması sonucu belirlenen en iyi BBT koşulu kullanılarak 9. sınıfta öğrenim gören 54 öğrenciye gerçek uygulama yapılmıştır. Uygulamada OKS ve Özel Okullar Sınavının matematik maddeleri kullanılarak 3PL modele uygun olarak oluşturulan soru havuzu kullanılmıştır. Elde edilen sonuçlara göre BBT uygulamasıyla kağıt-kalem sınavına göre daha az maddeyle benzer yetenek kestirimleri elde edilmiştir. Bayes kestirim yöntemi ML'den daha iyi sonuçlar verirken sabit hataya ve sabit madde sayısına dayalı sonlandırma kurallarıyla benzer ve yüksek verimli sonuçlar elde edildiği görülmüştür. Ayrıca BBT ile elde edilen yetenek kestirimleri ile dış ölçüt olarak belirlenen öğrencilerin matematik ders notları arasında da yüksek korelasyon olduğu sonucuna varılmıştır.

Cömert (2008)'in çalışması psikometrik bir çalışmadan çok BBT yazılımı geliştirmeye yöneliktir. Araştırmacının bu amaçla geliştirdiği yazılım MTK stratejilerini kullanmaz. Çalışmada kullanılan soru havuzundaki maddelerin güçlüğü, herhangi bir kalibrasyon yapılmadan araştırmacı tarafından 1-10 arasında değer verilerek belirlenmiştir. Araştırmada Thorndike (1982)'nin çalışmasında yer alan sıralı, çok düzeyli, iki aşamalı, piramitsel, kendinden seçimli ve değişmeli olmak üzere altı uygulama stratejisi kullanılmıştır.

Kalender (2011), Öğrenci Seçme Sınavı fen bilimleri alt testini kullanarak gerçekleştirdiği BBT uygulamasını kağıt-kalem uygulamasıyla karşılaştırmıştır. Araştırmanın ilk kısmında, farklı yıllarda ÖSS fen bilimleri alt testinde uygulanan maddeler kullanılarak oluşturulan madde havuzu için en uygun BBT algoritmasının belirlenmesi amacıyla *post-hoc* simülasyon çalışması yapılmıştır. Yetenek kestirim yöntemlerinden MLE ve EAP, madde seçim yöntemi olarak maksimum olabilirlik, sonlandırma kurallarından ise standart hata ve sabit madde sayısına dayalı

kuralların hangisinin en iyi sonuçları verdiği tespit edilmiştir. İkinci aşamada ise, araştırmacı tarafından geliştirilen yazılım kullanılarak *post-hoc* simülasyon sonucu belirlenen en uygun BBT koşulu çerçevesinde 33 kişiye gerçek uygulama yapılmıştır. Bu bireylere uygulanan kağıt-kalem testi ise 30 maddeden oluşmaktadır. Elde edilen sonuçlara göre BBT ve kağıt-kalem uygulamalarıyla elde edilen yetenek kestirimleri arasında oldukça yüksek bir korelasyon bulunmuş, BBT ile yapılan uygulamalarda uygulanan madde sayısının geleneksel yöntemle göre %50 oranında daha az olduğu saptanmıştır.

Kaskatı (2011) çalışmasında Romatoid Artrit hastalarının özürüllük durumunun değerlendirilmesi için Rasch modele dayalı bir BBT uygulaması yapmıştır. Likert tipindeki 113 madde kullanılarak 270 bireye yapılan uygulama sonucu bir madde havuzu oluşturulmuştur. Madde havuzu iki boyutlu olmakla birlikte, bu boyutlarda yer alan maddeler kendi içinde uygulanarak kalibre edilmiştir. Bir başka ifadeyle, tek boyutlu iki madde havuzu oluşturulmuştur. Kalibrasyon sürecinin ardından havuzda iki boyut için toplam 99 madde yer almıştır. Araştırmanın simülasyon kısmında 1000 birey için gerçek madde parametrelerine dayalı yanıt üretilerek, EAP ve ML yetenek kestirim yöntemleriyle elde edilen yetenek kestirimleri karşılaştırılmıştır. Gerçek zamanlı BBT uygulamasında ise 29 hastaya uygulama yapılmıştır. Kağıt-kalem formatındaki testlerde ilk boyut için 15 madde uygulanırken, ikinci boyut için uygulanan madde sayısı 14'tür. İlk boyut için BBT uygulaması ile uygulanan madde sayısı kağıt-kalem uygulamasına göre %50 azalırken, ikinci boyut için bu değer yaklaşık %30'dur. İki boyut için de elde edilen korelasyon değerlerinin oldukça yüksek olduğu sonucuna ulaşılmıştır.

Bulut ve Kan (2012) tarafında yapılan çalışma, Akademik Lisansüstü Eğitime Giriş Sınavının (ALES) BBT olarak uygulanabilirliği üzerinedir. Bu amaçla araştırmacılar bir *post-hoc* simülasyon çalışması gerçekleştirmişlerdir. Araştırma kapsamında 2008 yılında uygulanan ALES'e giren öğrencilerden rastgele 10000 kişinin yanıtları kullanılarak sayısal-1, sayısal-2 ve sözel testleri ayrı ayrı tek boyutlu 3PL modele göre kalibre edilerek madde havuzu oluşturulmuştur. Yetenek kestirimi olarak uygun yöntem olarak EAP belirlenmiş olsa da simülasyonlarla bu yöntem ML, WLS ve MAP ile elde edilen sonuçlarla karşılaştırılmıştır. Madde seçim yöntemi olarak maksimum bilgi kullanılırken sonlandırma kuralı üç farklı standart hata değerine dayalı olarak yapılmıştır. Simülasyonlar sonucunda bireylerin BBT ve kağıt-kalem uygulamasıyla

kestirilen yetenek düzeyleri arasında yüksek korelasyon elde edildiği görülmüştür. Her ne kadar yetenek skalasının iki ucundaki bireylerin madde havuzundaki maddelerin tamamını yanıtlaması gerekse de, orta düzeyde ve onlara yakın yetenekteki bireylerin yanıtladığı madde sayısının kağıt-kalem sınavlarına göre oldukça az olduğu sonucuna ulaşılmıştır.

Sulak (2013) araştırmasında, Han (2010) tarafından geliştirilen simülasyon programını kullanarak bir simülasyon çalışması yürütmüştür. Çalışma kapsamında beş farklı madde seçim yöntemi, iki yetenek kestirim yöntemi ve üç farklı sonlandırma kuralının kullanıldığı toplam 30 koşul elde edilmiştir. 250 hipotetik madde havuzuna ait parametre ve 2000 birey için theta üretilmiştir. Elde edilen sonuçlara göre, sabit madde sayısı ile durdurma kuralına göre ML yetenek kestirim yöntemi kullanıldığında elde edilen standart hata değerlerinin, EAP ile elde edilen standart hata değerlerinden daha yüksek bulunmuştur. ML yöntemine dayalı yetenek kestirimi kullanıldığı durumlarda ise kullanılan madde sayısı 30'un altında iken a-tabakalama yöntemiyle, 30'dan büyük olduğunda ise KL madde seçme yöntemiyle elde edilen standart hata değerleri daha büyüktür.

Eroğlu (2013) simülasyona dayalı olan çalışmasında Han (2010) tarafından geliştirilen SimulCAT programını kullanarak 250 ve 500 maddeden oluşan iki farklı madde havuzu için 3PL modele göre parametre ve 1000 kişi içinse theta üretilmiştir. Bu verilere dayanarak; sabit madde sayısına dayalı sonlandırma (15 ve 20), standart hataya dayalı sonlandırma (.385, .315 ve .220), standart hata ve en az madde sayısı olarak 10 maddenin belirlendiği durum, theta değerinde yakınsama (.05 ve .02) ve theta değerinde yakınsama ile en az madde sayısı olarak 10 maddenin kullanılması durumlarının belirlendiği farklı sonlandırma kuralları çerçevesinde simülasyon çalışması yürütülmüştür. Bu simülasyonlarda, yetenek kestirim yöntemi olarak da ML ve EAP yöntemlerinin performansları karşılaştırılmıştır. Elde edilen sonuçlara göre; sabit madde sayısının kullanıldığı durumda 20 madde ile sonlandırma, standart hata için .220 değeri ve theta değerindeki yakınsama içinse .02'nin kullanıldığı koşullarda yanlışlık ve RMSE değerleri daha düşükken, uyum katsayılarının önemli farklılıklar göstermediği görülmüştür. Bu durumlara en az madde koşulunun eklendiği koşullarda ise ölçme kesinliği artmıştır. Diğer bir bulgu ise madde havuzuyla ilgili olup, büyük madde havuzunda hata istatistiklerinin düştüğü bulgulanmıştır. Yetenek kestirim

yöntemlerinden EAP yönteminin kullanıldığı koşullardaki ölçme kesinliği, ML yönteminin kullanıldığı durumlardan daha yüksektir.

Kezer (2013), İngilizce kelime bilgisini ölçmek amacıyla gerçek-zamanlı bir BBT uygulaması yapmıştır. Bu amaçla araştırmacı tarafından PHP programlama dili kullanılarak çevrimiçi bir ortam oluşturulmuştur. Madde havuzu için kelime bilgisine yönelik hazırlanmış 100 madde, 994 öğrenciye uygulanarak soruların kalibrasyonu yapılmıştır. Bireyler için teste başlangıç noktası olarak yetenek ölçeğinin .5 sağı ve solundaki aralık belirlenmiş, ilk yedi madde yanıtlanana kadar bireylerin yetenek kestirimi yapılmamıştır. Söz konusu yedi maddenin yanıtlanmasının ardından bir sonraki madde, verilen cevaba göre daha zor ya da daha kolay olarak belirlenmiştir. Yedi maddenin yanıtlanması sonucu ML'ye göre yapılan yetenek kestirimi sonrası yeni madde seçimi maksimum bilgi yöntemine göre yapılmıştır. Sonlandırma kuralı olarak ise, son iki yetenek kestirimi arasındaki farkın .01'in altında olması ve .5 sabit hataya göre sonlandırma kurallarından herhangi birinin önce gerçekleşmesi belirlenmiştir. Araştırmanın ikinci kısmı ise; farklı başlatma, yetenek kestirimi ve sonlandırma kurallarının kullanıldığı toplam 18 koşul için gerçekleştirilen simülasyon çalışmasıdır. Araştırma sonucu bulgulara göre gerçek zamanlı BBT uygulaması ile kağıt-kalem sınavına kıyasla daha az madde ve zaman harcanarak yetenek kestirimlerinin yapıldığı, iki uygulama ile kestirilen yetenek düzeyleri arasındaki pozitif ve yüksek bir korelasyon olduğu görülmüştür.

Özbaşı (2014) tarafından yapılan çalışma, bilgisayar okur-yazarlığı testinin tek boyutlu BBT olarak uygulanması üzerinedir. Çalışma, simülasyon ve gerçek zamanlı uygulamaları içermektedir. 1366 kişiye yapılan uygulama sonucu 136 maddelik bir havuz oluşturulmuştur. Han (2010) tarafından geliştirilen SimulCAT simülasyon programı kullanarak, BBT uygulamasının farklı yetenek kestirim ve sonlandırma kuralı çerçevesinde nasıl performans gösterdiği incelenmiştir. Gerçek zamanlı BBT uygulamasında ise, Kalender (2011) tarafından geliştirilen BBT programı kullanılmıştır. Elde edilen sonuçlara göre, simülasyon uygulamasında EAP ve ML yetenek kestiriminin kullanıldığı koşullarda elde edilen değerlerin farklılık gösterdiği görülmüştür. Bunun yanında, sabit maddeye dayalı sonlandırma kuralı ve EAP yetenek kestiriminin kullanıldığı BBT koşuluyla elde edilen yetenek kestirimlerinin gerçek theta değerlerine daha yakın olduğu sonucuna varılmıştır. Elde edilen başka bir bulgu, EAP yetenek kestirim yöntemi kullanılarak elde edilen

güvenirlik değerlerinin daha iyi olduğu yönündedir. Gerçek-zamanlı uygulama sonucunda ise; gerçek ve BBT ile kestirilen theta değerleri arasında pozitif ve anlamlı bir korelasyon olduğu, bu yetenek kestirim değerleri arasındaki farkın $\theta < 0$ olduğu durumda büyükken, $0 < \theta < 2$ olduğu durumda ise azaldığı sonucuna ulaşılmıştır.

Boztunç-Öztürk (2014) tarafından yapılan çalışmada, madde kullanım sıklığını kontrol etmeye yönelik farklı yöntemlerin ölçme kesinliğine etkisi incelenmiştir. Araştırmada, farklı madde seçim yöntemleri ve farklı özellikteki madde havuzları kullanılmıştır. Madde kullanım sıklığını kontrol etmek için kullanılan yöntemler ise; bilgi koşullu tesadüfi, Sympson-Hetter ve azalarak kaybolma yöntemleridir. Araştırma kapsamında kullanılan madde seçim yöntemleri maksimum bilgi, a-tabakalama ve aşamalı en çok bilgi oranı yöntemidir. Elde edilen bulgulara göre, a-tabakalama yönteminin madde kullanım sıklığının kontrol edilmediği durumlarda madde havuzunu en efektif kullanan madde seçim yöntemi olduğu sonucuna ulaşılmış; bununla beraber madde kullanım sıklığının kontrol edildiği durumlarda, edilmediği durumlara göre hata istatistiklerinde büyük farklılıkların bulunmadığı sonucuna ulaşılmıştır.

Özdemir (2015) tarafında yapılan çalışma, ülkemizde çok boyutlu BBT uygulamalarına yönelik yapılan ilk çalışma olma özelliğini taşımaktadır. Bu çalışmada araştırmacı, Hacettepe Üniversitesi YDO tarafından uygulanan yeterlik sınavının verilerini kullanarak bir post-hoc simülasyon çalışması yürütmüştür. R'de tanımlı MAT (2011) paketi kullanılarak yürütülen bu çalışmada madde-içi ve maddeler-arası boyutluluk modelleri kullanılarak farklı madde seçim, yetenek kestirim ve sonlandırma kuralları çerçevesinde toplam 72 koşulun performanslarını karşılaştırmıştır. Farklı boyutluluk modellerine göre kalibre edilmelerinin ardından 2PL modele göre kalibre edilen maddelerle, 565 ve 559 maddelik iki havuz oluşturulmuştur. Yetenek kestirim yöntemi olarak MAP ve Fisher'in puanlama yöntemi; madde seçimi için A-optimality, D-optimality ve seçkisiz yöntem; sonlandırma kuralı olarak ise sabit madde sayısı ve hata varyansı durdurma kuralı kullanılmıştır. Elde edilen sonuçlara göre, A-optimality madde seçim yöntemi, her iki boyutluluk modelinde de en iyi istatistikleri vermiştir. Genel olarak, çok boyutlu BBT uygulaması ile kağıt-kalem testine göre daha güvenilir kestirimlerin daha az madde kullanılarak elde edildiği görülmüştür. Tüm koşullar karşılaştırıldığında ise en iyi

koşulun; A-optimality madde seçim yöntemi, MAP yetenek kestirim yöntemi ve boyutluluk modeli açısından da madde-içi boyutluluk modelinin kullanıldığı koşulda elde edildiği bulgulanmıştır.

Aybek (2016) çalışmasında Kendini Değerlendirme Envanteri'nin BBT olarak uygulanabilirliğini sınamıştır. Araştırma sonunda amaç, bir BBT uygulamasıyla öğrencilerin profillerini ortaya koymaktır. Her biri 10 maddeden oluşan 23 boyutluk ölçeğin BBT uygulamasında, her bir boyut için ayrı ayrı analiz yapılmıştır. Araştırmanın post-hoc simülasyon uygulaması için yanıtlanan madde sayısı en az üç olacak biçimde .30, .40, .50 standart hataya dayalı sonlandırma, dört farklı madde seçim ve EAP yetenek kestiriminin kullanıldığı koşullardan elde edilen istatistikler ile gerçek ve kestirilen theta değerleri arasındaki korelasyonlara bakılmıştır. Simülasyon sonuçlarına dayalı olarak gerçekleştirilen gerçek zamanlı BBT uygulamasında; madde seçim yöntemi olarak maksimum bilgi, EAP yetenek kestirimi ve .40 standart hataya dayalı sonlandırma kuralı belirlenmiştir. Uygulama, R'de çalışan Concerto platformunda gerçekleştirilmiştir. Bulgulara göre, simülasyon çalışmalarında kullanılan madde seçim yöntemlerinin değişimiyle elde edilen istatistikler arasında önemli farklılıklar görülmemektedir. Gerçek zamanlı BBT uygulaması ile kağıt-kalem uygulamasına göre %52 daha az madde yanıtlanmış, testleri yanıtlama süresinde ise %60'lık bir azalma görülmüştür. 23 boyuttan her biri için kağıt-kalem ve BBT uygulamasıyla elde edilen sonuçlardan üçü arasında düşük korelasyon olduğu görülürken, iki teste göre bireylerin belirlenen profilleri incelendiğinde ise öğrencilerin 20'sinde uyum sağlandığı sonucuna ulaşılmıştır.

Gündeğer & Doğan (2016) tarafından yapılan çalışma, Türkiye'de bireyselleştirilmiş bilgisayarlı sınıflama testleri üzerine yapılmış ilk çalışmadır. Bu çalışma, madde havuzunun farklı dağılımlar göstermesi ve farklı büyüklüklere sahip olması durumunda test uzunluğu ve ortalama sınıflama doğruluğunun nasıl etkilendiği inceleyen bir monte carlo simülasyon çalışmasıdır. Bu amaçla, sivri ve basık dağılım gösteren dörder tane farklı büyüklükteki madde havuzu ve üç farklı madde seçim yönteminin etkisi incelenmiştir. R'de yapılan simülasyon uygulaması için kodlar araştırmacı tarafından yazılmıştır. Madde havuzundaki madde parametreleri, 3PL modele göre üretilmiştir. Madde seçim yöntemi olarak; tesadüfi, maksimum bilgi ve KL yöntemleri kullanılmıştır. Ortalama sınıflama doğruluğu için, Cohen'in Kappa istatistiğinden yararlanılmıştır. Toplam 24 koşulun elde edildiği çalışma sonuçlarına

göre en uzun testler, tesadüfi madde seçim metodunun kullanıldığı durumda elde edilmiştir. Diğer iki madde seçim yöntemiyle elde edilen sonuçlar ise birbirine oldukça benzerdir. Madde havuzu büyüdükçe test uzunluğuyla birlikte sınıflama doğruluğunun da azaldığı sonucuna varılmıştır. Madde havuzunun yapısı bakımından, sınıflama doğruluğunu sivri dağılımların arttırdığı görülmüştür. Tüm koşullarda .94 ile .99 arasında değişen oldukça yüksek sınıflama doğruluğu değerleri elde edilmiştir.

Akyıldız & Şahin (2016) Anadolu Üniversitesi YDYO İngilizce Yeterlilik Sınavının (İYS) kelime bilgisi bölümünü gerçek zamanlı BBT olarak uyguladıkları çalışmalarında, Chalmers (2016) tarafından geliştirilen R’de tanımlı mirtCAT paketini kullanmışlardır. Her eğitim-öğretim yılı başı ve sonunda gerçekleştirilen ve her biri 100’er maddeden oluşan İYS’nin farklı yıllardaki formlarından elde edilen kelime bilgisi maddeleri ayrı ayrı kalibre edilmiştir. 25’er sorudan oluşan her bir madde setine ait verilerin kalibrasyonu ile madde ve yetenek parametrelerinin değişmezliği sayıtlarının test edilmesi için flexMIRT ve jmetrik kullanılmıştır. Tüm veri setleri için en uygun modelin 3PL olduğu bulgulanmıştır. 100 maddeden oluşan soru havuzuyla gerçekleştirilen gerçek zamanlı BBT uygulaması için başlangıç noktası $\theta=0$, madde seçim yöntemi olarak maksimum bilgi, yetenek kestirim yöntemi olarak EAP ve sonladırma kuralı olarak ise .45 standart hata değeri belirlenmiştir. Linux işletim sistemine sahip bir bilgisayarda kurulan sunucuyla söz konusu paket R’de tanımlı shiny (2014) web arayüzü ile birlikte kullanılmıştır. Elde edilen sonuçlara göre testin sonlandırılması için gerekli madde sayısının 4 ile 13 arasında değiştiği görülmüştür. Madde seçim yöntemi olarak maksimum bilgi kullanılması ve kimi maddelerin bazı yetenek düzeylerinde çok bilgi vermesi nedeniyle bazı bireylerin testlerinin çok erken sonuçlandığı bulgusuna ulaşılmıştır.

Yurtdışında, çok boyutlu BBT uygulamalarına yönelik çalışmalar 1970’lerde başlamış olmakla birlikte, literatür açısından kritik önem taşıyan çalışmaların 1990’lı yılların sonlarında başladığı görülmektedir. Son 10 yılda çok boyutlu BBT’nin farklı bileşenleri üzerine birçok çalışma yapılmış olmakla birlikte, özellikle bifaktör model kullanılarak çok boyutluluğun göz ardı edilmeden genel bir yetenek kestirimi yapılan çalışmaların sayısının az olduğu görülmektedir. Bifaktör modele dayalı olarak yapılan BBT çalışmalarından ilki olan Weiss & Gibbons (2007)’a ait çalışmanın, her ne kadar gerçek-zamanlı bir uygulama yapılmış olsa da tam anlamıyla bir çok

boyutlu uygulama olmadığı görülmektedir. Tüm boyutlara ait kestirimlerin simültane olarak yapıldığı çalışmalar olan Seo (2012) ve Seo & Weiss (2015)'e ait çalışmaların ise, *monte carlo* simülasyon çalışmaları oldukları bilinmektedir. Bu nedenle, bifaktör modele dayanan gerek *post-hoc* simülasyon, gerekse gerçek-zamanlı uygulamaya dönük araştırmalara ihtiyaç duyulduğu düşünülmektedir.

Yurtiçinde yürütülen çalışmalara dair yapılan literatür taraması, boyutluluk ayrımı yapmaksızın tüm BBT uygulamalarını kapsamaktadır. Ülkemizde BBT'ye yönelik olan ilk çalışma, Kaptan (1993) tarafından yapılmış bir gerçek-zamanlı BBT uygulamasıdır. Ancak bu çalışma sonrasında yürütülen araştırmaların uzun bir süre zarfında sayısının oldukça az olduğu, yapılan sınırlı sayıdaki çalışmalar arasında da psikometrik yönü oldukça zayıf olan, MTK'nın kullanılmadığı ve ağırlıklı olarak BBT uygulamalarını bir bilgisayar yazılımı olarak gören uygulamaların yer aldığı dikkat çekmektedir (Cömert, 2008 vb.). Bununla birlikte 2010'lu yılların başından itibaren BBT çalışmalarının ivme kazandığı dikkat görülmektedir. Bu dönemde, gerçek-zamanlı BBT uygulamalarına yönelik yaptıkları çalışmalarda bazı araştırmacılar, kendi geliştirdikleri yazılımları kullanmışlardır (Kalender, 2011, Kezer, 2013). Son dönemde ise, açık kaynak kodlu bir istatistik programı olan R'nin kullanımındaki artışın BBT uygulamalarına da yansısıyla Concerto platformuna dayalı bir uygulama Aybek (2016) tarafından gerçekleştirilmiştir. Bununla birlikte, ülkemizde henüz çok boyutlu gerçek-zamanlı bir BBT uygulamasının yapılmamış olmasının önemli bir eksiklik olduğu değerlendirilmektedir.

3. YÖNTEM

Bu bölüm; araştırmanın türü, verilerin elde edilmesi ve verilerin analizi kısımlarından oluşmaktadır.

3.1. Araştırmanın Türü

Bu araştırma, Anadolu Üniversitesi YDYO tarafından hazırlık eğitimlerini tamamlayan öğrencilere yönelik olarak hazırlanan ve uygulanan İngilizce Yeterlik Tespit Sınavı (YTS) verilerini kullanılarak gerçek-zamanlı çok boyutlu BBT uygulamasını gerçekleştirmeyi amaçlamaktadır. Bu doğrultuda, farklı zamanlarda kağıt-kalem formatında uygulanmış YTS'lerden elde edilen gerçek veri setleri kullanılarak, gerçek-zamanlı uygulamada kullanılacak çok boyutlu BBT algoritmasının belirlenmesi amacıyla bir *hibrit* simülasyon çalışması gerçekleştirilmiştir. Bir sonraki adımda ise belirlenen bu algoritma kullanılarak gerçek-zamanlı çok boyutlu BBT uygulaması yapılmıştır. Bu anlamda; literatürde yer alan yöntem ve tekniklerin gerçek veri üzerinde sınanması gerçekleştirildiğinden, bu araştırmanın temel araştırma türünde olduğu değerlendirilmektedir (Karasar, 2010).

3.2. Verilerin Elde Edilmesi

Araştırma kapsamında kullanılan veri setleri iki kısımdan oluşmaktadır. İlk veri seti, Anadolu Üniversitesi YDYO'da İngilizce hazırlık eğitimi alan öğrencilere dört farklı dönemin sonunda kağıt-kalem formatında uygulanan YTS'lerden elde edilmiştir. Bu sınavlardan elde edilen veriler kullanılarak madde havuzu oluşturulmuş ve *hibrit* simülasyon uygulaması gerçekleştirilmiştir. İkinci veri seti ise, simülasyon uygulamasındaki madde havuzu kullanılarak gerçek-zamanlı BBT uygulaması yapılan bireylerden elde edilen yetenek kestirimleri ile birlikte madde havuzuna dair bazı istatistikler ve bu bireylerin bir kısmına uygulanan kağıt-kalem sınavından elde edilen puanlardan meydana gelmektedir.

Madde havuzunun oluşturulmasında ve simülasyon uygulamasında verileri kullanılan YTS, hazırlık eğitimini tamamlayan öğrencilere her eğitim-öğretim yılı sonunda (Haziran ayında), bu sınavda başarılı olamayanlar için ise bir sonraki eğitim-öğretim yılı başında (Eylül ayında) olmak üzere yılda iki kez uygulanmaktadır. Her sınav öğrencilerin İngilizce gramer, kelime bilgisi, okuma ve

dinleme becerilerini ölçen ve her bir beceri için eşit sayıda madde içeren toplam 100 sorudan oluşmaktadır. Araştırma kapsamında 2014–2016 yılları arasında uygulanan dört YTS'ye ait maddeler kullanılmıştır. Ancak, okuma becerisini ölçen maddelerin aynı paragrafa bağlı olarak birden fazla maddenin yanıtlanması gerektirmesi ve çalışma kapsamında kullanılan Chalmers (2016) tarafından geliştirilmiş R'de tanımlı bir BBT paketi olan mirtCAT'ın dinleme becerisini ölçen maddelerin kullanılmasına olanak sağlamaması nedeniyle bu çalışma, İngilizce YTS'de yer alan gramer ve kelime becerilerini ölçmeye yönelik hazırlanmış maddelerle yürütülmüştür. Dolayısıyla çalışma kapsamında, 100 kelime bilgisi 100 gramer olmak üzere toplam 200 maddeden elde edilen veriler kullanılmıştır. Söz konusu dört YTS'ye 2014 Eylül ayında 415, 2015 Haziran ayında 692, 2015 Eylül ayında 798 ve 2016 Haziran ayında ise 1153 öğrenci girmiştir.

İkinci veri seti, Anadolu Üniversitesi YDYO'da öğrenim gören öğrencilere yapılan gerçek-zamanlı uygulama sonucu elde edilen; madde havuzuna dair bazı istatistikler, söz konusu gerçek-uygulamanın yapıldığı 99 öğrenciden elde edilen yetenek kestirimleri ve bu bireylerden 32'sine uygulanan kağıt-kalem testi sonuçlarından oluşmaktadır. Uygulamanın yapıldığı öğrencilerin tamamı A kurunda öğrenim görmektedir. Bu öğrencilerden 46'sı, ilk dönem itibarıyla A kuruna yükselmiş olup, bu nedenle Şubat ayında hazırlık eğitimini geçmek amacıyla yapılan ara dönem yeterlilik sınavına girme hakkı kazanmışlardır. Kalan 53 öğrenci ise, Mart ayı itibarıyla A kuruna yükseldiklerinden, ara dönemde hazırlık eğitimini geçmek amacıyla uygulanan yeterlilik sınavına girme hakları olmamıştır. Şubat ayında hazırlık eğitimini geçmek amacıyla yapılan sınava girme hakkı olan söz konusu 46 öğrenciden 32'sinin sınav sonuçları sağlıklı olarak elde edilebildiğinden, çevrimiçi olarak uygulanan gerçek-zamanlı çok boyutlu BBT uygulaması ile kağıt-kalem testi sonuçları arasındaki korelasyon, yalnızca bu 32 öğrenci için elde edilmiştir.

3.3. Verilerin Analizi

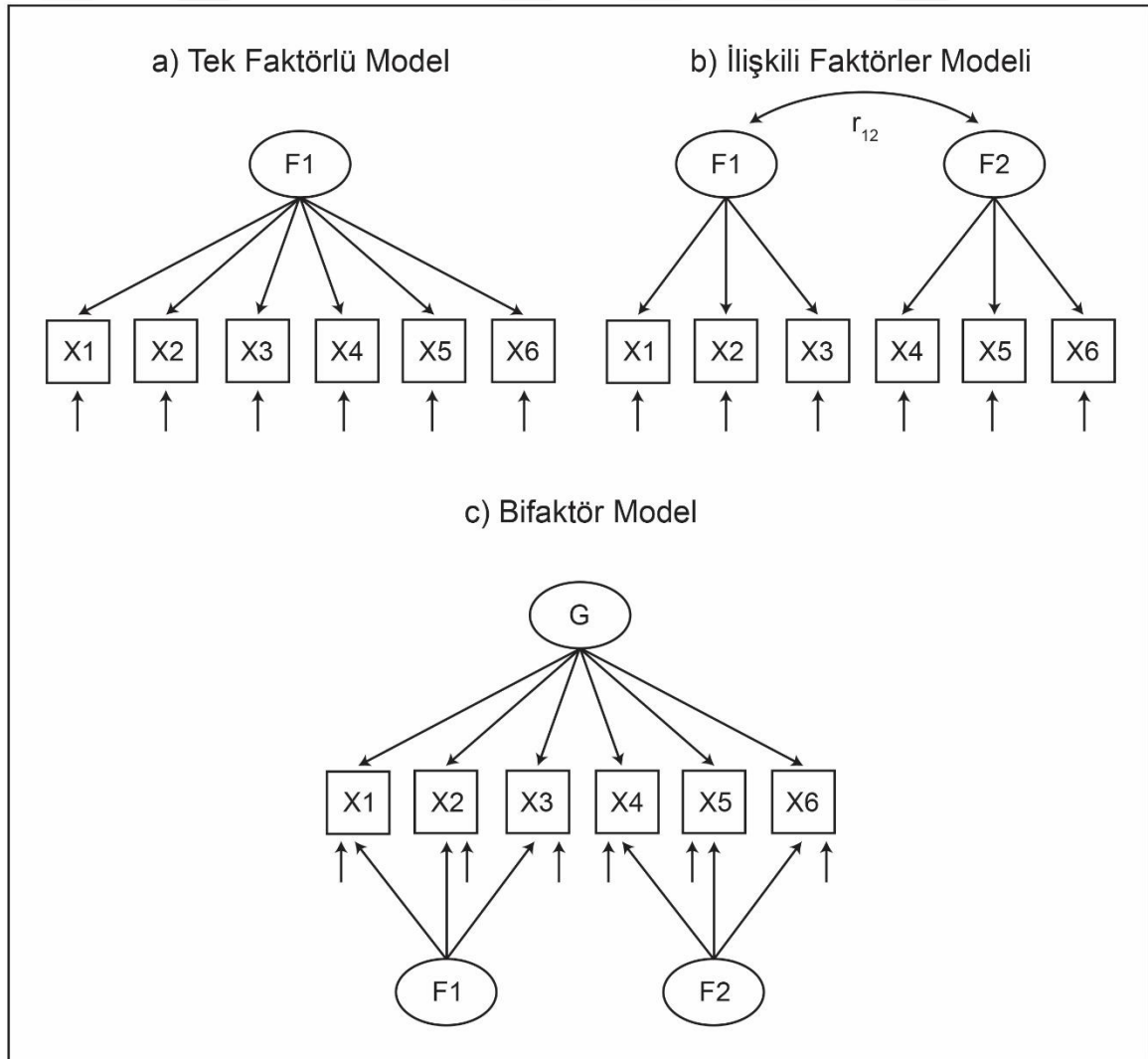
Çalışmanın amacı doğrultusunda verilerin analiz süreci, temel olarak üç aşamada incelenebilir. İlk aşama, çok boyutlu BBT uygulamasında kullanılacak madde havuzunun oluşturulmasına yönelik analizleri içermektedir. İkinci aşamada, 36 farklı koşul için *hibrit* simülasyon uygulaması yapılmış ve her bir koşul için ilgili istatistikler karşılaştırılarak gerçek-zamanlı çok boyutlu BBT uygulaması için en uygun koşul belirlenmiştir. Son aşamada ise belirlenen en uygun koşul çerçevesinde gerçek-

zamanlı uygulama yapılarak madde havuzunun kullanım durumu ve bireylerin bu uygulamayla kestirilen yetenek düzeyleri ile kağıt-kalem uygulaması sonucu kestirilen yetenek düzeyleri arasındaki ilişkiler hesaplanmıştır.

3.3.1. Madde Havuzunun Oluşturulması

3.3.1.1. Model – Veri Uyumunun Test Edilmesi

Simülasyon ve gerçek-zamanlı BBT uygulamasında kullanılacak madde havuzunun oluşturulması için, öncelikle her bir sınavdan elde edilen veri seti ayrı ayrı sırasıyla tek boyutlu model, iki boyutlu model ve üç boyutlu bifaktör modele uygun faktör analizine tabi tutulmuştur. Söz konusu modellerin teorik şekilleri aşağıda görülmektedir.



Şekil 3.1: Araştırma Kapsamında Kullanılan Faktör Analitik Modellerin Hipotetik Gösterimi

Dört madde setine uygulanan modellerden ilki, tek boyutlu/faktörlü modeldir. Tek boyutlu modelde, tüm maddelerin altında ortak bir varyans kaynağı söz konusudur

ve dolayısıyla tüm maddeler tek faktöre yük verirler. Her ne kadar araştırma kapsamındaki madde setlerinin çok boyutlu olduğu öngörülse de tutumluluk prensibi gereği öncelikle tek boyutlu model uygulanmıştır.

Uygulanan ikinci model, ilişkili faktörler modeli/iki boyutlu modeldir. Bu modelde tüm maddeler arasında ortak bir varyans kaynağı yoktur. Ayrıca maddeler, boyutlardan yalnızca birine, önsel olarak ölçtüğü düşünülen boyuta yük verirler. Bu modele, boyutlar arasında korelasyona izin verilmesi nedeniyle ilişkili faktörler modeli (correlated traits model) de denir. Madde düzeyinde boyutluluk bakımından ise bu modeller, MTK literatüründe maddeler arası boyutluluk (between-item dimensionality) gösteren modeller ya da çoklu-tek boyutlu (multi-unidimensional) modeller olarak anılırlar (Wang & Chen, 2004; Sheng & Winkle, 2007).

Tüm sınavlardan elde edilen verilere uygulanan son model ise bifaktör modeldir. Bifaktör modellerde; tüm maddelerin altında yatan ortak bir varyans kaynağı, “genel” bir yetenek/faktör (general trait) vardır. Bunun yanında her bir madde grup/spesifik/nuisance faktör olarak görülen bir faktöre daha yük verir. Bir başka ifadeyle, her bir madde genel faktörün yanında başka bir grup faktörünü daha ölçer. Böylece söz konusu grup faktörlerinin daha dakik/saf (pure) ölçülmesi sağlanmış olur. Bifaktör model, hipotetik olarak çok boyutlu olmakla birlikte, baskın bir genel faktör barındıran yapılar için uygundur. Bir başka ifadeyle bu model kullanılarak yapının çok boyutluluğu göz ardı edilmeden genel bir yetenek de kestirilmiş olur. Bifaktör modellerde, grup faktörleri arasında korelasyon olmadığı gibi genel faktörden de bağımsızdırlar. Eğer, analiz sonucu genel faktördeki yükler küçük, grup faktördekiler büyükse maddeler arası boyutluluk söz konusudur. Tam tersine genel faktördeki yükler büyükken grup faktördekiler küçükse, bu durum bifaktör modelin bir göstergesidir (Reise, Morizot & Hays, 2007; Gustafson & Aber-Bengtsson, 2010; Reise, Moore & Haviland, 2010; Seo, 2012; Seo & Weiss, 2015). Bu araştırma kapsamında tüm veri setlerine sırasıyla tek boyutlu model, iki boyutlu (ilişkili faktörler) model ve üç boyutlu bifaktör model uygulanarak elde edilen bulgular raporlanmıştır.

Geleneksel faktör analitik yöntemler, korelasyon matrisine dayalıdır. MTK’ya göre ise bireyin yetenek düzeyi ile soruyu doğru yanıtlama ihtimali arasında non-lineer (eğrisel) bir ilişki bulunur (Seo, 2012). Bu nedenle faktör analizleri, R’de tanımlı MTK temelli faktör analizi ve yetenek kestirimi yapma imkanı veren mirt (Chalmers,

2012) paketi kullanılarak yapılmıştır. Tüm analizler, 2PL modele göre gerçekleştirilmiştir. Analizlerin sonucunda elde edilen model-veri uyumlarına dair istatistikler aşağıda Tablo 3.1’de görülmektedir.

Tablo 3.1: Madde Setlerinin Farklı Faktör Analitik Yöntemlere Göre Kalibre Edilmesi Sonucu Elde Edilen Uyum İstatistikleri

Sınav Dönemi	FA Türü	-2LL	AIC	BIC	r	Δ -2LL	Δ sd
2014 Eylül	Tek Faktör	21294.68	21494.7	21897.5	-		
	İki Faktör	21290.94	21492.9	21899.8	.82	3.73	1
	3F Bifaktör	21464.16	22068.4	21907.3	-	130.52*	50
2015 Haziran	Tek Faktör	37591.34	37791.3	38245.3	-		
	İki Faktör	37588.7	37790.7	38249.2	.79	2.63	1
	3F Bifaktör	37442.48	37742.48	38423.42	-	148.86*	50
2015 Eylül	Tek Faktör	43822.1	44022.17	44490.39	-		
	İki Faktör	43729.32	43931.33	44404.22	.77	92.78*	1
	3F Bifaktör	43500.42	43800.42	44502.74	-	228.91*	49
2016 Haziran	Tek Faktör	61643.88	61843.88	62348.89	-		
	İki Faktör	61587.68	61789.67	62299.74	.85	56.2*	1
	3F Bifaktör	61482.12	61782.78	62540.3	-	105.55*	49

*p<.05

Yukarıdaki tabloda, farklı zamanlarda yapılmış dört sınav verisinin farklı faktör modellerine ait uyum değerleri yer almaktadır. Her bir veri setine tutumluluk prensibi gereği öncelikle tek boyutlu MTK modeli uygulanmıştır. Sonrasında ise sırasıyla iki boyutlu ve üç boyutlu-bifaktör model uygulanmıştır. Aynı veri seti üzerine uygulanan iki farklı modelin, model-veri uyumlarının karşılaştırılması amacıyla uyum istatistiklerinden yararlanılır. Uyum istatistiklerinin karşılaştırılması planlanan iki modelden ilkinin ikinci model içinde yuvalanmış (nested) olması durumunda, ilk model ve ikinci modelin -2LL değerleri ve serbestlik dereceleri arasındaki fark hesaplanır. Eğer, ilk modeli kapsayan ikinci modelin -2LL değerinde meydana getirdiği düşüş ilgili serbestlik derecesinde istatistiksel olarak anlamlıysa, ikinci modelin veri uyumunun ilk modelden daha iyi olduğu sonucuna varılır (Embretson

& Reise, 2000). Aynı veri setine uygulanan yukarıdaki modellerden tek boyutlu model iki boyutlu ve üç boyutlu bifaktör modelde, iki boyutlu model ise üç boyutlu bifaktör modelde yuvalanmıştır. Bu nedenle, her bir veri seti için model-veri uyumunun karşılaştırılmasında sırasıyla tek boyutlu model ile iki boyutlu model, iki boyutlu model ile de üç boyutlu bifaktör model karşılaştırılmıştır. İki boyutlu modelle tek boyutlu model arasında anlamlı farklılık elde edilmediği durumlarda ise, tek boyutlu modelle üç boyutlu bifaktör model karşılaştırılmıştır. Bu doğrultuda, ilk iki veri setinde iki boyutlu modelin -2LL değerlerinde meydana getirdiği düşüş istatistiksel olarak anlamlı olmadığından, söz konusu veri setlerinde tek boyutlu modelle üç boyutlu bifaktör model karşılaştırılmıştır. Son iki veri setinde ise, iki boyutlu modelin tek boyutlu modellere göre istatistiksel olarak daha uyumlu sonuçlar vermesinin ardından, iki boyutlu model üç boyutlu bifaktör modelle karşılaştırılmıştır. Elde edilen sonuçlara göre tüm veri setleri için en iyi model-veri uyumuna üç boyutlu bifaktör modelin sahip olduğu sonucuna ulaşılmıştır. Bu aşamadan sonra dört YTS'den elde edilen veriler, 2PL üç boyutlu bifaktör modele göre kalibre edilerek madde havuzu oluşturulmuştur.

3.3.1.2. Madde Havuzunda Yer Alacak Maddelerin Kalibrasyonu

Tüm veri setleri için en iyi uyum değerlerine üç boyutlu bifaktör modelin sahip olduğunun görülmesinin ardından, havuzu oluşturan maddelerin kalibrasyonu yapılmıştır. Öncelikle, MTK'nın temel sayıtlarından madde ve yetenek parametrelerinin değişmezliği varsayımı tüm madde setleri için ayrı ayrı test edilmiştir. Madde parametrelerinin değişmezliğinin test edilmesi amacıyla R'de "sample" komutu kullanılarak birey sayısı kadar, her bir bireye bir tanesi atanacak şekilde sıra gözetmeksizin rastgele sayı üretilmiştir. Bu sayıların tek-çift olmasına bağlı olarak bireyler rastgele (random) eşit iki gruba ayrılmışlardır. Sonraki aşamada üç boyutlu bifaktör modele göre bu iki birey grubundaki maddeler için madde parametreleri kestirilmiş ve rastgele ayrılmış bu iki gruptan aynı maddeler için elde edilen parametreler arasındaki Pearson momentler çarpımı korelasyon katsayısı hesaplanmıştır. Madde parametrelerinin kestirilmesi işlemi R'de, mirt (Chalmers, 2012) paketi kullanılarak yapılmıştır. Elde edilen sonuçlar Tablo 3.2'de verilmiştir.

Tablo 3.2: Madde Setlerinin Madde Parametrelerinin Değişmezliğine İlişkin Korelasyon Değerleri

	Parametreler Arası Korelasyon Değeri (r) ²			
	a1	a2	a3	b
2014 Eylül	.66*	.61*	.49*	.87*
2015 Haziran	.88*	.75*	.66*	.98*
2015 Eylül	.75*	.66*	.46*	.98*
2016 Haziran	.70*	.61*	.55*	.90*

*p<.05

Tablo 3.2 incelendiğinde, 2014 Eylül ayında gerçekleştirilen YTS maddeleri için rastgele belirlenmiş iki farklı gruptan elde edilen parametreler arasındaki korelasyon değerlerinin sırasıyla; a1 parametresi için .66, a2 için .61, a3 için .49 ve güçlük parametresi içinse .87 olduğu görülmektedir. 2015 yılı Haziran ayına ait YTS verileri üzerinde yapılan analizlere göre, farklı iki gruptan elde edilen değerler arasında a1 parametresi için .88, a2 için .75, a3 için .66 ve d parametresi içinse .98'lik korelasyon değerleri elde edilmiştir. 2015 Eylül sınavından elde edilen verilerin, rastgele iki gruba bölünmesinin ardından her bir madde için iki gruptan da elde edilen parametrelerin sırasıyla .75, .66, .46 ve .98 olduğu görülmüştür. Araştırma kapsamındaki son veri seti olan 2016 Haziran YTS üzerinde yapılan analizler sonucu, iki farklı gruptan aynı maddeler için elde edilen parametreler arasındaki korelasyonların sırasıyla .70, .61, .55 ve .90 olduğu görülmüştür. Tüm veri setleri için elde edilen değerler .05 düzeyinde istatistiksel olarak anlamlıdır. Tüm veri setlerinde güçlük parametresi için elde edilen korelasyonlar çok yüksekken, ayırt edicilik parametreleri için elde edilen korelasyonların değişkenlik gösterdiği görülmektedir. Tüm bu veriler ışığında, tüm veri setleri için madde parametrelerinin değişmezliği varsayımının sağlandığı sonucuna varılmıştır.

² Tüm analizler için a1 parametresi "genel yetenek" (general trait) boyutuna ait ayırt edicilik parametresidir. Bifaktör modellerde genel yetenek, tüm maddelerin altında yatan ortak varyans kaynağı olduğundan, a1 parametresi tüm maddeler için kestirilir. Diğer ayırt edicilik parametreleri ise grup/spesifik faktörlere aittirler ve bu nedenle sadece ait oldukları grup faktörleri için kestirilirler. Bu çalışmada, 50 maddelik soru setlerinden her birinin ilk 25 maddesi gramer, ikinci 25 maddesi ise kelime bilgisini ölçmeye yönelik maddelerden oluşmaktadır. Dolayısıyla a2 parametresi gramer, a3 parametresi ise kelime bilgisi grup faktörlerinin ayırt edicilik parametreleridir ve sadece bu faktörler için kestirilmişlerdir.

Her biri 50'şer maddeden oluşan farklı zamanlarda uygulanmış dört YTS'nin madde parametrelerinin değişmezliğine ilişkin analizlerin ardından, bir sonraki adımda her bir sınav verisinin MTK'nın bir diğer varsayımı olan yetenek parametrelerinin değişmezliği varsayımına uygunluğu test edilmiştir. Bu amaçla, her bir veri seti için aşağıdaki adımlar izlenmiştir:

- Gramer ve kelime bilgisi becerilerini ölçmek amacıyla yazılan 25'er madde, kendi içinde bir grupta 12, diğerinde 13 madde olacak şekilde, madde sıra sayılarının tek-çift olması durumuna göre iki gruba ayrılmıştır.
- Her bir bireyin, her bir boyut için bu iki farklı madde grubuna göre yetenekleri kestirilmiştir. Üç boyutlu bifaktör modele göre her bir birey için biri genel, biri gramer diğeri ise kelime bilgisi becerisine ait olmak üzere üç farklı yetenek düzeyi/theta (θ) kestirilmiştir. θ kestirim işlemi, mirt (Chalmers, 2012) paketi kullanarak R'de yapılmıştır.
- Her bir birey için iki farklı madde setinden elde edilen üç θ değeri arasındaki korelasyon hesaplanmıştır. Elde edilen sonuçlar, Tablo 3.3'de görülmektedir.

Tablo 3.3: Madde Setlerinin Yetenek Parametrelerinin Değişmezliğine İlişkin Korelasyon Değerleri

	Yetenek Kestirimleri Arası Korelasyon Değeri (r)		
	θ_1 (genel yetenek)	θ_2 (gramer)	θ_3 (kelime bilgisi)
2014 Eylül	.84*	.54*	.55*
2015 Haziran	.82*	.64*	.44*
2015 Eylül	.62*	.59*	.50*
2016 Haziran	.68*	.51*	.62*

*p<.05

Elde edilen sonuçlara göre dört madde setinin de değişmezlik varsayımlarını karşıladıkları sonucuna ulaşılmıştır. Madde setlerinin her birinin üç boyutlu bifaktör modele göre MTK varsayımlarını karşıladığının anlaşılmasının ardından, çok boyutlu BBT madde havuzunun oluşturulması amacıyla tüm madde setleri için madde parametreleri kestirilmiştir. Elde edilen sonuçlara göre parametre değerlerinin ± 4 aralığının dışında olan herhangi bir maddeye rastlanmamıştır. Bu durum sonrasında 200 maddenin tamamının havuzda yer almasına karar verilmiştir. Söz konusu maddelerin güçlük parametreleri -2.628 ile 3.041 arasında

değişmektedir. Tüm madde setleri için kestirilen madde parametreleri EK-3'te görülmektedir.

3.3.2. Simülasyon Süreci

Bir BBT sisteminin geliştirilmesinde, madde havuzunun oluşturulmasının ardından gerçek zamanlı BBT uygulamasında kullanılacak algoritmanın belirlenmesi gerekir. Gerçek zamanlı BBT'ler için olmazsa olmaz bu üç aşamanın dışında madde teşhir oranı (item exposure rate), içerik ağırlıklandırılması (content balancing), ön-BBT (pre-CAT) uygulamaları da algoritmaya dahil edilebilir (Chalmers, 2016). Algoritmanın gerçek zamanlı uygulama öncesinde belirlenmesi içinse simülasyon uygulaması yapılır. Gerçek-zamanlı uygulamada farklı başlatma, geçici yetenek kestirimi ve sonlandırma kurallarından hangisinin tercih edileceğine, maddelerin kağıt-kalem formatında uygulanması sonucu elde edilen verilerin kullanılmasına dayanan *post-hoc* simülasyon uygulamasıyla karar verilir (Weiss, 2004).

Post-hoc simülasyonlar esnasında; bireyin BBT havuzunu oluşturan maddelere kağıt-kalem formatındaki uygulamada verdiği yanıtlar, BBT uygulamasında ilgili maddeye verdiği yanıt olarak kabul edilir (Nydick & Weiss, 2009). Bu nedenle *post-hoc* simülasyonlara “gerçek veri” simülasyonları da denir (Thompson & Weiss, 2011). Ancak *post-hoc* simülasyonların bir BBT çıktısını doğru kestirebilmesi, tüm maddelerin tüm bireyler tarafından yanıtlanmasına bağlıdır (Weiss & Gibbons, 2007; Gibbons ve diğ., 2008). Bununla birlikte çoğu zaman bireylerin tüm maddeleri cevapladığı eksiksiz bir yanıt matrisi elde etmek mümkün olmaz. Böyle durumlarda, aralıklı/seyrekle (sparse) veri matrisinin tamamlanmasında *hibrit* (*hybrid*) simülasyonların kullanılması uygun olacaktır. *Hibrit* simülasyonlar, *monte carlo* ve *post-hoc* simülasyonları beraber kullanarak, “eğer bireylerin tamamı, soru havuzundaki tüm maddeleri yanıtlasaydı ne olurdu?” sorusuna yanıt ararlar. Bu şekilde; madde havuzundaki maddelerin bir kısmını yanıtlamamış farklı gruplardaki bireylerin, maddelerin tamamını yanıtlamalarına gerek kalmadan, bu soru setinin BBT olarak nasıl çalıştığını test etmiş olacaklardır. *Hibrit* simülasyon çalışmaları, BBT çalışmasında kullanılan maddelerin önceden uygulanmış olma durumuna göre iki şekilde yürütülürler (Nydick & Weiss, 2009).

Bu çalışmada kullanılan madde havuzunda yer alan maddeler, madde setlerinin tamamı önceden uygulandığı için Nydick & Weiss (2009) tarafından önerilen bu iki süreçten ikincisi uygulanmış olup, bu uygulama basamakları aşağıda özetlenmiştir:

1. Araştırma kapsamındaki dört madde seti, üç boyutlu bifaktör modele göre kalibre edilerek madde havuzu oluşturulmuştur.
2. Yanıtladıkları madde setindeki maddelere verdikleri yanıtlara dayanarak, her bir birey için üç boyutlu bifaktör modele göre yetenek kestirimi yapılmıştır. Bu doğrultuda her bir birey için biri genel, biri gramer ve diğeri kelime bilgisi becerisine ait olmak üzere üç farklı θ kestirilmiştir.
3. Bu durumda, araştırma kapsamında verileri kullanılan dört farklı YTS'ye giren bireyler alt alta sıralandığında, her biri yalnız bir madde setini yanıtladığı ve bu nedenle yanıt matrisinde diğer üç madde setine verdikleri yanıtlar boş olduğundan, *post-hoc* uygulamasına geçilebilmesi için her bireyin kalan üç madde setine verdiği yanıtlar üretilmiştir. Aşağıdaki tabloda, madde havuzunu oluşturan maddelere ait birey-yanıt matrisi yer almaktadır.

Tablo 3.4: Eksik Birey-Yanıt Matrisinin Gösterimi

	50 Soru 2014 Eylül	50 Soru 2015 Haziran	50 Soru 2015 Eylül	50 Soru 2016 Haziran
2014 Eylül 415 Birey	X			
2015 Haziran 692 Birey		X		
2015 Eylül 798 Birey			X	
2016 Haziran 1153 Birey				X

Yukarıdaki tabloda toplam 16 hücre yer almaktadır. “X” simgesi bulunan hücreler, satırdaki bireylerin sütundaki madde setini yanıtlamış olduğu anlamına gelir. Diğer 12 hücre içinse, yanıt üretme işlemi gerçekleştirilmiştir.

4. Yanıt üretme işlemi, her bir madde seti için ayrı ayrı yapılmıştır. Bu amaçla öncelikle her madde setinin parametreleri girilmiş ve bununla birlikte bireylerin yanıtladıkları madde setinden elde edilen yetenek kestirimleri göz önüne alınarak, yanıtlamadıkları madde setleri için yanıt üretilmiştir. Her bir 50 soruluk madde seti için yanıt üretme işlemi, soldan sağa doğru

gerçekleştirilmiştir. Örneğin, 2014 Eylül ayında YTS'ye giren bireylerin öncelikle yetenek kestirimi yapılmış, sonra sırasıyla 2015 Haziran, 2015 Eylül ve 2016 Haziran sınavlarındaki maddeler için, bu maddelerin parametrelerine dayanılarak yanıtları üretilmiştir. Aynı süreç, diğer tüm eksik yanıt matrisi için tekrarlanmış, sonuçta toplam 12 yanıt üretme işlemi gerçekleştirilmiştir. Yanıt üretme işleminin tamamı, R'de tanımlı mirtCAT (Chalmers, 2016) paketindeki "generate.mirt_object" fonksiyonu kullanılarak yapılmıştır. Her bir madde seti için eksiksiz yanıt matrisi elde edildikten sonra, tüm setler bir araya getirilmiştir.

5. Bu şekilde oluşturulan eksiksiz yanıt matrisi, 3058 bireyin 200 maddeye verdikleri yanıtlardan meydana gelmektedir.
6. Elde edilen eksiksiz yanıt matrisi kullanılarak *post-hoc* simülasyonlar gerçekleştirilmiştir.

Araştırmanın bir sonraki aşamasında farklı yetenek kestirim, madde seçim yöntemleri ve sonlandırma kuralları kombine edilerek toplam 36 farklı durum için *post-hoc* simülasyon çalışılması yürütülmüştür. Her bir simülasyonun sonucunda araştırmanın bağımlı değişkenleri olan; gerçek ve kestirilen θ değerleri arasındaki korelasyon ile RMSD, yanlışlık (bias), standart hata ve yanıtlanan ortalama madde sayısı hesaplanmıştır.

Testle ölçülmesi amaçlanan her bir boyut için, bireylere ait gerçek θ değerleri BBT uygulaması sonrası elde edilen θ değerleri arasındaki tutarlılığa dair elde edilen istatistiklerin ilki RMSD (Root Mean Squared Difference-ortalama kareköklerin sapması)'dir. Bu değer, gerçek ve kestirilen θ farklarının karelerinin ortalamasının kare köküdür. Yanlışlık (bias) değeri ise, gerçek ve kestirilen θ değerleri arasındaki farkın ortalamasıdır. Standart hata (ölçmenin standart hatası) değeri, her bir boyuta ilişkin güvenirliliğin karesinin 1'den çıkarılmasıyla elde edilen değer karekökünün, her bir boyuta ilişkin kestirilen yetenek parametresinin standart sapmasıyla çarpılmasından elde edilir. RMSD ve standart hatanın formülleri aşağıda görülmektedir.

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

$$\text{ÖSH}(\theta) = \sqrt{\frac{1}{I(\theta)}} = \sqrt{\frac{1}{E_{\theta} \left[\frac{d \log f(Y; \theta)}{d\theta} \right]}}$$

Yanlılığın (bias) mutlak değerinin, RMSD'nin ve standart hatanın sıfıra yakın olması; yanıtlanan madde sayısının ise azalması BBT uygulamalarında ölçme kesinliğini artırır (Segall, 2005; Weiss, 2011). Bunların yanında gerçek ve kestirilen θ değerleri arasındaki korelasyonun yüksek olması beklenir. Her bir koşul için yapılan simülasyon uygulaması sonucunda genel yetenek, gramer ve kelime bilgisi θ kestirimleri için sonlandırma kuralının niteliğine göre standart hata – θ veya yanıtlanan madde sayısı – θ grafikleri oluşturulmuştur. Söz konusu grafikler, R'de tanımlı lattice (Sarkar, 2011) paketi kullanılarak elde edilmiştir. Bu grafikler, EK-4'te yer almaktadır. Bu çalışmanın simülasyon aşamasında kullanılan BBT bileşenleri aşağıda açıklanmaktadır.

1. **Başlatma Kuralı:** Tüm koşullar, mirtCAT paketinin sınırlılığı nedeniyle ortalama yetenek düzeyine ($\theta = 0$) yakın ve ayırt ediciliği yüksek bir maddeyle başlamıştır. Bu niteliklere en uygun maddenin, madde havuzundaki 117. madde olduğuna karar verilmiştir. Söz konusunun maddenin güçlük parametresi -.013, ayırt edicilik parametreleri ise sırasıyla 1.328 ve .232'dir.
2. **Yetenek Kestirimi (Geçici):** Bu çalışmada Bayes yetenek kestirim yöntemlerinden EAP ve MAP kullanılmıştır. ML yetenek kestirimi araştırma kapsamında kullanılan R'de tanımlı mirtCAT paketinde bifaktör model için kullanılabilir değildir.
3. **Madde Seçim Kuralı:** Çalışma kapsamında, çok boyutlu BBT uygulamalarında kullanılabilecek farklı madde seçim yöntemlerinden D-rule, W-rule, T-rule ve KL (Kullback Leibner) kullanılmıştır. Bunlar arasından W-rule ve T-rule, çok boyutlu uygulamalarda herhangi bir boyutun “daha önemli” görülmesi durumunda, madde seçiminde bu boyutlara istenilen düzeyde ağırlıklandırma yapma özelliğine sahiptir (Chalmers, 2016). Bu nedenle bu iki madde seçim yöntemi; boyutlar için özel bir ağırlıklandırma yapılmaması ve genel yetenekten başlayarak (.8, .1, .1) şeklinde ağırlıklandırma yapılması durumu olmak üzere, tüm sonlandırma kuralları ve yetenek kestirim

yöntemleri için ikişer kez kullanılmıştır. Bu doğrultuda toplam altı madde seçim yöntemi kullanılarak simülasyon gerçekleştirilmiştir.

4. Sonlandırma Kuralı: Bu çalışmada, BBT literatüründeki sonlandırma kurallarından bireyin; önceden belirlenen sabit sayıda madde yanıtlamasıyla, yetenek kestiriminin önceden belirlenen bir standart hata değerinin altına düşmesiyle ve bir önceki yetenek kestirimiyle son yetenek kestirimi arasında önceden belirlenmiş bir farkın altına düşmesiyle sonlandırma olmak üzere üç farklı yöntem kullanılmıştır. Sabit madde sayısı olarak 20 madde, standart hata değeri olarak .4 ve yetenek kestirimleri arasındaki fark olarak .05 kullanılmıştır. Standart hataya dayalı sonlandırma yöntemi, sadece genel yetenek için uygulanmıştır. Alanyazın incelendiğinde, sonlandırma kuralı olarak belirlendiği durumlarda, .20'den .50'ye kadar farklı standart hata değerlerinin kullanıldığı görülmektedir. Bu araştırmanın gerçekleştirilmesi öncesinde yapılan deneme uygulamalarında, standart hata değerinin .30 ve .35 olarak kabul edildiği durumlarda bir çok birey için testin çok fazla madde ile ya da belirlenen maksimum madde sayısına ulaşınca sonlandırılabilirdiği görülmüştür. Bu durumdan yola çıkarak, madde havuzunun özellikleri doğrultusunda standart hataya dayalı sonlandırma kuralı için söz konusu standart hata değerinin .40 olarak belirlenmesine karar verilmiştir.

5. Yanıtlanacak Minimum ve Maksimum Madde Sayıları: Madde havuzunun çok geniş olmadığı durumlarda, özellikle yetenek skalasının iki ucundaki bireylerin yetenek düzeyine yönelik yeterli sayıda madde olmayabilir. Bu durumda, söz konusu bireyler diğerlerinden çok daha fazla madde yanıtlamak durumunda kalacaklar, kimi zamanda madde havuzundaki maddelerin tamamını yanıtlamaları bile standart hataya dayalı durdurma kuralı için yeterli olmayacaktır. Böyle durumlarda sonlandırma kuralının yanında yanıtlanacak en az ve en fazla madde sayısının da eklenmesi yararlı olacaktır (Frey, 2009). Araştırma kapsamında kullanılan madde havuzu çok büyük olmadığından, kullanılan tüm algoritmalarda yanıtlanacak madde sayısı en az 10, en çok 60 olarak belirlenmiştir. Burada belirlenen yanıtlanacak en düşük madde sayısı belirlenirken, İngilizce hazırlık sınıflarında derse giren öğretim elemanlarının görüşleri dikkate alınmıştır.

Simülasyon çalışması kapsamında elde edilen 36 koşulu oluşturan bileşenler, aşağıdaki tabloda özetlenmiştir.

Tablo 3.5: Simülasyon Çalışmasındaki 36 Koşulu Oluşturan BBT Bileşenleri

BBT Bileşenleri	Kullanılan Yöntem	Koşul sayısı
Yetenek Kestirimi	EAP ve MAP	2
Madde Seçimi	D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule	6
Sonlandırma Kuralı	Standart hata (.40), θ değişimi ($\Delta\theta < .05$) ve sabit madde sayısı ($k=20$)	3

Tablo 3.5'te görüldüğü üzere; yetenek kestirimi için iki, madde seçimi için altı ve testi sonlandırma için üç yöntem kullanılarak toplam 36 koşul elde edilmiştir.

3.3.3. Gerçek-Zamanlı Çok Boyutlu BBT (real-time MCAT) Uygulama Süreci

Bu bölümde, gerçek-zamanlı çok boyutlu BBT uygulamasının gerçekleştirilme süreci üzerinde durulmuştur. Öncelikli olarak R'de tanımlı mirtCAT (Chalmers, 2016) paketi özelleştirilerek, çok boyutlu BBT uygulaması çevrimiçi kullanılabilir hale getirilmiştir. Sonraki aşamada ise, Anadolu Üniversitesi YDYO öğrencilerinden A kurunda bulunan 99 kişiye gerçek-zamanlı uygulama yapılmıştır.

3.3.3.1. mirtCAT Paketinin Özelleştirilerek Yazılımın Oluşturulması

Simülasyon uygulamasının ardından, elde edilen sonuçlar yorumlanmış ve gerçek-zamanlı çok boyutlu BBT uygulaması için en uygun koşul belirlenmiştir. Belirlenen bu koşul kullanarak gerçek-zamanlı çok boyutlu BBT uygulamasına geçilmiştir. Bu amaçla, simülasyon çalışması için de kullanılan mirtCAT paketi kullanılmıştır. Üç boyutlu bifaktör modele uygun kodların oluşturulmasının ardından uygulama, R'de tanımlı shiny (2014) web arayüzü kullanılarak bireysel bilgisayarda çalıştırılmıştır. Lokalde birkaç pilot uygulamanın yapılması ve uygulamanın hatasız çalıştığının

görülmesinin ardından, öncelikle <https://www.shinyapps.io/> üzerinden gerçek-zamanlı uygulama yürütülmeye çalışılmıştır. Ancak, gerçek-zamanlı bir uygulamada aynı anda 1’den fazla uygulamanın yapılması için shiny tarafından sağlanan sunucu hizmet bedelinin çok yüksek olması nedeniyle araştırmacı tarafından bir sunucu kiralanmış, yazılan kodlar bu doğrultuda özelleştirilerek her bir uygulama sonucunun bir veri tabanına kaydedilmesi sağlanmıştır. Gerçek-zamanlı çok boyutlu BBT uygulamasına erişim <http://adaptiftest.net/mcat/> adresinden sağlanmıştır. Bireyler adrese ulaştıklarında karşılarına çıkan ekran aşağıda görülmektedir.

Şekil 3.2: Gerçek-Zamanlı Çok boyutlu BBT Uygulamasının Karşılama Ekranı

Testi alan bireylere verilen kısa yönergenin ardından, “next” tuşuna basılmasıyla bireyler, demografik bilgilerini girdikleri aşağıda görülen ekranla karşılaşmaktadır.

Şekil 3.3: Bireylerin Demografik Bilgilerini Girdikleri Ekran Görüntüsü

Bu aşamadan sonra “next” tuşuna iki kez basılmasıyla bireyler yanıtlayacakları ilk maddeyle karşılaşmaktadır. İlgili ekran görüntüsü aşağıda yer almaktadır.

Bireyselleştirilmiş İngilizce Sınavı

Geliştiren
Murat Doğan ŞAHİN

To progress through the interface, click on the action button below.

Next

My roommate _____ worried because he _____ the bills yet.

- ☐ is looking / did not pay
- ☐ looked / has not paid
- ☐ was looking / did not pay
- ☐ looks / has not paid

Şekil 3.4: Maddenin Yanıtlanmadan Önceki Ekran Görüntüsü

Yukarıdaki görüntüde next butonunun aktif olmadığı görülmektedir. Bunun sebebi; lokal uygulamada karşılaşılmamış olmasına rağmen çevrimiçi uygulamada, madde yanıtlanmadan next tuşuna basılmasının uygulamanın hata vermesine neden olmasıdır. Bu hata sonucunda uygulama kaldığı yerden devam etmediği gibi, sunucu yenilenmeden yeni bir sınav uygulamasının da yapılmasına imkan vermediğinden, bir “javascript” yazılarak söz konusu tuşun madde yanıtlanınca aktif olması sağlanmıştır. Maddenin yanıtlandıktan sonraki ekran görüntüsü aşağıda görülmektedir.

Bireyselleştirilmiş İngilizce Sınavı

Geliştiren
Murat Doğan ŞAHİN

To progress through the interface, click on the action button below.

Next

My roommate _____ worried because he _____ the bills yet.

- ☐ is looking / did not pay
- ☐ looked / has not paid
- ☐ was looking / did not pay
- ☒ looks / has not paid

Şekil 3.5: Maddeye Yanıt Verildikten Sonraki Ekran Görüntüsü

Maddenin yanıtlanmasının ardından bireyin geçici yetenek kestiriminin yapılarak bir sonraki soruya geçilmesini sağlayan next tuşu aktif hale gelmektedir. Yapılan geçici yetenek kestirimi ile birlikte standart hata kestirimine göre sonlandırma kuralının gerçekleşmesi durumunda bireye bildirilen sonuç ekranı aşağıda görülmektedir.

Bireyselleştirilmiş İngilizce Sınavı

Geliştiren
Murat Doğan ŞAHİN

To progress through the interface, click on the action button below.

Next

Testimizi tamamladığınız için teşekkür ederiz.

Puanınız -3 ile 3 arasında hesaplanmaktadır, 0 ortalama puan demektir.

Sizin Puanınız: -0.182, Puanınızın Standart Hatası = 0.383

Lütfen Next'e tıklayarak sınavınızı bitiriniz.

Şekil 3.6: Testin Sonlanmasının Ardından Görülen Sonuç Ekranı

Sınavın sonlandırılmasıyla bireye yukarıda görülen, yalnızca genel yetenek (general trait) düzeyi için yapılan kestirim ve bu kestirimin standart hatasına ilişkin bilgi verilmiştir. Next tuşuna basılmasıyla birlikte ise, ekran inaktif hale gelmekte ve bireyin sınava ilişkin bilgileri sunucudaki veri tabanına kaydedilmektedir. Yukarıdaki sınav sonucuna ilişkin veri tabanına yansıyan ayrıntılı sonuçlar aşağıda görülmektedir.

```
$final_estimates
      Theta_1      Theta_2      Theta_3
Estimates -0.1816384 -0.04847894 0.1140159
SEs        0.3834855  0.73712675 0.5953089

$raw_responses
[1] "was looking / did not pay"
[2] "make a fortune"
[3] "how "
[4] "incurable"
[5] "such/that"
[6] "The judge told them they would not have to go to jail"
[7] "when"
[8] "enemies"
[9] "what"
[10] "suffering"
[11] "violent"

$scored_responses
[1] 0 1 0 0 1 1 0 0 1 1 1

$items_answered
[1] 117 132 106 82 4 11 112 149 12 189 140
```

```

$thetas_history
      Theta_1      Theta_2      Theta_3
[1,]  0.0000000  0.0000000  0.000000e+00
[2,] -0.4583913 -0.08008080 -4.957016e-07
[3,] -0.2144413 -0.09757380  3.184561e-01
[4,] -0.4454139  0.15743365  3.795707e-01
[5,] -0.5389794  0.14241232  5.250023e-01
[6,] -0.3648825  0.37752266  5.221224e-01
[7,] -0.2034398  0.15118375  5.312646e-01
[8,] -0.3476555  0.03180295  5.225508e-01
[9,] -0.4481992  0.03116828  3.093520e-01
[10,] -0.4033644 -0.06954755  3.036985e-01
[11,] -0.2531225 -0.05374009  3.488615e-02
[12,] -0.1816384 -0.04847894  1.140159e-01

$thetas_SE_history
      Theta_1      Theta_2      Theta_3
[1,]  1.0000000  1.0000000  1.0000000
[2,]  0.8470180  0.9956790  1.0000000
[3,]  0.7674453  0.9951322  0.8963076
[4,]  0.6920992  0.9259844  0.8840643
[5,]  0.6463972  0.9286774  0.7847002
[6,]  0.5856929  0.8628013  0.7713149
[7,]  0.5342033  0.7982516  0.7607976
[8,]  0.4710904  0.7646190  0.7700495
[9,]  0.4519367  0.7652708  0.6918357
[10,] 0.4380200  0.7307912  0.6861287
[11,] 0.4068815  0.7345326  0.6112182
[12,] 0.3834855  0.7371268  0.5953089

$demographics
      Adınız      Bölüm
1 Celal SALİK Basın yayın

$item_time
[1] 4.538 2.843 2.331 2.046 2.492 2.646 3.085 2.809 4.128 3.401 3.596

```

Şekil 3.7: Bireyin Veri Tabanına Yansıyan Ayrıntılı Sınav Sonucu

Yukarıda görülen sonuçlar, mirtCAT paketi tarafından verilebilen tüm analizleri içermektedir. Bunların ayrı bir dosyaya yazılması amacıyla, gerçek-zamanlı çok boyutlu BBT uygulaması için yazılan R kodları özelleştirilmiştir. Her bir bireyin sınav sonucu, “final estimates” başlığı altında öncelikle tüm boyutlardaki yetenek kestirimleri ve bu kestirimlerin standart hatalarını içermektedir. Bir sonraki adımda, örnekte görülen bireyin yanıtladığı 11 maddede işaretlediği seçenekler görülmektedir. Bunun hemen altında “scored responses” başlığı altında ise, verilen yanıtların doğru-yanlış olma durumunu özetleyen “1-0” şeklinde kodlanmış bir veri elde edilir. Bu bilginin ardından, “items answered” başlığında, yanıtlanan maddelerin soru dosyasındaki numaraları verilmektedir. “Thetas history” kısmında, bireyin her bir maddeye verdiği yanıt ardından üç boyut için de yapılan kestirimler; hemen altındaki “thetas SE History” başlığında ise, söz konusu yetenek kestirimlerine ait standart hata değerleri yer alır. Bunların altında bireylerin demografik bilgileri yer alırken, son olarak “item time” başlığı altında bireyin her bir maddeyi yanıtlarken harcadığı süre, birimi saniye olacak biçimde verilmiştir.

3.3.3.2. Gerçek-Zamanlı Uygulamanın Yapılması

Gerçek-zamanlı çok boyutlu BBT uygulaması, Anadolu Üniversitesi YDYO'da A kurunda İngilizce hazırlık eğitimi gören 99 öğrenciye, internet erişim hızının yüksek olduğu YDYO'ya bağlı bilgisayar laboratuvarlarında yapılmıştır. Uygulama, A kurunda yer alan iki farklı gruba uygulanmıştır. İlk grup; A kurunda yer alma hakkını ilk dönem elde eden, dolayısıyla ara dönemde (Şubat ayı başında) hazırlık sınıfını geçme amacıyla yapılan sınava girme hakkı olan öğrencilerden oluşmaktadır. Buradaki amaç, bireylerin kağıt-kalem testi sonuçlarıyla gerçek-zamanlı çok boyutlu BBT uygulaması sonuçlarını karşılaştırabilmektir. Söz konusu öğrenci grubu, üç ayrı sınıfta öğrenim görmektedir. İkinci grup ise, A kuruna ikinci dönemde Mart ayı itibariyle başlayan öğrencilerin yer aldığı üç ayrı sınıfta öğrenim gören öğrencilerden oluşmaktadır. Dolayısıyla bu öğrenci grubu, dönem arasında hazırlık geçme amacıyla uygulanan yeterlilik sınavına girme hakkına sahip olmamışlardır. Bu nedenle bu öğrenci grubuna yapılan uygulama, herhangi bir kağıt-kalem uygulaması ile karşılaştırılmamıştır. İlk öğrenci grubunda 46, ikinci grupta ise toplam 53 öğrenci sınavı almıştır. Öğrencilerin 47'si kız, 52'si ise erkektir. Her bir sınıfta uygulama kendi laboratuvar ders saatlerinde yapılmıştır. Öğrencilere çevrimiçi uygulamanın yapıldığı bilgisayarların tamamı aynı teknik özellikleri taşımakta olup, bağlantı ve kullanım konusunda uygulama esnasında problem yaratacak bir duruma mahal vermemek için gerekli kontroller yapılmıştır. Uygulama öncesinde bireylere sınavın yapılma amacı ayrıntılı olarak anlatılarak ders sorumlularının da desteğiyle ciddi bir sınav ortamı yaratılarak, tüm öğrencilerin sınavı aynı koşullarda almaları sağlanmıştır. Her bir öğrenci için uygulama yaklaşık olarak 9 – 13 dakika arası sürmüştür. Yalnızca 45 madde yanıtlayan bir öğrencinin uygulama süresi 18 dakikaya ulaşmıştır.

4. BULGULAR

Araştırmanın bu bölümünde, alt problemlerin çözümlenmesine yer verilmiştir. Bu bağlamda ilk olarak; farklı yetenek kestirim yöntemleri, madde seçim yöntemleri ve sonlandırma kuralları kullanılarak elde edilen 36 farklı koşul için gerçekleştirilen *hibrit* simülasyon sonuçları yer almaktadır. Her bir koşul, sonlandırma kuralına göre sınıflandırılarak elde edilen değerler karşılaştırılmıştır. Bir sonraki aşamada, öncelikle araştırma kapsamında kullanılan üç sonlandırma kuralının her biri için en iyi koşul belirlenmiş, sonrasında ise bu üç koşul karşılaştırılarak gerçek-zamanlı uygulamada hangisinin kullanılacağına karar verilmiştir. Son aşamada ise, simülasyon sonuçlarına dayalı olarak belirlenen algoritmaya göre gerçek-zamanlı çok boyutlu BBT uygulaması yapılarak sonuçları raporlanmıştır.

4.1. Alt Problem 1: Standart hataya dayalı sonlandırma kuralının kullanıldığı çok boyutlu BBT uygulamalarında farklı yetenek kestirim yöntemleri (EAP ve MAP) ve madde seçim kurallarına (D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule) göre gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanıtlanan ortalama madde sayısı, yanlılık (bias) ve RMSD değerleri nasıldır?

Araştırmanın bu bölümünde, genel yetenek için kestirilen standart hatanın .4'ün altına düşmesiyle testin sonlandırılmasına dayanan çok boyutlu BBT uygulaması çerçevesinde farklı yetenek kestirim ve madde seçim yöntemlerinin kullanılmasıyla elde edilen bulgular yer almaktadır. İlk kısımda EAP, ikinci kısımda ise MAP yetenek kestirim yönteminin kullanıldığı koşullara ait sonuçlar incelenmiştir.

Sabit hataya dayalı sonlandırma kuralı çerçevesinde yetenek kestirimi olarak EAP yöntemi kullanıldığında D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule madde seçim yöntemlerinin kullanılmasıyla elde edilen değerlere Tablo 4.1'de yer verilmiştir.

Tablo 4.1: Standart Hataya Dayalı Durdurma Kuralının Kullanıldığı Koşullarda Elde Edilen Korelasyon, Yanlılık, RMSD, Standart Hata ve Yanıtlanan Madde Sayısı İstatistikleri

Sonlandırma Kuralı	Yetenek Kestirimi	Madde Seçim Yöntemi	r			Yanlılık (bias)			RMSD			Std.Hata			Ortalama Madde Sayısı
			$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	k
Standart Hataya Dayalı Sonlandırma (SH < .4)	EAP	D-rule	.91	.63	.66	-.0441	.0472	.0454	.38	.55	.54	.37	.69	.66	16.1
		KL	.92	.41	.58	-.0029	-.0037	-.0288	.35	.61	.55	.38	.87	.77	17.7
		W-rule	.93	.41	.58	-.0013	-.0030	.0293	.35	.61	.55	.37	.87	.78	17.8
		W-rule (.8,.1,.1)	.93	.33	.47	-.0074	.0115	.0200	.34	.62	.59	.35	.91	.83	12.9
		T-rule	.93	.44	.57	-.0273	.0249	.0476	.35	.60	.58	.35	.86	.71	13.0
		T-rule (.8,.1,.1)	.93	.35	.51	-.0081	.0187	.0328	.35	.62	.60	.34	.90	.78	12.7
	MAP	D-rule	.90	.60	.62	-.0326	.0370	.0721	.39	.56	.54	.39	.71	.65	13.4
		KL	.92	.41	.56	.0012	-.0007	.0019	.36	.61	.55	.39	.87	.78	15.2
		W-rule	.92	.40	.57	.0026	.0018	.0027	.36	.61	.55	.39	.87	.78	15.2
		W-rule (.8,.1,.1)	.92	.30	.46	.0079	.0064	.0329	.35	.62	.58	.36	.92	.84	11.4
		T-rule	.92	.42	.55	-.0195	.0084	.0806	.35	.60	.57	.36	.86	.69	11.5
		T-rule (.8,.1,.1)	.92	.33	.49	.0037	.0050	.0482	.35	.62	.58	.35	.90	.77	11.33

Tabloda 4.1’de EAP yetenek kestirim yönteminin kullanıldığı durumlarda genel yetenek düzeyi (θ_9) için elde edilen sonuçlar incelendiğinde, bireylerin genel yeteneklerine ait gerçek θ değerleri ile kestirilen θ değerleri arasındaki korelasyonların tüm koşullarda birbirine çok yakın olup .91 ile .93 arasında değiştiği görülmektedir. Kestirilen ortalama yanlılık değerleri arasında -.0013 ile 0’a en yakın değer W-rule ile elde edildiği görülmektedir. .34 ile .38 arasında değişen RMSD değerlerinin en küçük olanı ise ağırlıklandırılmış W-rule madde seçim yöntemiyle elde edilmiştir. Sonlandırma kuralı sabit .4 standart hataya göre yapıldığından, tüm madde seçim yöntemleriyle elde edilen standart hata ortalamalarının birbirine yakın ve söz konusu değer altında elde edilmiş olması beklenen bir durumdur. Son olarak testin sonlandırılmasına kadar yanıtlanan madde sayılarının ortalamaları incelenmiş olup, bu değerlerin 12.7 ile 17.8 arasında değiştiği ve bu sayının ağırlıklandırılmış T-rule madde seçim yönteminin kullanıldığı koşulda en az olduğu sonucuna ulaşılmıştır.

Gramer (θ_2) için elde edilen sonuçlara göre, gerçek ve kestirilen θ değerleri arasındaki korelasyonların .33 ile .63 arasında değiştiği görülmektedir. Bunlar arasında en yüksek değer D-rule madde seçim yöntemiyle elde edilmiş olup, diğer yöntemlerle elde edilen korelasyon değerleri birbirine yakın ve D-rule ile elde edilen değere göre oldukça düşüktür. Yanlılık değerleri arasından 0’a en yakın olanın (-.0030) ile W-rule yöntemine ait olduğu sonucuna ulaşılmıştır. Tabloya göre RMSD değerleri .55 – .62 arasında olup, en düşük değer D-rule madde seçim yönteminin kullanıldığı koşulda elde edildiği bulgulanmıştır. Testin sonlandırılmasıyla elde edilen standart hataların ortalamasına bakıldığında ise, .69 ile .91 arasında değişen değerler arasında en düşük olanın D-rule yöntemiyle elde edildiği, korelasyon değerindeki duruma benzer olarak diğer madde seçim yöntemleriyle elde edilen hata değerlerinin birbirine oldukça yakın ve D-rule ile elde edilenlere göre çok büyük oldukları sonucuna ulaşılmıştır.

Kelime bilgisi (θ_3) için elde edilen değerlere bakıldığında, gerçek ve kestirilen θ değerleri arasındaki korelasyonun .47 ile .66 arasında değiştiği, en yüksek korelasyonun D-rule madde seçim yönteminin tercih edildiği koşulda elde edildiği görülmektedir. Yanlılık değerleri arasında .0200 ile 0’a en yakın olanın ağırlıklandırılmış W-rule madde seçim yönteminin kullanıldığı koşulda ortaya çıktığı bulgusuna ulaşılmıştır. RMSD değerleri .54 ile .60 arasında değişmekte olup en

düşük değerin D-rule ile elde edildiği görülmüştür. Son olarak ortalama standart hata değerleri incelenmiş, .66 ile .83 arasında değişen bu değerlerden en düşüğünün de yine D-rule madde seçim yöntemiyle elde edildiği sonucuna ulaşılmıştır.

Tablo 4.1’de MAP yetenek kestirim yönteminin kullanıldığı koşullara ait bulgular incelendiğinde; genel yetenek düzeyi (θ_g) için gerçek ve kestirilen θ değerleri arasındaki korelasyon katsayılarının D-rule ile .90, diğer tüm yöntemlerle ise .92 olarak kestirildiği görülmektedir. .0012 ile en düşük yanlışlık değerine, KL madde seçim yönteminin kullanıldığı koşulda ulaşılmıştır. .35 ile .39 arasında değişen RMSD ve standart hata değerleri arasında en düşüğü iki istatistik için de ağırlıklandırılmış T-rule madde seçim yönteminin kullanıldığı koşullarda elde edilmiştir. Son olarak yanıtlanan madde sayılarının ortalamasının 11.3 ile 15.2 arasında birbirine oldukça yakın değerler olduğu sonucuna varılmıştır.

Gramer (θ_2) için elde edilen sonuçlara göre, yetenek kestirimleri arasındaki korelasyon değerleri .30 ile .60 arasında değişmektedir. EAP ile elde edilen sonuçlara benzer olarak burada da en yüksek değer D-rule ile elde edilirken diğer madde seçim yöntemlerinin kullanıldığı koşullardaki korelasyon değerleri birbirine yakın ve düşüktür. Yanlışlık değerleri arasında 0’a en yakın olan -.0007 değeri, KL madde seçim yöntemiyle elde edilmiştir. RMSD ve standart hata değerleri ise sırasıyla .56 ile .62 ve .71 ile .92 arasında değişmekte olup, iki istatistik için de en düşük değerlerin D-rule madde seçim yönteminin kullanıldığı koşulda elde edildiği görülmektedir.

Kelime bilgisi (θ_3) için elde edilen değerler incelendiğinde, gerçek ve kestirilen θ değerleri arasındaki korelasyonların .46 ile .62 arasında değiştiği, en yüksek değerin D-rule madde seçim yöntemiyle elde edildiği bulgusuna ulaşılmıştır. Yanlışlık değerleri içerisinde ise 0’a en yakın değerin KL madde seçim yöntemiyle elde edilen .0019’luk değer olduğu görülmektedir. RMSD değerleri ise .54 ile .58 arasında değişmekte olup en düşük değer D-rule ile elde edilmiştir. En düşük değerin yine D-rule ile elde edildiği standart hata değerlerinin ise .65 - .84 aralığında olduğu bulgusuna ulaşılmıştır.

4.2. Alt Problem 2: θ deęişimine dayalı sonlandırma kuralının kullanıldığı çokboyutlu BBT uygulamalarında farklı yetenek kestirim yöntemleri (EAP ve MAP) ve madde seçim kurallarına (D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule) göre gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanıtlanan ortalama madde sayısı, yanlılık (bias) ve RMSD deęerleri nasıldır?

Araştırmanın bu bölümünde, bireyin yanıtladığı son madde sonrası elde edilen yetenek kestirimiyle bir önceki yetenek kestirimi arasındaki farkın her bir boyut için .05'in altına düşmesiyle testin sonlandırılmasına dayanan çok boyutlu BBT uygulaması çerçevesinde farklı yetenek kestirim ve madde seçim yöntemlerinin kullanılmasıyla elde edilen bulgular yer almaktadır. İlk olarak EAP, ikinci olarak ise MAP yetenek kestirim yönteminin kullanıldığı koşullara ait sonuçlar raporlanmıştır.

θ deęişimine dayalı sonlandırma kuralı çerçevesinde EAP yetenek kestirim ve farklı madde seçim yöntemlerinin kullanıldığı BBT koşullarına göre Tablo 4.2'de yer alan deęerler incelendiğinde, genel yetenek düzeyi (θ_g) için gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonların birbirine çok yakın sonuçlar verdiği görülmektedir. KL ve W-rule madde seçim yöntemleriyle elde edilen deęerler .92 iken, dięer tüm yöntemlerle elde edilen deęerler .94 olarak bulunmuştur. Yanlılık deęerleri arasında en küçük deęer, -.0031 ile ağırlıklandırılmış W-rule madde seçim yönteminin kullanıldığı koşulda elde edilmiştir. RMSD deęerleri incelendiğinde ise korelasyon deęerlerine benzer bir sonuçla karşılaşılmış, birbirine oldukça yakın olan bu deęerlerden KL ve W-rule madde seçim yöntemiyle elde edilenler .35 olarak bulunurken, dięer madde seçim yöntemleriyle bu deęer .32 olarak hesaplanmıştır. Standart hata ortalamaları, .30 ile .37 arasında deęişmektedir. En düşük deęerlerin, D-rule ve T-rule yöntemleriyle elde edildiği bulgulanmıştır. Son olarak yanıtlanan madde sayılarının ortalamaları incelendiğinde, en düşük deęerler ağırlıklandırılmış madde seçim yöntemleriyle elde edildiği sonucuna ulaşılmıştır. En yüksek deęer ise, D-rule ile elde edilmiştir.

Tablo 4.2: θ Değişimine Dayalı Durdurma Kuralının Kullanıldığı Koşullarda Elde Edilen Korelasyon, Yanlılık, RMSD, Standart Hata ve Yanıtlanan Madde Sayısı İstatistikleri

Sonlandırma Kuralı	Yetenek Kestirimi	Madde Seçim Yöntemi	r			Yanlılık (bias)			RMSD			Std.Hata			Ortalama Madde Sayısı
			$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	k
θ Değişimine Dayalı Sonlandırma Kuralı ($\Delta\theta < .05$)	EAP	D-rule	.94	.71	.71	-.0301	-.0678	.0515	.32	.53	.50	.30	.63	.62	24.1
		KL	.92	.42	.59	.0073	.0006	-.0289	.35	.60	.54	.37	.87	.77	18.1
		W-rule	.92	.41	.59	.0113	.0043	-.0282	.35	.61	.54	.37	.88	.77	18.0
		W-rule (.8,.1,.1)	.94	.39	.53	-.0031	.0173	.0152	.32	.62	.57	.31	.88	.79	16.1
		T-rule	.94	.50	.62	-.0164	.0230	.0473	.32	.59	.55	.30	.80	.68	17.1
		T-rule (.8,.1,.1)	.94	.41	.55	-.0095	.0216	.0337	.32	.62	.58	.30	.86	.75	16.3
	MAP	D-rule	.94	.70	.70	-.0069	.0639	.0768	.32	.52	.49	.33	.63	.61	21.4
		KL	.92	.41	.59	.0208	.0013	-.0072	.36	.60	.54	.38	.87	.76	17.2
		W-rule	.92	.40	.59	.0218	.0037	-.0032	.36	.61	.54	.38	.87	.76	17.1
		W-rule (.8,.1,.1)	.94	.36	.53	.0108	.0078	.0320	.32	.62	.56	.32	.88	.77	15.1
		T-rule	.94	.50	.63	.0030	.0086	.0684	.32	.59	.53	.31	.79	.65	16.6
		T-rule (.8,.1,.1)	.94	.42	.53	.0088	.0055	.0570	.32	.61	.57	.32	.85	.74	15.0

Gramer (θ_2) için tabloda yer verilen bulgulara göre, gerçek ve kestirilen θ değerleri arasındaki korelasyon değerleri arasında en büyük değer .71 ile D-rule madde seçim yönteminin kullanıldığı koşulda elde edilmiştir. Diğer madde seçim yöntemleriyle elde edilen değerlerin, bu değere göre oldukça düşük olduğu dikkat çekmektedir. Yanlılık değerlerinden 0'a en yakın olanı, .0006 ile KL madde seçim yöntemine aittir. .53 ile .62 arasındaki RMSD değerlerinden en düşüğü, D-rule madde seçim yöntemiyle elde edilmiştir. Son olarak elde edilen standart hata değerlerinin .63 ile .88 arasında değiştiği görülmüştür. Tüm bu değerler ışığında, θ_2 için elde edilen istatistiklerin neredeyse tamamında en iyi değerlerin D-rule madde seçim yönteminin kullanıldığı koşulda elde edildiği sonucuna ulaşılmıştır.

Kelime bilgisi (θ_3) için elde edilen değerler incelendiğinde; gerçek ve kestirilen θ değerleri arasındaki korelasyonlar .53 ile .71 arasında değişirken, en yüksek değer D-rule madde seçim yöntemiyle elde edildiği görülmektedir. Yanlılık değerlerine bakıldığında ise 0'a en yakın değer W-rule madde seçim yöntemine ait olduğu bulgulanmıştır. RMSD'nin .50 ile .58 arasında değişim gösteren değerleri arasında en düşük değer D-rule madde seçim yöntemiyle elde edilmiş, .62 ile .79 arasındaki ortalama standart hata değerlerinde en düşüğü de yine D-rule madde seçim yönteminin kullanıldığı koşulda ortaya çıkmıştır. Elde edilen sonuçlara göre dikkat çeken bir diğer husus ise; D-rule madde seçim yönteminin kullanıldığı koşulda gramer ve kelime bilgisi boyutlarına ait korelasyon, RMSD ve standart hata değerlerinin birbirine oldukça yakınken, diğer madde seçim yöntemlerinde ise bu boyutlara ait söz konusu istatistikler arasında önemli farklılıklar görülmüş olmasıdır.

Çalışma kapsamındaki altı madde seçim yöntemlerinin her birinin θ değişimine dayalı sonlandırma kuralı ve MAP yetenek kestirimiyle kullanıldığı farklı çok boyutlu BBT koşullarına ait Tablo 4.2'de yer alan değerler incelendiğinde; genel yetenek düzeyi (θ_9) için gerçek ve kestirilen θ değerleri arasındaki korelasyon değerlerinin EAP yöntemiyle elde edilen değerlerle bire bir aynı olduğu görülmektedir. Yanlılık değerleri incelendiğinde ise, -.0069 ile .0218 arasında değişken bu değerler arasında 0'a en yakın olanın T-rule madde seçim yönteminin uygulandığı koşulda elde edildiği bulgulanmıştır. RMSD'nin .32 ile .36 arasında değişen değerleri arasında en yüksek değerler KL ve W-rule madde seçim yöntemleriyle elde edilmişken, diğer tüm koşullarda bulgularanan değerler birbirine eşit ve .32'dir. Standart hatanın .31 ile .38 arasında değişen değerlerinden en düşüğü T-rule

madde seçim yöntemiyle elde edilmiştir. Test sonlandığında yanıtlanmış olan madde sayılarının ortalamasına bakıldığında ise T-rule ile elde edilen değerin en düşük olduğu sonucuna varılmıştır.

Gramer (θ_2) için elde edilen değerler incelendiğinde, gerçek ve kestirilen θ değerlerinin .36 ile .70 arasında değiştiği görülmektedir. En yüksek korelasyon değeri, EAP yönteminde olduğu gibi diğer yöntemlerle elde edilen değerlerden oldukça yüksek olup D-rule ile elde edilmiştir. Yanlılık değerleri incelendiğinde ise bu değerler arasında da en düşüğünün .0013 ile KL madde seçim yöntemine ait olduğu bulgulanmıştır. Sırasıyla .52 ile .62 ve .63 ile .88 arasında değer alan RMSD ve standart hata kestirimleri için de en düşük değerler yine D-rule madde seçim yöntemiyle elde edilmiştir.

Kelime bilgisi (θ_3) için tabloda yer verilen bulgulara göre, gerçek ve kestirilen θ değerleri için .53 ile .70 arasında değişen korelasyonlar arasında en yüksek değerin D-rule madde seçim yöntemiyle elde edildiği görülmektedir. En düşük yanlılık değeri, -.0032 ile W-rule madde seçim yöntemiyle elde edilmiştir. RMSD'nin .49 ile .57 arasında değişen değerleri ve .61 ile .77 arasında değişen standart hata değerleri arasında da en düşük değerlerin D-rule madde seçim yönteminin kullanıldığı koşullarda elde edildiği sonucuna ulaşılmıştır.

4.3. Alt Problem 3: Sabit madde sayısına dayalı sonlandırma kuralının kullanıldığı çok boyutlu BBT uygulamalarında farklı yetenek kestirim yöntemleri (EAP ve MAP) ve madde seçim kurallarına (D-rule, KL, W-rule, ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule) göre gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon, yanıtlanan ortalama madde sayısı, yanlılık (bias) ve RMSD değerleri nasıldır?

Araştırmanın bu bölümünde, her bireyin 20 maddeyi yanıtlamasıyla testin sonlandırılmasına dayanan BBT uygulaması çerçevesinde farklı yetenek kestirim ve madde seçim yöntemlerinin kullanılmasıyla elde edilen bulgular yer almaktadır. Söz konusu bulgulara, Tablo 4.3'te yer verilmiştir. İlk kısımda EAP, sonrasında ise MAP yetenek kestirim yöntemlerinin kullanıldığı koşullara ait sonuçlar raporlanmıştır.

Tablo 4.3: Sabit Madde Sayısına Dayalı Durdurma Kuralının Kullanıldığı Koşullarda Elde Edilen Korelasyon, Yanlılık, RMSD, Standart Hata ve Yanıtlanan Madde Sayısı İstatistikleri

Sonlandırma Kuralı	Yetenek Kestirimi	Madde Seçim Yöntemi	r			Yanlılık (bias)			RMSD			Std.Hata		
			$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$
Sabit Madde Sayısına Dayalı Sonlandırma Kuralı (k = 20)	EAP	D-rule	.93	.69	.69	-.0315	.0637	.0481	.33	.53	.51	.34	.65	.63
		KL	.93	.42	.60	.0085	-.0023	-.0344	.33	.61	.54	.37	.87	.76
		W-rule	.93	.42	.61	.0108	-.0012	-.0336	.33	.61	.53	.35	.87	.76
		W-rule (.8,.1,.1)	.94	.42	.60	-.0093	.0183	.0115	.30	.62	.55	.29	.85	.72
		T-rule	.95	.53	.66	-.0154	.0277	.0346	.30	.58	.53	.29	.78	.66
		T-rule (.8,.1,.1)	.95	.50	.58	-.0097	.0231	.0266	.30	.59	.57	.28	.81	.72
	MAP	D-rule	.93	.69	.69	-.0093	.0607	.0751	.33	.52	.50	.33	.64	.61
		KL	.93	.43	.61	.0170	-.0054	-.0086	.33	.60	.53	.35	.86	.75
		W-rule	.93	.42	.61	.0184	.0009	-.0061	.33	.61	.53	.35	.86	.75
		W-rule (.8,.1,.1)	.95	.42	.61	.0078	.0071	.0255	.29	.62	.53	.30	.84	.70
		T-rule	.95	.53	.66	-.0027	.0144	.0601	.29	.58	.52	.29	.76	.64
		T-rule (.8,.1,.1)	.95	.50	.58	.0037	.0124	.0495	.29	.59	.55	.29	.80	.80

Tablo 4.3'te ilk olarak; sabit sayıdaki maddenin yanıtlanmasıyla testin sonlandırıldığı ve EAP kestiriminin kullanıldığı durumlarda farklı madde seçim yöntemlerinin kullanılmasıyla elde edilen değerler görülmektedir. Buna göre, gerçek ve kestirilen θ değerleri arasındaki korelasyonun genel yetenek düzeyi için (θ_g) .93 ile .95 arasında değiştiği görülmektedir. Söz konusu değerler birbirine oldukça yakındır. Yanlılık değerlerine bakıldığında 0'a en yakın değer .0085 ile KL madde seçim yöntemine ait olduğu sonucuna ulaşılmıştır. Ayrıca D-rule, KL ve W-rule ile elde edilen değerler .33 iken, diğer madde seçim yöntemlerinin kullanıldığı durumlara ait RMSD değerleri .30 olarak bulunmuştur. Son olarak .28 ile .37 arasında değişen standart hata değerlerinin incelenmesi sonucunda en düşük değerin ağırlıklandırılmış T-rule madde seçim yöntemiyle elde edildiği bulgulanmıştır.

Gramer (θ_2) için elde edilen değerler incelendiğinde, .42 - .69 aralığındaki gerçek ve kestirilen θ değerleri arasındaki korelasyonlardan en yüksek olan D-rule madde seçim yönteminin kullanılmasıyla elde edildiği görülmektedir. -.0012 ile 0'a en yakın yanlılık değeri ise, KL madde seçim yönteminin kullanıldığı koşula aittir. .53 ile .62 arasındaki RMSD değerlerinin en düşüğü D-rule madde seçim yöntemiyle elde edilmiştir. Son olarak .65 ile .87 arasında değişen ortalama standart hata değerlerinin en düşüğünün de yine D-rule madde seçim yöntemiyle elde edildiği sonucuna ulaşılmıştır.

Kelime bilgisi (θ_3) için tablo 6'da yer verilen bulgulara göre, .58 ile .69 arasında değişen gerçek ve kestirilen θ değerleri arasındaki korelasyonların en yükseği D-rule ile elde edilirken, .0115 ile en düşük yanlılık değerlerinin ağırlıklandırılmış W-rule ile elde edildiği görülmektedir. .51 ile .57 arasındaki RMSD değerlerinden en düşüğü de yine D-rule madde seçim yönteminin kullanıldığı koşula aittir. Son olarak testin sonlandırılmasıyla elde edilen standart hata değerleri incelenmiş olup, .63 ile .75 arasında değişen bu değerlerin en düşüğünün yine D-rule madde seçim yönteminin kullanıldığı koşulda elde edildiği sonucuna varılmıştır.

Tablo 4.3'te, sabit sayıdaki maddenin yanıtlanmasıyla testin sonlandırıldığı ve MAP kestiriminin kullanıldığı durumlarda, farklı madde seçim yöntemlerinin kullanılmasıyla elde edilen bulgular incelendiğinde; genel yetenek düzeyi için (θ_g) gerçek ve kestirilen θ değerleri arasındaki korelasyonun .93 ile .95 arasında birbirine oldukça yakın değerler aldığı görülmekte olup, en düşük değerler T-rule ile birlikte

ağırlıklandırılmış madde seçim yöntemleriyle elde edildiği görülmektedir. Yanlılık değerleri incelendiğinde 0'a en yakın değer $-.0027$ ile ağırlıklandırılmış T-rule madde seçim yöntemine ait olduğu sonucuna ulaşılmıştır. RMSD değerleri incelendiğinde ise korelasyon değerlerine benzer bir sonuç görülmektedir. Tablodaki madde seçim yöntemlerinden ilk üçü ile elde edilen değerler birbirine eşit ve $.33$ olarak bulunmuş, diğer üç yöntem için ise $.29$ olarak kestirilmiştir. Son olarak $.29$ ile $.35$ arasında değişen standart hata değerlerinin incelenmesi sonucunda en düşük değerlerin ağırlıklandırılmış T-rule ve madde seçim yöntemlerinin kullanıldığı koşullarda elde edildiği sonucuna ulaşılmıştır.

Gramer (θ_2) için elde edilen değerler incelendiğinde, gerçek ve kestirilen θ değerleri arasındaki $.42 - .69$ aralığındaki korelasyonlardan en yüksek olan D-rule madde seçim yönteminin kullanılmasıyla elde edildiği görülmektedir. W-rule madde seçim yönteminin kullanıldığı koşulda, $.0009$ ile 0'a en yakın yanlılık değeri elde edilmiştir. RMSD'nin $.52$ ile $.62$ arasında değişen değerlerinin en düşüğü de D-rule madde seçim yöntemiyle elde edilmiş olup, $.64$ ile $.86$ arasında değişen ortalama standart hata değerleri için diğer değerlere benzer olarak en düşük değerin D-rule madde seçim yöntemleriyle elde edildiği bulgulanmıştır.

Kelime bilgisine (θ_3) ait tabloda yer alan bulgulara göre; $.58$ ile $.69$ arasında değişen gerçek ve kestirilen θ değerleri arasındaki korelasyonların en yüksek olan D-rule madde seçim yöntemine ait olduğu, $-.0061$ ile en düşük yanlılık değerlerinin ise W-rule ile elde edildiği bulgusuna ulaşılmıştır. RMSD'nin $.50$ ile $.55$ arasındaki değerleri ile testin sonlandırılmasıyla elde edilen ve standart hataya ait $.61$ ile $.80$ arasında değişen değerleri arasında en düşüğünün D-rule madde seçim yönteminin kullanıldığı koşulda elde edildiği görülmüştür.

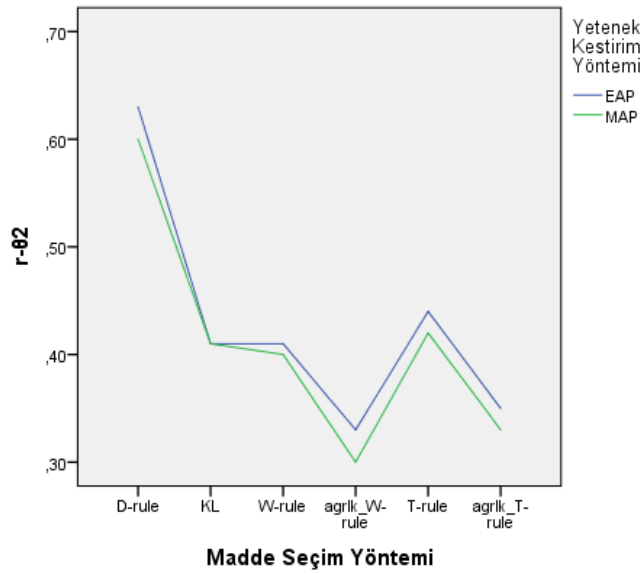
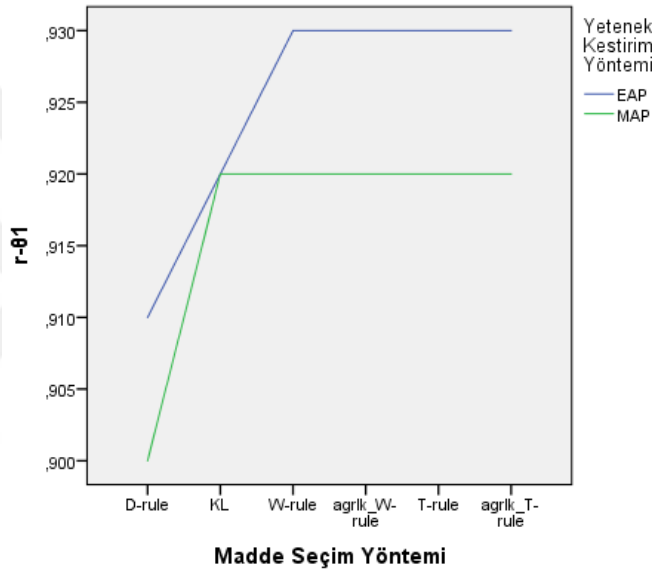
4.4. Alt Problem 4: Simülasyon sonuçlarına göre her bir sonlandırma kuralı için en uygun çok boyutlu BBT koşulu nedir?

Araştırmanın bu bölümünde, simülasyon çalışmaları sonucu edilen bulgulardan yola çıkarak her bir boyutta farklı madde seçim ve yetenek kestirim yöntemlerine göre; gerçek ve kestirilen θ değerleri arasındaki korelasyon, yanlılık, RMSD, standart hata ve yanıtlanan ortalama madde sayısı değerleri için elde edilen grafiklerin yorumlarına yer verilmiş ve buradan yola çıkarak, her bir sonlandırma kuralı çerçevesinde en uygun yetenek kestirimi ve madde seçim yöntemi ortaya

koyulmuştur. Son aşamada ise, söz konusu koşullar karşılaştırılarak gerçek-zamanlı çok boyutlu BBT uygulamasında kullanılacak algoritmaya karar verilmiştir. Bu amaçla elde edilmiş grafiklerden yalnızca korelasyon (r) ve yanıtlanan ortalama madde sayısı için elde edilmiş olanlar metin içerisinde yer almaktadır. Diğer tüm grafikler, EK-6'da görülebilir.

a. Standart hataya dayalı sonlandırma kuralı için en uygun yetenek kestirim ve madde seçim yöntemi hangisidir?

Standart hataya dayalı durdurma kuralı çerçevesinde her bir boyut için elde edilen korelasyon grafikleri aşağıda görülmektedir.





Şekil 4.1: Standart Hataya Dayalı Sonlandırma Kuralı için θ Değerleri Arasındaki Korelasyon Grafikleri

Şekil 4.1’de yer alan grafiklerden ilki, genel yetenek düzeyi için (θ_0) elde edilmiştir. Buna göre, farklı madde seçim yöntemleriyle θ_0 için elde edilen korelasyon değerleri arasında büyük farklılıklar olmadığı, elde edilen değerlerin .90-.93 arasında değiştiği görülmektedir. KL hariç tüm madde seçim yöntemlerinde, EAP ile elde edilen korelasyonların MAP ile elde edilenlere göre daha yüksek olduğu görülmektedir.

θ_2 için elde edilen ikinci grafiğe bakıldığında, D-rule ile elde edilen korelasyon değerinin diğer değerlerden oldukça büyük olduğu dikkat çekmektedir. En düşük korelasyonlar, madde seçim yönteminde genel yeteneğe öncelik veren ağırlıklandırılmış yöntemlerde görülmektedir. Yetenek kestirim yönteminin farklılaşmasının ise korelasyon değerleri üzerinde büyük bir farklılığa neden olmadığı sonucuna ulaşılmıştır.

Son olarak, θ_3 için elde edilen grafik incelenmiştir. Buna göre, D-rule yöntemiyle elde edilen korelasyonun en yüksek değere sahip olduğu görülmektedir. Bir önceki grafikte olduğu gibi, ağırlıklandırılmış yöntemler yine en düşük korelasyona sahiptirler. Yetenek kestirim yöntemine göre elde edilen korelasyon değerleri arasında önemli bir farklılık olmamakla birlikte, tüm madde seçim yöntemlerinde EAP ile elde edilen korelasyonların daha büyük olduğu görülmektedir.

Sonuç olarak; gerçek ve kestirilen yetenek düzeyleri arasındaki korelasyon değerlerinin genel yetenek için madde seçim yöntemine göre en fazla .3’lük küçük bir fark gösterdiği, bununla beraber özellikle ikinci boyut olmak üzere diğer iki

boyutta, D-rule ile elde edilen korelasyonların diğer yöntemlerle elde edilenlere göre oldukça yüksek olduğu görülmüştür. Bu nedenle, tüm boyutlarda gerçek ve kestirilen θ değerleri arasındaki korelasyonlar açısından D-rule madde seçim yönteminin kullanılmasının uygun olduğu görülmüştür. Yetenek kestirim yönteminin farklılaşmasıyla elde edilen korelasyon değerlerinin ise birbirlerine oldukça yakın oldukları söylenebilir.

Yanlılık değerlerinin her bir boyut için elde edilen grafikleri, EK-6.1'de yer almaktadır. Yanlılık değerlerine yönelik grafiklerden ilkinde, θ_9 için elde edilen yanlılık değerleri görülmektedir. Buna göre; 0'a en yakın değer KL madde seçim yöntemiyle elde edildiği, yetenek kestiriminin farklılaşmasının bu değer için 0'dan önemli uzaklaşmaya neden olmadığı görülmüştür. D-rule ile elde edilen değerler, diğer yöntemlerle elde edilenlere nispeten en yanlı değerlerdir.

İkinci grafikteki yanlılık değerleri, θ_2 için elde edilmiştir. Burada da ilk boyutta olduğu gibi KL ile elde edilen değerler en yansız değerlerdir. W-rule ile elde edilen değerler de KL ile elde edilenlere oldukça yakındır. Diğer yöntemlere nazaran 0'dan en çok sapma gösteren değer ise, ilk boyutta olduğu gibi D-rule ile elde edilmiştir.

Son grafikte θ_3 için elde edilen değerler yer almaktadır. MAP yetenek kestirim yönteminin kullanıldığı durumlarda KL ve W-rule madde seçim yöntemleriyle elde edilen değerlerin 0'a en yakın değerler olduğu sonucuna ulaşılmıştır.

Elde edilen sonuçlara göre, yanlılık değerlerinin 0'a yakınlığı göz önüne alınarak KL madde seçim ve MAP yetenek kestirim yöntemlerinin kullanıldığı koşullarda en düşük yanlılık değerlerinin elde edildiği sonucuna ulaşılmıştır.

Post-hoc simülasyonlar sonucu tüm yetenek düzeyleri için standart hataya dayalı sonlandırma kuralı çerçevesinde farklı madde seçim ve yetenek kestirim yöntemleri kullanıldığında elde edilen RMSD değerleri EK-6.1'de görülmektedir.

RMSD değerlerine ait grafiklerden ilki, θ_9 için elde edilmiştir. Buna göre, RMSD değerleri .3'lük bir aralıkta değişmektedir. T-rule haricindeki yöntemlerde EAP ile elde edilen değerler, MAP ile elde edilenlere göre daha düşüktür. En yüksek değerler ise D-rule madde seçim yönteminin kullanıldığı koşula aittir.

İkinci yetenek düzeyi için elde edilen grafikte ise en düşük RMSD değerinin elde edildiği D-rule haricindeki tüm durumlarda yetenek kestirim yönteminin farklılaşmasının elde edilen değerleri etkilemediği görülmektedir. En yüksek

değerler ise, ağırlıklandırılmış madde seçim yöntemlerinin kullanıldığı koşullara aittir.

Üçüncü ve son grafik θ_3 için elde edilmiştir. En düşük değer D-rule ile elde edilmiş olup, bu değer yetenek kestirim yöntemine göre farklılaşmadığı bulgusuna ulaşılmıştır. İkinci yetenek düzeyi için olduğu gibi burada da en yüksek RMSD değerleri ağırlıklandırılmış yöntemlerle elde edilmiştir.

Bu sonuçlar doğrultusunda RMSD değerleri açısından; θ_g için elde edilen en büyük değer D-rule madde seçim yönteminin kullanıldığı duruma ait olmakla birlikte bu değer diğerlerinden farkının oldukça küçük olduğu, diğer boyutlar içinse D-rule ile elde edilen değerlerin diğer değerlerden oldukça düşük olduğu görülmüştür. Ağırlıklandırılmış yöntemler genel yetenek düzeyi için nispeten düşük değerler verse de, diğer boyutlar için elde ettikleri hata değerleri oldukça büyüktür. Bu nedenle, üç boyut için elde edilen RMSD değerleri açısından D-rule madde seçim yönteminin kullanılmasının uygun olduğu sonucuna varılmıştır.

Her bir yetenek düzeyinde farklı madde seçim ve yetenek kestirim yöntemlerinin kullanılmasıyla standart hata değerlerinin değişimini gösteren EK-6'da yer alan grafiklere dair yorumlar aşağıda yer almaktadır.

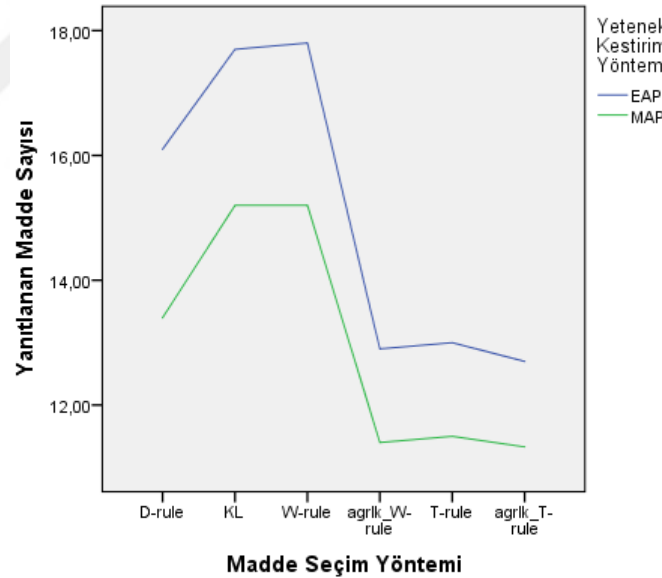
Standart hatayla ilgili EK-6.1'deki grafiklerden ilki, genel yetenek düzeyi için elde edilmiştir. Araştırma kapsamında standart hataya dayalı sonlandırma kuralı, yalnızca genel yetenek düzeyi için .4 olarak baz alınmış olduğundan, grafikte yer alan tüm koşullar için kestirilen standart hata ortalamalarının .4'ün altında olduğu görülmektedir. Bunun yanında tüm madde seçim yöntemlerinin EAP ile kullanıldığı koşullarda standart hata değeri, MAP kullanıldığı koşullara göre daha küçüktür. En düşük değer ise, ağırlıklandırılmış T-rule madde seçim yönteminin kullanıldığı koşulda elde edilmiştir.

İkinci grafikte θ_2 elde edilen standart hata değerleri yer almaktadır. θ_g için elde edilen değerlerin aksine yetenek kestirim yönteminin farklılaşmasının sadece iki madde seçim yöntemi için .1'lik bir farklılığa neden olduğu görülmektedir. Diğer madde seçim yöntemleriyle elde edilen standart hata değerlerinin hepsi .85'in üzerindeyken, D-rule madde seçim yöntemiyle çok daha düşük bir standart hata değerinin elde edildiği dikkat çekmektedir. En yüksek hata değerleri ise, ağırlıklandırılmış yöntemlerin kullanıldığı duruma aittir.

Son grafikte θ_3 için elde edilen değerler incelendiğinde, D-rule madde seçim yönteminin kullanıldığı koşulda elde edilen standart hata değerinin en düşük değer olduğu bulgusuna ulaşılmıştır. Yetenek kestirim yönteminin farklılaşmasının ise elde edilen değerlere etkisinin çok düşük düzeyde olduğu görülmüştür.

Bu üç grafikten elde edilen sonuçlara göre; θ_g için elde edilen değerlerin birbirine yakın olduğu, θ_2 ve θ_3 için ise D-rule ile elde edilen değerlerin diğerlerine göre oldukça düşük olduğu bulgulanmıştır. Bu nedenle standart hata değerleri açısından üç boyut birlikte göz önüne alındığında, D-rule madde seçim yönteminin uygun olduğu sonucuna varılmıştır. Yetenek kestirim yönteminin farklılaşmasının ise, aynı madde seçim yöntemiyle elde edilen değerler üzerinde dramatik bir etkiye sahip olmadığı görülmüştür.

Şekil 4.2’de, standart hataya dayalı olarak sonlandırma kuralı çerçevesinde farklı yetenek ve madde seçim yöntemleri kullanıldığında, testin bitimiyle birlikte yanıtlanan madde sayıları yer almaktadır.



Şekil 4.2: Standart Hataya Dayalı Sonlandırma Kuralı için Yanıtlanan Madde Sayısı Grafiği

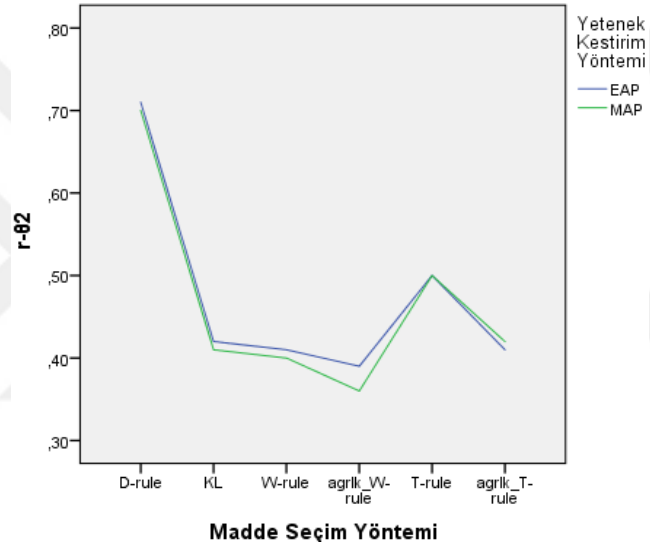
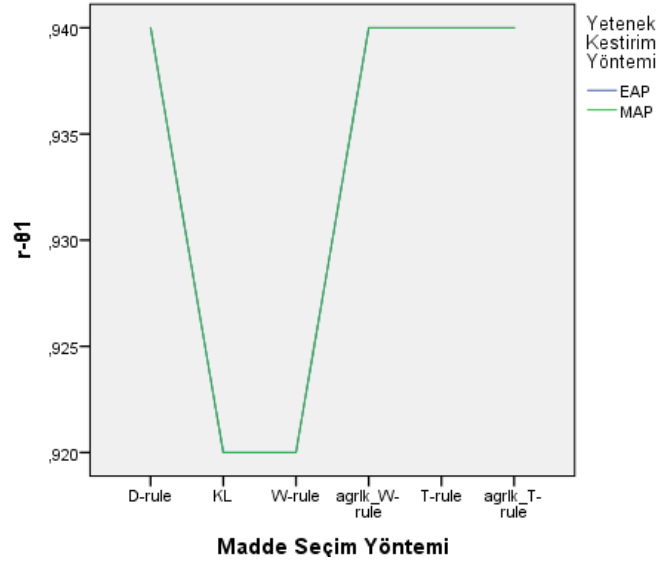
Yukarıdaki grafik incelendiğinde, tüm madde seçim yöntemleri için yanıtlanan madde sayılarının iki farklı yetenek kestirim yönteminde paralellik gösterdiği, MAP'ın kullanıldığı tüm koşullarda yanıtlanan madde sayısının daha az olduğu görülmektedir. Ağırlıklandırılmış yöntemlerin kullanıldığı koşullarda yanıtlanan madde sayısının daha düşük olduğu dikkat çekmektedir. Bu durumun nedeni olarak, standart hataya dayalı sonlandırma kuralının yalnızca genel yetenek için koyulmuş

olması ve ağırlıklandırılmış yöntemlerdeki bu ağırlıklandırmanın da genel yetenek düzeyi için yapılmış olması düşünülmektedir. Ayrıca T-rule madde seçim yöntemiyle yanıtlanan madde sayılarının ortalaması ağırlıklandırılmış yöntemlere oldukça yakındır. Yanıtlanan madde sayıları dikkate alındığında, standart hataya dayalı durdurma kuralının kullanılması durumunda en uygun çok boyutlu BBT koşulunun MAP yetenek kestirimi ve ağırlıklandırılmış T-rule madde seçim yöntemi olduğu sonucuna ulaşılmıştır.

Standart hataya dayalı durdurma kuralı çerçevesinde farklı yetenek kestirim ve madde seçim yöntemlerinin kullanıldığı koşullarda elde edilen korelasyon, yanlılık, RMSD, standart hata ve yanıtlanan madde sayılarına ait grafikler yukarıda yorumlanmıştır. Buna göre; genel yetenek için elde edilen istatistiklerin genel olarak ağırlıklandırılmış yöntemler lehine olduğu, bununla birlikte farklı madde seçim yöntemleriyle elde edilen değerler arasında küçük farklar bulunduğu görülmektedir. İkinci ve üçüncü boyutlar içinse, D-rule madde seçim yöntemleriyle elde edilen değerler diğer değerlere göre oldukça farklıdır. İkinci ve üçüncü boyutlarda elde edilen hata değerlerini kayda değer miktarda düşürürken, genel yetenek düzeyi için diğer madde seçim yöntemlerine çok yakın sonuçlar vermesi nedeniyle standart hataya dayalı durdurma kuralı çerçevesinde D-rule madde seçim yönteminin kullanılmasının uygun olduğu sonucuna varılmıştır. Yetenek kestirim yöntemleri ise elde edilen tüm istatistiklerde birbirine oldukça yakın, kimi zaman aynı sonuçları vermişlerdir. Ancak, yanıtlanan madde sayılarına bakıldığında MAP yetenek kestiriminin kullanıldığı durumlarda yanıtlanan ortalama madde sayısının EAP ile yanıtlananlara göre yaklaşık %17 daha düşük olması nedeniyle standart hataya dayalı durdurma kuralı çerçevesinde D-rule madde seçim yöntemiyle MAP yetenek kestirim yönteminin kullanılmasının uygun olduğu sonucuna varılmıştır.

b. 0 değişimine dayalı sonlandırma kuralı için en uygun yetenek kestirim ve madde seçim yöntemi hangisidir?

Bu bölümde, çok boyutlu BBT uygulamasında, tüm boyutlar için son iki yetenek kestirimi arasındaki farkın .05'in altına düşmesiyle testin sonlandırması çerçevesinde farklı yetenek kestirim ve madde seçim yöntemlerinin kullanılmasıyla elde edilen grafiklerin yorumlarına yer verilmiştir. Şekil 4.3'te, gerçek ve kestirilen yetenek parametreleri arasındaki korelasyonların farklı koşullara göre değişimini gösteren grafikler yer almaktadır.



Şekil 4.3: θ Değişimine Dayalı Sonlandırma Kuralı için θ Değerleri Arasındaki Korelasyon Grafikleri

Şekil 4.3'te görülen grafiklerden ilki, genel yetenek düzeyi (θ_g) için elde edilmiştir. Buna göre, farklı madde seçim yöntemlerinin tamamıyla elde edilen korelasyon değerlerinin .90'dan büyük olduğu sonucuna varılmıştır. Ayrıca, bu değerlerin .2'lik oldukça dar bir aralıkta değişim gösterdiği görülmektedir. Tüm madde seçim yöntemlerinde iki yetenek kestirim yöntemi için de aynı sonuçların elde edildiği dikkat çekmektedir.

θ_2 için kestirilen değerlere bakıldığında, standart hataya dayalı durdurma kuralında olduğu gibi, D-rule ile elde edilen korelasyon değerinin diğer madde seçim yöntemleriyle elde edilen korelasyonlara göre oldukça yüksek olduğu görülmektedir. Yetenek kestirim yönteminin etkisinin ise önemsenmeyecek düzeyde olduğu sonucuna varılmıştır.

θ_3 için elde edilen son grafiğe göre, D-rule ile elde edilen değerlerin diğer boyutlara göre oldukça yüksek olduğu, en düşük değerlerin ise ağırlıklandırılmış yöntemlerle elde edildiği bulgusuna ulaşılmıştır.

Elde edilen sonuçlar doğrultusunda, korelasyon değerleri açısından tüm boyutlarda en yüksek değerini veren D-rule madde seçim yönteminin kullanılmasının uygun olduğu, yetenek kestirim yöntemlerinin ise neredeyse bire bir aynı sonuçları verdiği sonucuna varılmıştır.

EK-6.2'de yer alan yanlılık değerlerine ait grafikler incelendiğinde, θ_g için farklı yetenek kestirimleriyle her bir madde seçim yöntemiyle elde edilen değerlerin grafiklerinin paralel olduğu görülmektedir. 0'a en yakın değer ise, MAP yetenek kestirim ve D-rule madde seçim yöntemlerinin kullanıldığı koşulda elde edilmiştir.

θ_2 için elde edilen ikinci grafik incelendiğinde, D-rule haricindeki madde seçim yöntemleriyle MAP yetenek kestirim yönteminin kullanıldığı durumlarda elde edilen yanlılık değerleri 0'a oldukça yakındır. D-rule ile elde edilen değerlerin yetenek kestirimine göre farklılaşmakla birlikte, 0'a uzaklıklarının birbirine çok yakın olduğu görülmektedir. En yansız değer ise, KL ile elde edilmiştir.

θ_3 için elde edilen üçüncü grafik incelendiğinde ise, MAP ile kullanılan W-rule yöntemleri ile 0'a en yakın değerleri verdiği görülmektedir.

Elde edilen bulgulara göre, yanlılık değerleri açısından tüm boyutlar için KL madde seçim yönteminin optimum sonuçları verdiği sonucuna ulaşılmıştır.

RMSD değerlerinin farklı yetenek kestirim ve madde seçim yöntemlerine göre değişimini incelemek amacıyla elde edilen ve EK-6.2'de yer alan grafiklerin ilki, genel yetenek düzeyi için elde edilmiş olup, sonuçların korelasyon grafikleriyle benzerlik gösterdiği dikkat çekmektedir. En yüksek değerler KL ve W-rule ile elde edilirken, diğer dört yöntemle elde edilen değerler birbirine eşittir. Bu dört yöntemle elde edilen değerlere yetenek kestiriminin farklılaşmasının herhangi bir etkisi olmadığı görülmektedir.

İkinci grafik, θ_2 için elde edilmiştir. En düşük değerlerin D-rule ile elde edildiği, en yüksek hata değerlerine ise ağırlıklandırılmış yöntemlerin kullanıldığı koşullarda ulaşıldığı görülmektedir.

θ_3 için elde edilen üçüncü grafiğe bakıldığında en düşük değer MAP yetenek kestirimiyle D-rule madde seçim yöntemlerinin kullanıldığı koşulda elde edildiği görülmektedir. En yüksek RMSD değerleri ise, yine ağırlıklandırılmış madde seçim yöntemlerinin kullanıldığı koşullarda elde edilmiştir.

Elde edilen grafiklerden çıkan sonuçlara göre, D-rule madde seçim yöntemi ve MAP yetenek kestiriminin kullanıldığı koşullarda en düşük RMSD değerlerine ulaşıldığı sonucuna varılmıştır.

Standart hata değerlerinin, yetenek kestirim ve madde seçim yöntemlerine göre değişimine ait EK-6.2'de yer alan grafiklerin yorumları aşağıda görülmektedir.

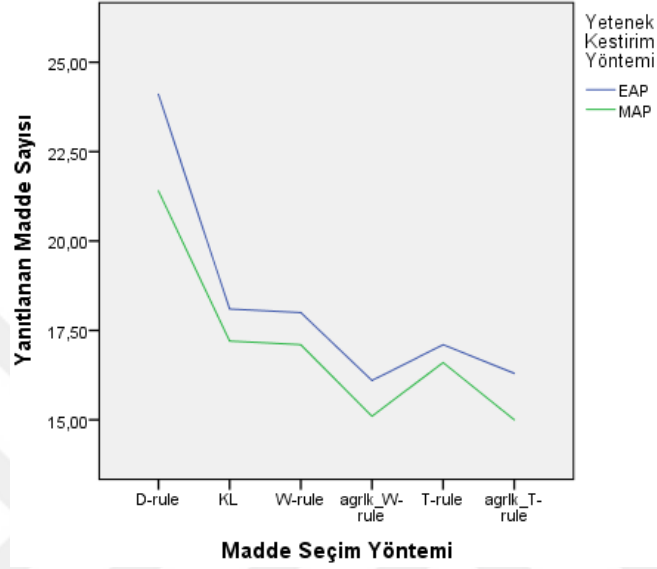
Standart hata istatistikleri için elde edilen grafiklerden ilki, genel yetenek (θ_g) için elde edilmiştir. Bu grafikteki değerlere göre, EAP ile elde edilen standart hata değerleri tüm madde seçim yöntemlerinde MAP ile elde edilenlere göre daha düşüktür. En düşük standart hata değerleri ise D-rule ile birlikte T-rule madde seçim yöntemleriyle elde edilmiştir.

İkinci grafik θ_2 için elde edilmiştir. Grafik incelendiğinde, D-rule ile elde edilen standart hata ortalamasının iki yetenek kestirim yöntemi ile de aynı ve diğer madde seçim yöntemlerine göre oldukça düşük olduğu görülmüştür.

Son grafik θ_3 içine elde edilmiş olup, en düşük standart hata değerinin MAP yetenek kestirimi ve D-rule madde seçim yöntemiyle elde edildiği bulgulanmıştır. Kalan beş madde seçim yönteminden T-rule kullanıldığı koşul haricindeki diğer tüm koşullarda ise, birbirine yakın ve yüksek standart hata değerlerinin elde edildiği görülmektedir.

Standart hata değerleri için elde edilen grafiklere dayanarak, üç boyut için de uygun madde seçim yönteminin D-rule olduğu, yetenek kestirim yöntemine göre elde edilen değerlerde büyük bir farklılaşma olmadığı sonucuna ulaşılmıştır.

Şekil 4.4'te, farklı yetenek kestirim ve madde seçim yöntemlerinin kullanıldığı koşullarda yanıtlanan ortalama madde sayıları yer almaktadır.



Şekil 4.4: : θ Değişimine Dayalı Sonlandırma Kuralı için Yanıtlanan Madde Sayısı Grafiği

Şekil 4.4'deki grafik incelendiğinde, MAP yetenek kestirim yönteminin kullanıldığı durumlarda tüm madde seçim yöntemleriyle yanıtlanan madde sayılarının EAP ile yanıtlanan madde sayılarına göre daha düşük olduğu görülmektedir. En düşük değer ise, ağırlıklandırılmış T-rule madde seçim yönteminin kullanıldığı koşulda elde edilmiştir. Diğer tüm istatistikler için en iyi değerleri veren D-rule madde seçim yönteminin kullanıldığı koşulda ise yanıtlanan madde sayısının en yüksek değeri almış olması dikkat çekmektedir.

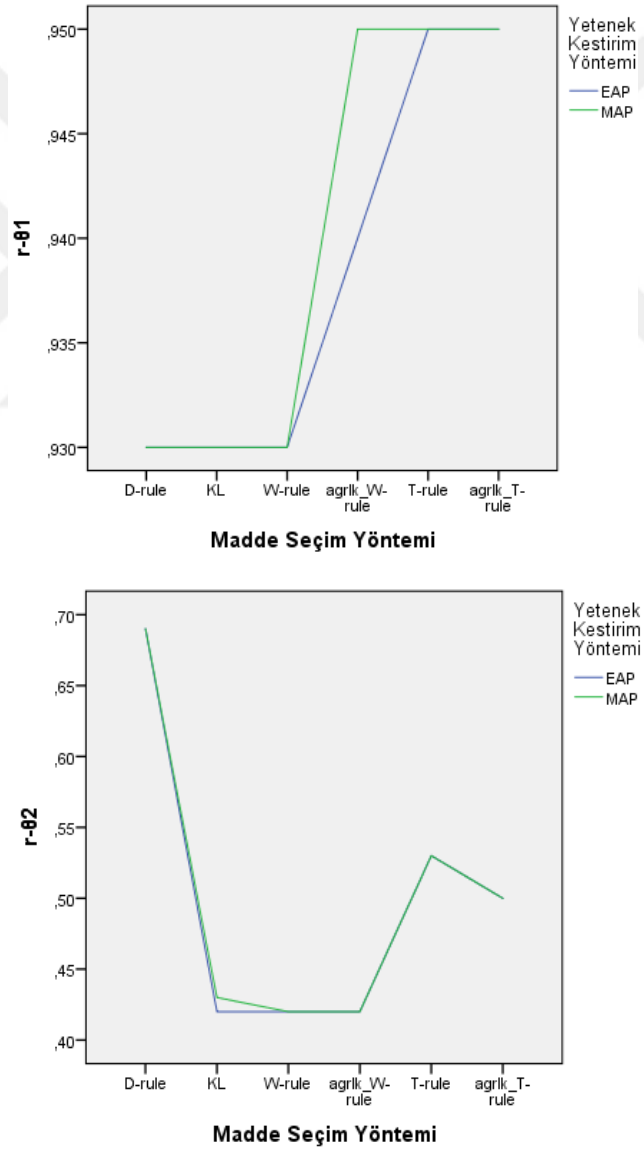
θ değişimine dayalı sonlandırma kuralı çerçevesinde, tüm boyutlar için çok boyutlu BBT uygulaması sonrası tüm koşullarda elde edilen istatistikler yukarıdaki şekillerde yer alan grafiklerde sunulmuştur. Yanıtlanan madde sayısı hariç tüm diğer durumlar için genel olarak D-rule madde seçim yönteminin daha iyi değerler vermesi nedeniyle söz konusu madde seçim yönteminin en uygun yöntem olduğu düşünülmektedir. D-rule madde seçim yönteminin kullanıldığı durumlarda; yetenek kestirim yöntemleri arasında elde edilen istatistiklere göre önemli farklılıklar bulunmamakla birlikte, benzer korelasyon ve hata istatistiklerinin daha az madde

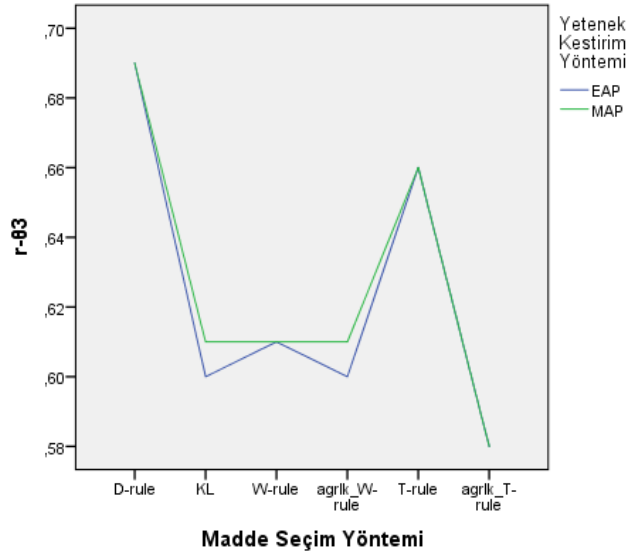
yanıtlanmasıyla elde edilmiş olması nedeniyle MAP yetenek kestirim yönteminin kullanılmasının uygun olduğu sonucuna ulaşılmıştır.

c. Sabit madde sayısına dayalı sonlandırma kuralı için en uygun yetenek kestirim ve madde seçim yöntemi hangisidir?

Araştırmanın bu bölümünde, 20 maddenin yanıtlamasıyla testin sonlandırılması çerçevesinde farklı yetenek kestirim ve madde seçim yöntemlerinin kullanıldığı koşullarda elde edilen çok boyutlu BBT istatistiklerinden yola çıkarak söz konusu sonlandırma kuralı için en uygun koşul belirlenmiştir.

Şekil 4.5'te, gerçek ve kestirilen θ değerleri arasındaki korelasyonlar yer almaktadır.





Şekil 4.5: Sabit Madde Sayısına Dayalı Sonlandırma Kuralı için θ Değerleri Arasındaki Korelasyon Grafikleri

Şekil 4.5'te yer alan grafiklerden ilki, genel yetenek θ_g için elde edilmiştir. Bu grafiğe göre, elde edilen tüm korelasyon değerleri çok yüksek ve birbirine oldukça yakındır. Ağırlıklandırılmış W-rule haricindeki tüm madde seçim yöntemlerinde yetenek kestirim yönteminin farklılaşması, elde edilen korelasyonları etkilememiştir.

İkinci grafik, θ_2 için elde edilmiş korelasyon grafiğidir. Buna göre, en yüksek korelasyon değerinin D-rule madde seçim yöntemiyle elde edildiği sonucuna ulaşılmış, yetenek kestirim yönteminin bu değer üzerinde bir etkisi olmadığı görülmüştür. Diğer madde seçim yöntemleriyle elde edilen korelasyon değerleri ise, D-rule ile elde edilen değere göre oldukça düşüktür.

θ_3 için elde edilen üçüncü grafiğe göre, en yüksek değer D-rule ile elde edilmiştir. İki yetenek kestirim yöntemiyle de söz konusu madde seçim yöntemi kullanılarak aynı değer elde edildiği görülmektedir.

Elde edilen grafiklerin yorumlanması sonucunda tüm boyutlar için gerçek ve kestirilen θ değerleri arasındaki korelasyonlar açısından D-rule madde seçim yönteminin kullanılmasının uygun olduğu sonucuna ulaşılmıştır.

Yanıluluk değerlerine dair EK-6.3'te yer verilen grafiklerden ilki, genel yetenek (θ_g) için elde edilmiştir. Bu grafiğe göre, diğer madde seçim yöntemleriyle elde edilen değerlerle karşılaştırıldığında 0'a en yakın yanıluluk değerlerinin D-rule ve KL ile elde edildiği görülmektedir.

θ_2 için elde edilen ikinci grafik incelendiğinde, KL ve W-rule ile 0'a en yakın değerlerin elde edildiği görülmektedir.

Son grafikte, θ_3 ile ilgili yanlılık değerlerine yer verilmiştir. Bu değerlerden MAP yetenek kestirim yöntemiyle birlikte KL ve W-rule madde seçim yöntemlerinin en yansız değerlere sahip olduğu görülmektedir.

Yorumlanan tüm bu veriler ışığında, sabit madde sayısına dayalı sonlandırma kuralı kullanıldığında tüm boyutlar için yanlılık değerleri açısından en uygun koşulun KL madde seçim yönteminin kullanıldığı durum olduğu sonucuna varılmıştır.

RMSD değerlerinin tüm boyutlarda farklı yetenek kestirim ve madde seçim yöntemlerinin kullanılmasıyla aldığı değerlerin değişimine ait grafikler EK-6.3'te görülmektedir. Söz konusu grafiklerden ilki, genel yetenek (θ_g) için elde edilmiştir. Grafikteki değerler yorumlandığında; en düşük değerlerin, MAP yetenek kestirim yöntemiyle birlikte ağırlıklandırılmış W-rule, T-rule ve ağırlıklandırılmış T-rule'nin kullanılmasıyla elde edildiği görülmektedir.

İkinci grafik, θ_2 için elde edilmiş olup en düşük RMSD değerinin D-rule madde seçim yöntemi ve MAP yetenek kestirimin kullanıldığı koşulda elde edildiği görülmektedir. Yetenek kestirim yöntemi, madde seçim yöntemlerinin yarısında fark yaratmazken, diğer kısmında MAP ile elde edilen değerlerin daha düşük olduğu sonucuna varılmıştır.

Son grafik, üçüncü boyut için elde edilmiştir. Grafikteki değerler incelendiğinde, W-rule haricindeki tüm madde seçim yöntemleri için MAP ile elde edilen RMSD değerlerinin EAP ile elde edilenlere göre daha düşük olduğu görülmektedir. En düşük değer ise, D-rule madde seçim yöntemiyle elde edilmiştir.

RMSD değerlerine ait grafikler yorumlandığında, tüm boyutlarda en düşük RMSD değerlerinin MAP yetenek kestirim yöntemiyle birlikte D-rule madde seçim yöntemleriyle elde edileceği sonucuna ulaşılmıştır.

Sabit madde sayısına dayalı sonlandırma kuralı çerçevesinde tüm boyutlarda farklı koşullar için elde edilen standart hata değerlerini gösteren EK'te yer alan grafiklere ait yorumlar aşağıda görülmektedir.

EK-6.3'te yer alan, sabit madde sayısına dayalı sonlandırma kuralının kullanıldığı durumlarda standart hata değerleri için elde edilen grafiklerden ilki, genel yetenek

(θ_9) için farklı yetenek kestirim ve madde seçim yöntemleri kullanıldığında elde edilen değerleri vermektedir. Buna göre, en düşük değerler EAP yetenek kestirim yöntemi ile birlikte ağırlıklandırılmış T-rule madde seçim yöntemiyle elde edilmiştir.

İkinci boyut için elde edilen grafik incelendiğinde, farklı yetenek kestirim yöntemlerinin kullanıldığı koşullarda EAP ve MAP ile elde edilen değerlerin birbirine oldukça yakın olduğu, en düşük standart hata değerine ise D-rule madde seçim yönteminin kullanıldığı koşulda ulaşıldığı sonucuna varılmıştır.

Üçüncü ve son grafikte, θ_3 elde edilen standart hata değerleri yer almaktadır. Bu değerler incelendiğinde; D-rule madde seçim ve MAP yetenek kestirim yöntemleriyle elde edilen standart hata değerinin, tüm madde seçim yöntemleri arasında en düşük değer olduğu görülmektedir. Tüm madde seçim yöntemlerinde MAP ile elde edilen değerlerin küçük farklarla da olsa EAP ile elde edilenlere göre daha düşük olmasına rağmen, ağırlıklandırılmış T-rule yönteminde bu durumun tersine MAP ile elde edilen standart hata değerinin EAP ile elde edilene göre oldukça fazla olduğu görülmektedir.

Yukarıda yorumlanmış olan değerler ışığında, standart hata değerleri açısından; genel yetenek düzeyinde D-rule ile elde edilen değer bir miktar yüksek olsa da diğer boyutlar için en iyi istatistiklerin bu madde seçim yöntemiyle elde edilmiş olması ve söz konusu genel yetenek düzeyi için elde edilen standart hata değerinin diğer yöntemlerle dramatik farklılıklar yaratmamış olması nedeniyle, D-rule madde seçim yönteminin tercih edilmesi gerektiği sonucuna varılmıştır. Yetenek kestirim yöntemi olarak ise, MAP'ın kullanılması uygun bulunmuştur.

d. Simülasyon sonuçlarına göre gerçek-zamanlı çok boyutlu BBT uygulaması için en uygun koşul hangisidir?

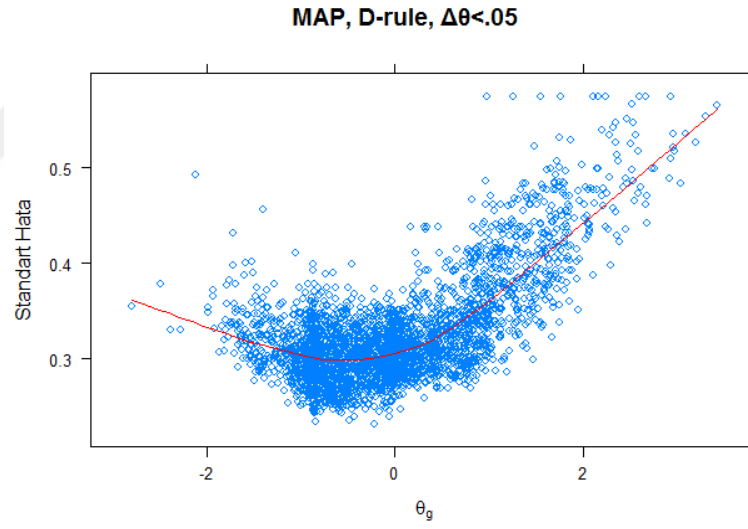
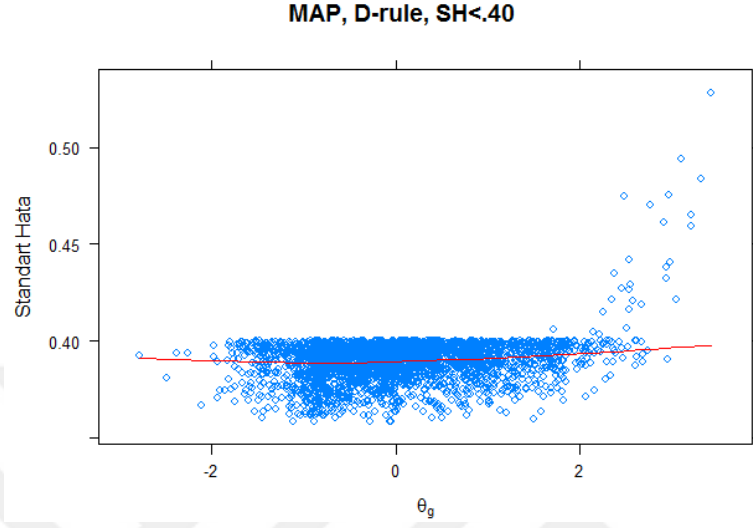
Araştırmanın bu bölümünde, her bir sonlandırma kuralı için belirlenen en iyi koşullar karşılaştırılmıştır. Bu amaçla öncelikle, her bir sonlandırma kuralı için en iyi koşullara ait istatistikler aşağıdaki tabloda verilmiş ve bu değerler yorumlanmıştır. Bir sonraki aşamada ise, her bir koşulda tüm boyutlar için R'de elde edilen grafiklere yer verilmiş ve bu grafikler yorumlanmıştır.

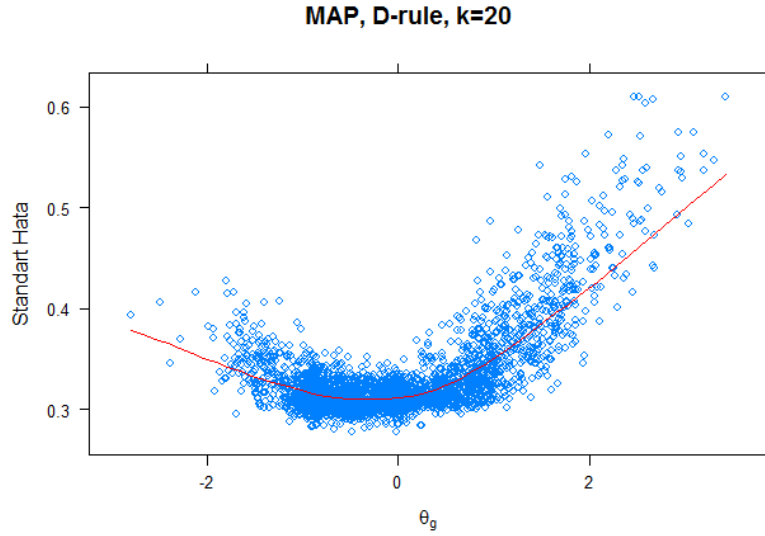
Tablo 4.4: Her Bir Sonlandırma Kuralı için En İyi İstatistikleri Veren Koşullara Ait İstatistikler

Sonlandırma Kuralı	Yetenek Kestirim Yöntemi	Madde Seçim Yöntemi	r			Yanlılık			RMSD			SH			k
			$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(g)}$	$\theta_{(2)}$	$\theta_{(3)}$	
Standart Hata	MAP	D-rule	.90	.60	.62	-.0326	.0370	.0721	.39	.56	.54	.39	.71	.65	13.4
θ Değişimi	MAP	D-rule	.94	.70	.70	-.0069	.0639	.0768	.32	.52	.49	.33	.63	.61	21.4
Sabit Madde	MAP	D-rule	.93	.69	.69	-.0093	.0607	.0751	.33	.52	.50	.33	.64	.61	-

Tablo 4.4'te, her bir sonlandırma kuralı için en iyi istatistikleri veren koşullara ait değerler yer almaktadır. Bu değerler incelendiğinde; θ değişimine dayalı ve sabit maddeye dayalı sonlandırma kurallarının kullanıldığı koşullarda, her bir boyut için elde edilen korelasyon ve hata istatistiklerinin birbirine yakın olduğu görülmektedir. Ayrıca, θ değişimine dayalı sonlandırma kuralı ile yanıtlanan madde sayısının ortalaması, sabit maddeye dayalı ($k=20$) sonlandırmada kullanılan madde sayısı kriterine oldukça yakındır. Bununla birlikte, standart hataya dayalı sonlandırma kuralı ile elde edilen korelasyon ve istatistikler, diğer yöntemlerle elde edilen değerlere göre farklılık göstermektedir. Elde edilen korelasyonlar nispeten daha küçük, hata istatistikleri ise daha büyüktür. BBT uygulamalarında ölçme kesinliği ve test uzunluğunun ölçme işleminin etkililiğini artırdığı bilinmektedir (Segall, 2005). Hatta Weiss (2011)'e göre ise BBT uygulamalarında ölçme kesinliği, madde sayısının az olmasını gerektirir. Bir başka ifadeyle, BBT uygulamalarının doğası ve amacı gereği yanıtlanan madde sayısı başlı başına ölçme işleminin etkililiğini belirler. Yukarıdaki değerler incelendiğinde; diğer yöntemlerle yanıtlanan madde sayısının, standart hata ile elde edilen değerlere göre yaklaşık %50-55 daha fazla olduğu görülmektedir. Bu nedenle, her ne kadar korelasyon ve hata istatistiklerinde standart hata sonlandırma kuralı ile elde edilen değerler diğer yöntemlerle elde edilen değerlere göre bir miktar fark olsa da yanıtlanan madde sayısında yarattığı büyük farklılık nedeniyle elde edilen istatistiklere dayanarak standart hataya dayalı durdurma kuralının gerçek-zamanlı çok boyutlu BBT uygulamasında kullanılmasının daha uygun olacağı değerlendirilmiştir.

Gerçek-zamanlı uygulamada hangi koşulun kullanılacağına karar vermek için bir de, her bir boyutta elde edilen yanıtlanan madde sayısı – θ ve standart hata – θ grafikleri yorumlanmıştır. EK-4'te yer alan grafiklerden sadece genel yetenek için elde edilenler, aşağıda da sunulmuştur.





Şekil 4.6: θ_g için Elde Edilen Standart Hata – θ Grafikleri

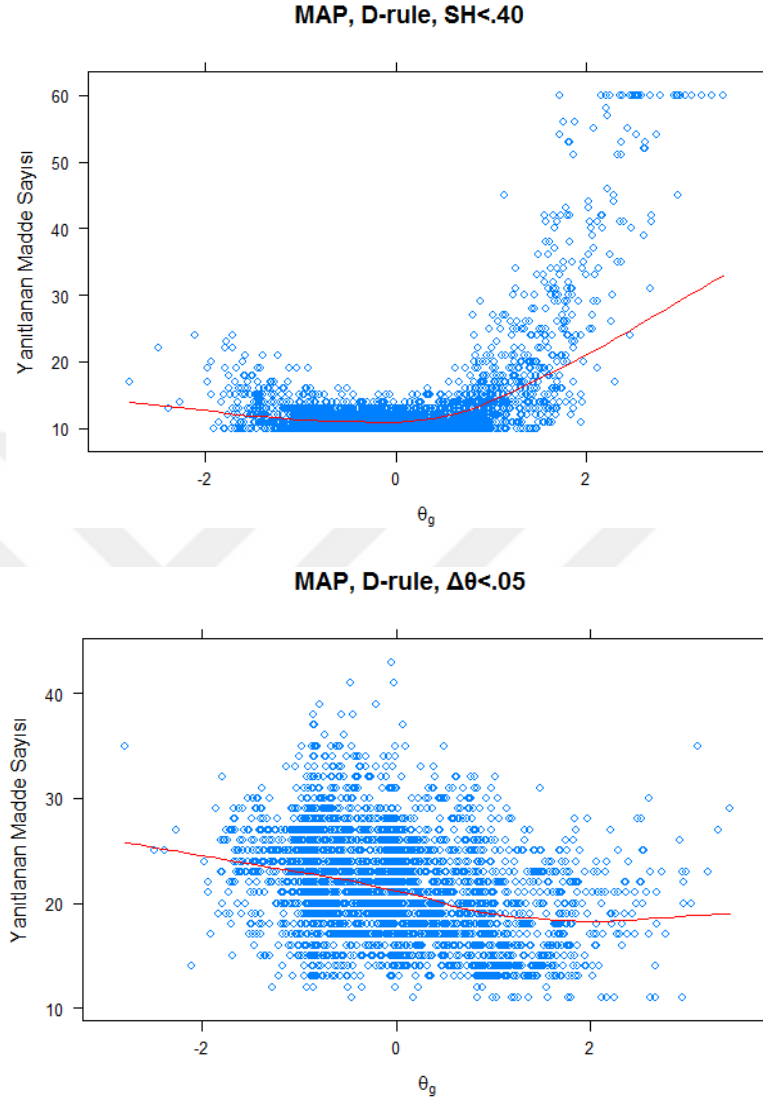
Şekil 4.6'da yer alan grafiklerden ilki, genel yetenek için standart hataya dayalı sonlandırma kuralı ile elde edilmiştir. Sonlandırma standart hatanın .4 değerinin altına düşmesiyle gerçekleştirildiğinden, yalnızca maksimum (60) madde sayısının yanıtlanmış olmasına rağmen bu standart hata değerinin altına düşmeyen kestirimlerin .4'ün üzerinde olduğu görülmektedir. Söz konusu yüksek standart hata değerlerinin, gerçek yetenek düzeyi yüksek olan bireylerde görüldüğü dikkat çekmektedir.

İkinci ve üçüncü grafikler θ değişimine dayalı ve sabit madde sayısına dayalı sonlandırma kuralları için elde edilmiş olup, bu grafiklerin birbirine benzer olduğu görülmektedir. Öyle ki, iki grafikte de yetenek ölçeğinin orta kısmında kestirilen standart hata değerleri oldukça küçük bir ranja sahiptir. Ancak, bireylerin gerçek yetenek düzeyi arttıkça standart hata değerinin de arttığı, elde edilen değerlerin .6'ya kadar yükseldiği görülmektedir.

EK-4.5'te görülen ve gramer becerisi için elde edilen grafikler incelendiğinde; standart hataya dayalı durdurma kuralının kullanıldığı koşulda ortalama hata değerleri .7'nin üzerinde seyrederken, diğer iki sonlandırma kuralına göre .65'in altında seyrettiği görülmektedir.

Standart hata – θ_3 değerleri için elde edilen ve EK-4.6'da yer alan grafikler incelendiğinde, kelime bilgisi becerisi için elde edilen grafiklerin, üç koşul için de benzer olduğu görülmektedir. Düşük yetenek düzeylerinde kestirilen ortalama standart hata değerleri düşükken, orta yetenek düzeyine yaklaşıldıkça bu değer nispeten arttığı ve yetenek skalasının sonuna kadar sabit kaldığı söylenebilir.

Sonraki şekillerde, değişken madde sayısına dayalı sonlandırma kurallarının kullanıldığı durumlarda, her bir boyut için yanıtlanan madde sayısının yetenek düzeyine göre değişimini gösteren grafikler yer almaktadır.



Şekil 4.7: θ_g İçin Elde Edilen Yanıtlanan Madde Sayısı – θ Grafikleri

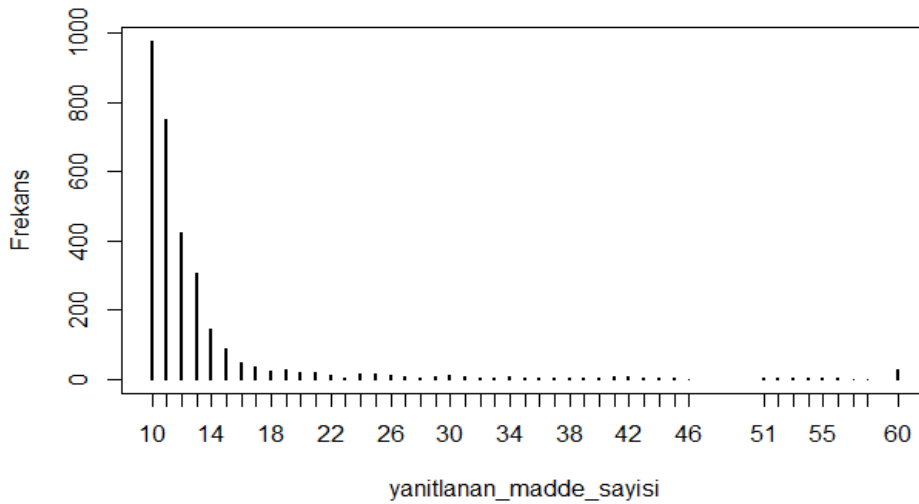
Şekil 4.7’de yer alan grafikler genel yetenek için elde edilmiştir. Standart hataya dayalı durdurma koşulunun kullanıldığı durumda; yetenek skalasının büyük bölümünde yanıtlanan ortalama madde sayısının 10’a yakın olduğu, yetenek düzeyinin 2’ye yaklaşmasıyla bu değerin arttığı görülmektedir. Özellikle çok yüksek yetenek düzeyindeki bireylerin, sonlandırma kuralının yanında belirlenen yanıtlanacak maksimum madde sayısına ulaştıkları görülmektedir. Bu bireylerin, standart hata grafiğinde .4’ün üzerinde oldukları görülen bireyler oldukları söylenebilir. θ değişimine dayalı sonlandırma kuralının kullanıldığı koşulda ise, tüm

yetenek skalasında yanıtlanan ortalama madde sayısının yüksek ve geniş bir ranja sahip olduğu dikkat çekmektedir.

Gramer becerisi için elde edilen, EK-4.2’de görülen yanıtlanan madde sayısı – θ grafikleri incelendiğinde, θ değişimine dayalı sonlandırma kuralının kullanıldığı koşulda genel yetenek ve gramer için elde edilen grafiklerin benzer olduğu söylenebilir. Standart hata dayalı durdurma kuralının kullanıldığı koşulda ise, tüm yetenek skalasında yanıtlanan ortalama madde sayısı 10’a oldukça yakın olmakla birlikte özellikle orta yetenek düzeyi için ranjin daha geniş olduğu görülmektedir. Dikkat çeken durum ise, genel yetenek için yetenek skalasının en yüksek kısmında bulunan bireylerin 60 madde yanıtlarken, burada 60 madde yanıtlayan bireylerin gramer becerisinin ± 1 aralığında olmasıdır.

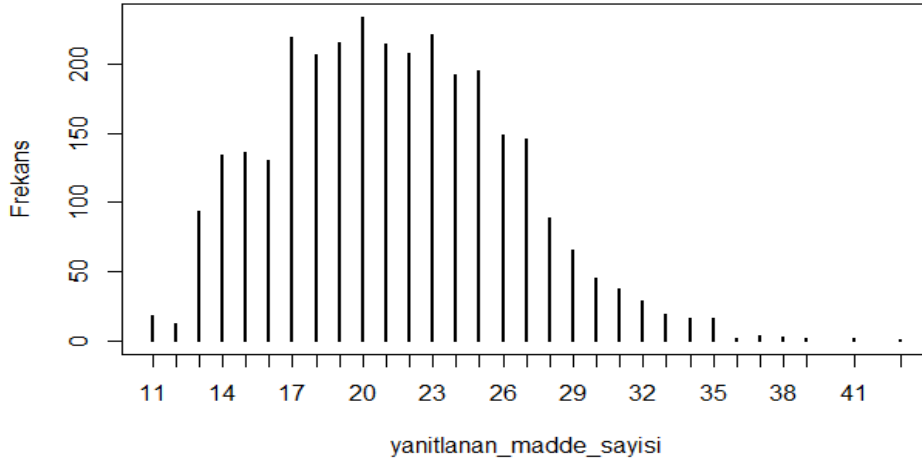
EK-4.3’te yer alan grafikler incelendiğinde; kelime bilgisi becerisi için elde edilen yanıtlanan madde sayısı – θ grafiklerinin, gramer becerisi için elde edilen grafiklerle oldukça benzer oldukları dikkat çekmektedir.

Simülasyon çalışmasında 3057 birey için yanıtlanan madde sayısının frekans değerlerine ilişkin grafikler aşağıda görülmektedir.



Şekil 4.8: Standart Hata Durdurma Kuralı, MAP ve D-rule’nin Kullanıldığı Çok Boyutlu BBT Koşulu için Yanıtlanan Madde Sayılarının Frekans Değerleri

Standart hataya dayalı sonlandırma yönteminin kullanıldığı koşullarda yanıtlanan madde sayılarının frekans değerleri incelendiğinde; bireylerin yaklaşık %30’unun, testi sonlandırmak için belirlenen en az madde sayısı olan 10 madde yanıtladıkları dikkat çekmektedir. Bu grafikten yola çıkarak, bireylerin yaklaşık %85’inin yanıtladıkları madde sayısının 10-15 arası olduğu söylenebilir.



Şekil 4.9: θ Değişimine Dayalı Durdurma Kuralı, MAP ve D-rule'nin Kullanıldığı Koşul için Yanıtlanan Madde Sayılarının Frekans Değerleri

Yukarıdaki grafiğe göre, θ değişimine dayalı durdurma kuralının kullanıldığı koşulda yanıtlanan madde sayısının bireylerin büyük bölümünde 17 ile 25 arasında değiştiği görülmektedir. Standart hataya dayalı durdurma kuralına ait koşuldan farkı ise bu koşulda yanıtlanan madde sayısı ranjının daha dar olmasıdır. Bu grafiklerden yola çıkarak, yanıtlanan madde sayısı açısından standart hataya dayalı durdurma koşulunun gerçek-zamanlı uygulamada kullanımının diğer yöntemlere göre daha verimli olacağı söylenebilir.

Araştırmanın bu bölümünde, her bir sonlandırma kuralı için simülasyonlar sonucu belirlenen en iyi koşullar için elde edilen istatistikler ve grafikler karşılaştırılmıştır. Elde edilen sonuçlara göre, gerçek-zamanlı çok boyutlu BBT uygulamasında yetenek kestirim yöntemi olarak MAP, madde seçim yöntemi olarak D-rule ve sonlandırma kuralı olarak da standart hatanın (.4) kullanılmasına karar verilmiştir.

4.5. Alt Problem 5: Gerçek-zamanlı çok boyutlu BBT uygulaması sonuçları nasıldır?

Bu bölümde, gerçek-zamanlı uygulama sonucu elde edilen sonuçlara yer verilmiştir. Sonuçlar, aşağıda görülen üç ayrı alt problemde ele alınmıştır.

a. Gerçek-zamanlı çok boyutlu BBT uygulaması sonucunda bireylerin yanıtladıkları madde sayılarının dağılımı nasıldır?

Gerçek-zamanlı uygulama, A kurunda yer alan toplam 99 öğrenciye yapılmıştır. Uygulamanın sonlandırılması için, genel yetenek kestiriminin standart hata değerinin .40'ın altına düşmesinin yanı sıra her bireyin en az 10 madde yanıtlaması sağlanmıştır. Bunun yanında yanıtlanacak maksimum madde sayısı olarak 60

belirlenmiş olmakla birlikte; gerçek-zamanlı uygulamaya katılan öğrencilerin tamamının yetenek kestiminin standart hatası, 60 madde yanıtlamaya gerek kalmadan .40'ın altına düşerek test sonlanmıştır. Aşağıda, yanıtlanan madde sayılarının frekanslarını gösteren tablo yer almaktadır.

Tablo 4.5: Katılımcıların Yanıtladıkları Madde Sayılarının Frekans Dağılımları

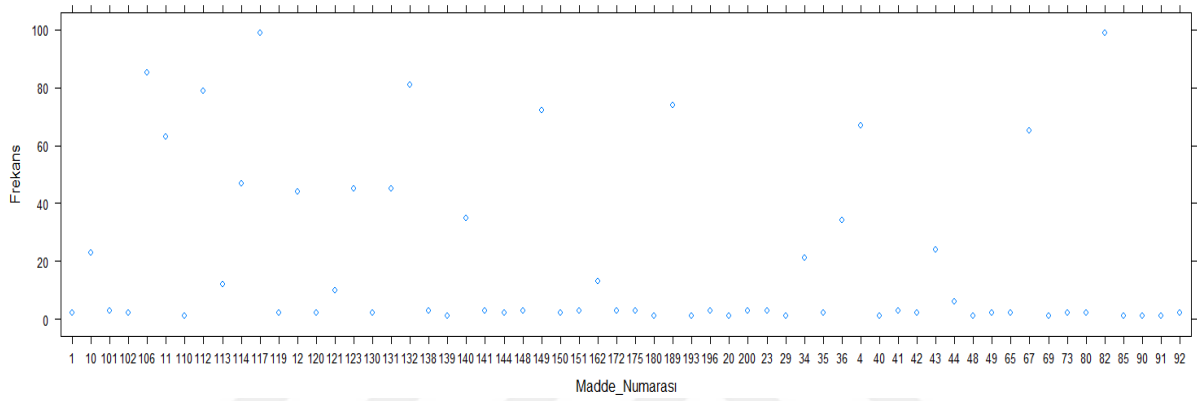
Yanıtlanan Madde Sayısı	Frekans	% (Yüzde)
10	25	25,3
11	24	24,2
12	25	25,3
13	12	12,1
14	4	4,0
15	5	5,1
16	1	1,0
22	1	1,0
32	1	1,0
45	1	1,0
Toplam	99	100,0

Yanıtlanan madde sayılarının frekanslarına bakıldığında; 10-12 madde yanıtlayanlarının sayısının 74 olduğu görülmektedir. 13 madde yanıtlayan kişi sayısı 12 iken, 14 madde yanıtlayanların sayısının 4, 15 madde yanıtlayanlarının sayısının ise 5 olduğu görülmektedir. 15 maddeden fazla yanıtlayanların sayısı ise sadece 4'tür. Söz konusu 4 kişi sırasıyla 16, 22, 32 ve 45 madde yanıtlamışlardır.

Elde edilen sonuçlara göre, gerçek-zamanlı uygulamaya katılan 99 öğrencinin yanıtladıkları madde sayısının ortalaması 12.3'tür. Bu değer, gerçek-zamanlı uygulama ile aynı koşulun (MAP, D-rule, SH<.4) kullanıldığı simülasyon uygulaması sonucu elde edilen ortalama madde sayısı olan 13.4'e oldukça yakın ve bir miktar daha düşük olduğu görülmüştür. Yanıtlanan madde sayısı 10 ile 45 arasında değişmektedir. Gerçek-zamanlı uygulama ile ölçülen gramer ve kelime bilgisi becerileri için kağıt-kalem formatında düzenlenen YTS'de bireylerin yanıtladıkları madde sayısı 50'dir. Buna göre; gerçek-zamanlı çok boyutlu BBT uygulamasının, kağıt-kalem formatındaki uygulama göre bireylerin ortalama %75.4 daha az madde yanıtladıkları sonucuna ulaşılmıştır.

b. Gerçek-zamanlı çok boyutlu BBT uygulaması sonucunda havuzda yer alan maddelerin kullanım oranları nedir?

İngilizce hazırlık eğitimi gören 99 öğrenciye yapılan gerçek-zamanlı çok boyutlu BBT uygulaması sonucunda, madde havuzunda yer alan 200 maddeden 60'ının kullanıldığı, 140 tanesinin ise hiç bir uygulamada yer almadığı görülmüştür. Bir başka ifadeyle, 99 kişiye yapılan gerçek-zamanlı çok boyutlu BBT uygulamasında madde havuzunun %30'u kullanılmıştır. Kullanılan 60 maddeden 37'sinin ise kullanım sayılarının 5'in altında olduğu dikkat çekmektedir.



Şekil 4.10: Madde Havuzunda Gerçek-Zamanlı Uygulamada Yer Alan Maddelerin Kullanım Miktarları

Yukarıda yer alan grafikte, madde havuzunda yer alıp gerçek-zamanlı uygulamada en az bir kez kullanılan 60 maddenin kullanılma sıklıkları görülmektedir. Söz konusu maddelerden her uygulamanın başlangıcında kullanılan 117. madde ile birlikte 82. maddenin de uygulamaların tamamında yer alması dikkat çekmektedir. Bu iki maddenin dışında en az 60 uygulamada kullanılan 8 madde yer almaktadır. 20 ve üzerinde birey tarafından yanıtlanan maddelerin sayısı ise 19'dur.

c. Bireylerin gerçek-zamanlı çok boyutlu BBT uygulamasıyla elde edilen genel yetenek yetenek düzeyleri ile kağıt-kalem testi sonucu elde edilen toplam puanları arasında anlamlı bir ilişki var mıdır?

Gerçek-zamanlı çok boyutlu BBT uygulaması, Anadolu Üniversitesi YDYO'da İngilizce hazırlık eğitimi gören toplam 99 kişiye yapılmıştır. Söz konusu 99 kişi, temel olarak iki gruptan oluşmaktadır. İlk grup, 2016-2017 öğretim yılı birinci döneminde A kuruna yükselmiş 46 öğrenciden oluşmaktadır. Bu nedenle bu grup, dönem arasında sadece A kurunda yer alan sınırlı sayıdaki öğrenci için uygulanan YTS'ye girme hakkına sahip olmuştur. Diğer grup ise A kuruna ikinci dönem itibariyle, bu

alıřma kapsamında yrtlen gerek-zamanlı uygulamanın yapıldığı Mart ayının bařında gemiř 53 ğrenciden oluřmaktadır. Her bir gruptaki ğrenciler,  farklı sınıfta ğrenim grmektedirler. Gerek-zamanlı uygulamanın yapılmasının ardından elde edilen yetenek kestirimlerinin kağıt-kalem testi sonularıyla karřılařtırılmasında kullanılan test, ğrencilerin ara dnemde girdikleri YTS'dir. Bu sınava sadece ilk dnem itibariyle A kurunda olan ğrenciler girebildiğinden, kestirilen yetenek ve kağıt-kalem testi korelasyon alıřması, alıřma kapsamındaki iki gruptan ilkiyle yapılmıřtır. Her ne kadar bu grupta 46 ğrenci bulunsa da ğrencilerin bir kısmı sınava girmediğinden, kimisi ise sınav sorularını yanıtlamadan ıktığından, uygulama kalan 32 ğrencinin verileriyle yrtlmřtr. Sz konusu 32 ğrencinin kağıt-kalem sınav sonuları, gerek-zamanlı ok boyutlu BBT uygulamasında kestirilen gramer ve kelime bilgisi becerilerine ynelik maddelere verdikleri yanıtlardan oluřmaktadır.

32 ğrenciyle yrtlen alıřmada bireylerin; gerek-zamanlı uygulama sonucundaki genel yetenek kestirimleriyle, kağıt-kalem formatındaki sınavın gramer ve kelime bilgisine ynelik maddelerinden elde ettikleri toplam puan arasındaki korelasyon hesaplanmıřtır. Elde edilen sonulara gre, kağıt-kalem formatında uygulanan sınavın gramer ve kelime bilgisi blmlerinin toplam puanlarıyla gerek-zamanlı ok boyutlu BBT uygulamasıyla kestirilen genel yetenek puanları arasında .77'lik anlamlı bir korelasyon bulunmuřtur.

5. TARTIŞMA

Bu çalışmada; üniversite İngilizce hazırlık sınıfı öğrencilerinin eğitimlerini bitirmelerinin ardından, lisans eğitimleri öncesi vermeleri gereken yeterlilik sınavının verileri kullanılarak gramer ve kelime bilgisi becerileriyle birlikte bir de genel yetenek ölçeği çok boyutlu bir BBT geliştirilmiştir. Bu amaçla, öncelikli olarak Anadolu Üniversitesi YDYO'da geçmiş yıllarda uygulanan YTS'lerdeki maddeler ve bu maddelere ait veriler kullanılarak bir madde havuzu oluşturulmuştur. Madde havuzunun oluşturulmasının ardından, gerçek-zamanlı çok boyutlu BBT uygulamasında kullanılacak algoritmayı belirlemek amacıyla, eldeki eksik veri matrisi de göz önüne alınarak bir *hibrit* simülasyon uygulaması yapılmıştır. Bu simülasyon uygulamasında farklı yetenek kestirim yöntemlerinin (EAP ve MAP), farklı madde seçim yöntemlerinin (D-rule, KL, W-rule, T-rule, ağırlıklandırılmış W-rule ve ağırlıklandırılmış T-rule) ve farklı sonlandırma kurallarının (standart hata, θ değişimi ve sabit madde) kullanıldığı toplam 36 farklı koşul yaratılmış; bu koşullarda her bir boyut için gerçek ve kestirilen θ değerleri arasındaki korelasyon ile birlikte yanlışlık, RMSD ve standart hata değerleri elde edilmiştir. Korelasyon değerleri ve hata istatistikleri ile birlikte yanıtlanan madde sayısı da BBT uygulamalarında ölçme kesinliğinin önemli bir göstergesi olduğundan, değişken madde sayısına dayalı sonlandırma kuralının kullanıldığı koşullarda yanıtlanan madde sayısı da raporlanmıştır. Bunların yanında, her bir çok boyutlu BBT koşulunun doğasına uygun olarak; yanıtlanan madde sayısı – θ grafikleri, standart hata – θ grafikleri ve yanıtlanan madde sayılarının frekansını gösteren grafikler elde edilmiştir. Tüm bu grafikler EK-4'te yer almaktadır. Simülasyonlara dayalı olarak en uygun çok boyutlu BBT algoritmasının belirlenmesinin ardından, bu algoritma kullanılarak gerçek-zamanlı çok boyutlu BBT uygulaması yapılmıştır. Bu uygulamanın yapıldığı öğrencilerden 32'sinin kağıt-kalem testiyle elde edilen toplam puanları ile gerçek-zamanlı uygulama sonucu elde edilen genel yetenek düzeyleri arasındaki korelasyon elde edilmiştir. Bunun yanında, uygulamanın yapıldığı toplam 99 kişinin sonuçlarına dayanarak havuzda yer alan maddelerin kullanım sıklıkları ve her bir bireyin yanıtladığı toplam madde sayısı da raporlanmıştır. Bu kısımda, elde edilen sonuçlar simülasyon ve gerçek-zamanlı uygulama olmak üzere iki başlık altında yorumlanmıştır.

5.1 Simülasyon Uygulaması Sonucu Elde Edilen Bulgulara İlişkin Yorumlar

Simülasyon sonuçlarına göre, araştırma kapsamında kullanılan üç sonlandırma kuralı için de en uygun koşulların D-rule madde seçim ve MAP yetenek kestirim yöntemleriyle elde edildiği görülmüştür. D-rule madde seçim yöntemi; genel yeteneğin kestirilmesinde diğer yöntemlerle çok yakın değerler vermekle birlikte, grup faktörleri açısından gerek korelasyon gerekse hata istatistikleri açısından çok daha iyi değerler vermiştir. Bu sonuçlar, Seo & Weiss (2015) tarafından yürütülen simülasyon çalışmasıyla benzerlik göstermektedir. Söz konusu çalışmada araştırmacılar, grup faktörleri için yapılan kestirimlerin önemsendiği durumlarda D-rule madde seçim ve MAP yetenek kestirim yönteminin kullanılmasını önermişlerdir.

Araştırmadan elde edilen sonuçlara göre, yanlılık değerleri karşılaştırıldığında genel olarak KL madde seçim yönteminin 0'a en yakın sonuçları verdiği görülmektedir. KL madde seçim yöntemiyle; gerek korelasyon değerleri, gerekse standart hata ve RMSD değerleri açısından hiçbir koşulda en iyi istatistiklerin elde edilmemiş olmasına rağmen, genel olarak en yansız sonuçların elde edilmiş olması dikkat çekmektedir. Literatürde, benzer bir durumla karşılaşıldığına dair bir bulguya rastlanmamıştır.

Bu çalışma kapsamında genel yetenek için elde edilen gerçek ve kestirilen yetenek parametreleri arasındaki korelasyon değerleri tüm koşullarda oldukça yüksekken, grup faktörleri için elde edilen korelasyon değerleri daha düşüktür. Bu durumun nedenlerinden ilkinin, bifaktör modelin doğası olduğu düşünülebilir. Öyle ki çok boyutlu bir modelin bifaktör yapıya uyum göstermesi için yapının genel bir yetenek kestiriyor olmasının yanı sıra, genel yeteneğe ait faktör yüklerinin grup faktörlerden daha yüksek olması gerektiği (Reise, Morizot & Hays, 2007) bilinmektedir. Bu durumun, bir yapının bifaktör modele uyum gösterirken grup faktörleri için verilen bilginin azalmasını beraberinde getireceği söylenebilir. Bunun yanında, Seo (2012) tarafından yürütülen çalışmada da bu çalışmaya benzer olarak genel yetenek için elde edilen korelasyon değerleri, grup faktörleri için elde edilen değerlere göre daha yüksek bulunmuştur.

Genel yetenek için elde edilen korelasyon değerleri yüksekken, grup faktörleri için elde edilen değerlerin düşük olmasının bir diğer nedeninin yanıtlanan madde sayısı olduğu düşünülmektedir. Weiss & Gibbons (2007) çalışmalarında; BBT'nin etkililiğini

artırma adına her bir grup faktörü için 20 ile 50 arasında madde kullanılması gerektiğini belirtmişlerdir. Buna bağlı olarak çok boyutlu bifaktör BBT uygulamalarına yönelik literatürdeki diğer iki çalışma olan Seo (2012) ve Seo & Weiss (2015) tarafından yürütülen çalışmalarda da her bir grup faktör için 20 madde kullanılmıştır. Ancak söz konusu çalışmalar, *monte carlo* simülasyon çalışmalarıdır. Bu çalışma kapsamında ise sabit madde sayısına dayalı sonlandırma koşulunda her bir faktör için 10'ar madde kullanılmış, diğer koşulların büyük bölümünde ise her bir grup faktörü için yanıtlanan madde sayısı 5'e kadar düşmüştür. Grup faktörler için elde edilen korelasyonların düşük ve hata istatistiklerinin yüksek olması bu duruma bağlı olabilir. Araştırmacı, bu çalışma sonrasında kağıt-kalem uygulamasında 50 madde kullanılan bir test için gerçek-zamanlı uygulama geliştirmeyi amaçlamıştır. BBT çalışmalarında yanıtlanan madde sayısının düşük olmasının ölçme kesinliği ile direkt ilişkili olmasından dolayı, her boyut için yanıtlanacak madde sayısının en az 20 olarak belirlenmesinin gerçek-zamanlı uygulamanın etkililiğini düşüreceği öngörülmüştür.

Ağırlıklandırmanın kullanıldığı madde seçim yöntemlerine bakıldığında; değişken madde sayısına dayalı sonlandırma yöntemlerinde, W-rule madde seçim yönteminde ağırlıklandırmanın kullanılması durumunda genel yetenek için elde edilen korelasyon değerlerinde bir artış ve hata istatistiklerinde de önemli bir düşüş sağlandığı görülmektedir. Yanıtlanan madde sayısı ise ortalama %20-25 arasında bir azalma sağlanmıştır. T-rule madde seçim yönteminde ağırlıklandırmanın kullanıldığı durumlarda ise gerek korelasyon ve hata istatistikleri, gerekse yanıtlanan ortalama madde sayıları açısından meydana gelen düşüşün W-rule ile elde edilen değerlere göre oldukça düşük olduğu görülmektedir. Bunun yanında, EK-4'teki grafiklere bakıldığında da W-rule madde seçim yönteminde ağırlıklandırma yapılması durumunda yanıtlanan madde sayılarının frekans dağılımlarında kayda değer değişiklik olduğu, T-rule yönteminde ise benzer bir sonuçla karşılaşmadığı görülmektedir. Yanıtlanacak minimum madde sayısının azaltılması ile birlikte, T-rule madde seçim yönteminde ağırlıklandırma kullanımının etkisinin artacağı düşünülebilir.

Simülasyon uygulaması sonrasında, her bir koşulun doğasına uygun olarak yanıtlanan madde sayısı – θ ve standart hata – θ grafikleri elde edilmiştir. EK-4'te yer alan bu grafikler incelendiğinde; genel yetenek için elde edilen standart hata –

θ grafiklerinin sonlandırma kuralına göre değiştiği görülmektedir. Buna göre, standart hataya dayalı durdurma kuralının uygulandığı koşullarda θ değeri 2'nin üzerine çıktığında, standart hata kestirimlerinin bir kısmının sonlandırma kuralı olan .4'ün altına düşmediği görülmektedir. θ değişimine dayalı sonlandırma kuralında ise, -1 ile +.5 yetenek aralığında standart hataların çok düşük olduğu, θ değerinin 1'in üzerine çıkmasıyla yetenek skalasının sonuna doğru ilerledikçe standart hata değerlerinin .5'in üzerine kadar çıktığı görülmektedir. Sabit madde sayısına dayalı sonlandırma kuralının kullanıldığı koşullarda da elde edilen grafikler benzerdir. Bu sonuçlardan yola çıkarak, genel yetenek için yüksek θ seviyelerinde madde havuzunun yeterli bilgi vermediği söylenebilir.

Yanıtlanan madde sayısı – θ grafiklerinden standart hataya dayalı durdurma kuralının kullanıldığı koşullar için elde edilen grafiklerde; genel yetenek için, θ değerinin 2'nin üzerine çıkmasıyla yanıtlanan madde sayısının 60'a ulaştığı görülmektedir. Söz konusu durumların, standart hata – θ grafiklerinde .4'ün üzerinde kestirilen standart hata değerlerine ait olduğu düşünülmektedir. Dikkat çeken bir diğer husus da θ değişimine dayalı sonlandırma kuralında, standart hataya dayalı sonlandırma kuralından farklı olarak yüksek θ düzeylerinde yanıtlanan madde sayılarının orta yetenek düzeyine göre daha az olmasıdır. Tam tersine, orta düzeydeki yetenek kestirimlerinde yanıtlanan madde sayısı çok daha fazladır. Söz konusu sonlandırma yönteminde orta yetenek düzeyinde standart hatanın çok düşük olmasının nedeni olarak bu durum görülebilir.

Gramer ve kelime becerisi içinse, yanıtlanan madde sayılarının ± 1 yetenek aralığında 60 rakamına ulaştığı dikkat çekmektedir. Buna bağlı olarak, genel yetenek seviyesi çok yüksek olan bireylerin θ_2 ve θ_3 için kestirilen yetenek düzeylerinin yetenek skalasının orta noktasına yakın olduğu sonucuna varılmıştır.

5.2 Gerçek-Zamanlı Çok Boyutlu Uygulaması Sonucu Elde Edilen Bulgulara İlişkin Yorumlar

Gerçek-zamanlı çok boyutlu BBT uygulaması sonucunda, 200 maddeden oluşan havuzun %30'luk kısmının kullanıldığı görülmüştür. Bu değere bakılarak, havuzun kullanım oranının düşük olduğu söylenebilir. Ancak yapılan çalışmalar, BBT uygulamalarının %50'sinde, madde havuzunun ancak %14'ünün kullanıldığına yöneliktir (Wainer, 2000). Simülasyon sonuçlarına dayanarak yüksek yetenek

düzeyinde, orta ve düşük yetenek düzeyine göre havuzun daha düşük bilgi verdiği göz önüne alındığında, yanıtlayıcıların yetenek düzeyinin orta ve düşük seviyede olmasının havuzun küçük bir kısmının kullanılmasına neden olduğu söylenebilir. Bunun yanında madde havuzunun kullanım oranının düşük olmasının bir başka nedeni de, yanıtlanan ortalama madde sayısının 12.3 gibi oldukça düşük bir değer olması ve çalışma kapsamında madde teşhir oranını kontrol edecek herhangi bir yöntem kullanılmamış olmasıdır. Genel olarak aynı yetenek düzeyinde olmaları ve madde teşhir oranının kontrol edilmemesi nedeniyle bireylerin yanıtladıkları sınırlı sayıdaki maddenin çoğu zaman ortak olduğu düşünülebilir. Yanıtlanacak minimum madde sayısının artırılması ve madde teşhir oranının kontrol edilmesine yönelik bir yöntemin kullanılması durumunda, madde havuzunda kullanılan maddelerin sayısının artacağı tahmin edilmektedir.

Gerçek-zamanlı uygulamada, madde havuzundaki maddelerden 60'ı kullanılmıştır. Bu 60 maddenin ise sadece 23 tanesi yapılan 99 uygulamadan 5 ve daha fazlasında yer almıştır. Daha önce de üzerinde durulduğu gibi, yanıtlanan madde sayısının ortalamasının oldukça düşük olması ve bireylerin yaklaşık %75'inin ortalama madde sayısından daha az madde yanıtlamasının bu durumun ortaya çıkmasında etkili olduğu söylenebilir.

Bireylerin, gerçek-zamanlı uygulama sonrası elde edilen genel yetenek düzeyleri kağıt-kalem formatında uygulanan sınavdan elde ettikleri toplam puanlar arasında .77'lik bir korelasyon elde edilmiştir. Bu değer, simülasyonlarda elde edilen .90'lık korelasyon değerine göre küçük olmasının bir kaç nedeni olabileceği düşünülmektedir. Bunlardan ilki, simülasyon uygulamasında kullanılan veri setinin 3057 bireyden oluşurken, gerçek-zamanlı uygulama için 32 kişiden elde edilen verilerin kullanılmasıdır. Korelasyon, kestirimin yapıldığı grubun büyüklüğünden etkilenen bir değerdir. Bu nedenle, 32 kişilik gruptan elde edilen değerlerin bir miktar grubun küçük olmasından etkilendiği söylenebilir. İkinci nedenin ise, kağıt-kalem testinin öğrencilerin Şubat ayında girdikleri YTS'den elde edilen puanlar olmasından kaynaklandığı düşünülmektedir.

6. SONUÇ ve ÖNERİLER

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

6.1. Sonuçlar

Araştırma kapsamında elde edilen tüm sonuçlar aşağıda maddelenmiştir.

- Değişken madde sayısına dayalı sonlandırma kuralının kullanıldığı simülasyon uygulamalarında, MAP yetenek kestirim yönteminin kullanıldığı tüm koşullarda yanıtlanan ortalama madde sayısı EAP yetenek kestirim yönteminin kullanıldığı koşullara göre daha azdır.
- Simülasyon uygulaması sonucunda yanıtlanan madde sayısı ortalaması 11.3 – 24.1 arasında değişmektedir. Kağıt-kalem uygulaması sonucu gramer ve kelime bilgisi becerilerini ölçmeye yönelik 50 madde yanıtladığı göz önüne alındığında, yanıtlanan madde sayısında kullanılan çok boyutlu BBT koşuluna göre %52 ile %78.4 arasında azalma sağlandığı sonucuna ulaşılmıştır.
- Standart hataya dayalı durdurma kuralının kullanılması durumunda ağırlıklandırılmış madde seçim yöntemlerinin kullanıldığı koşullarda; W-rule madde seçim yönteminde ağırlıklandırmanın kullanılması yanıtlanan ortalama madde sayısını ağırlıklandırma yapılmayan koşula göre azaltırken, T-rule yönteminde ağırlıklandırmanın kullanıldığı durumda yanıtlanan ortalama madde sayısında meydana gelen azalma ağırlıklandırma yapılmayan koşula göre çok daha düşüktür.
- Simülasyon sonucunda 36 koşulun tamamında, genel yetenek için gerçek ve kestirilen θ değerleri arasındaki korelasyon oldukça yüksek ve birbirine yakındır (.90 - .94 ranjında). Bunun yanında, diğer boyutlar için kestirilen korelasyon değerlerinin oldukça düşük olduğu görülmektedir. Bu duruma; madde havuzunun genel yetenek düzeyinde yüksek bilgi veren maddelerden oluşurken, diğer boyutlar içinse havuzda yer alan maddelerin verdiği bilginin düşük olması neden olabilir. Bu durumun, bifaktör modelin doğasına uygun olduğu düşünülebilir. Öyle ki; çok boyutlu bir yapının bifaktör modele uyum

göstermesi için, genel yetenek düzeyindeki faktör yüklerinin grup faktörlerine ait yüklerden daha büyük olması gerekir (Reise, Horizot & Hays, 2007).

- Simülasyonlarda kullanılan tüm madde seçim yöntemlerinde genel yetenek için elde edilen korelasyon büyük ve birbirine yakinken, D-rule haricindeki tüm madde seçim yöntemlerinde grup faktörleri için elde edilen korelasyon katsayılarının oldukça düşük olduğu görülmektedir..
- Madde seçiminde ağırlıklandırmanın kullanılabildiği yöntemler için, ağırlıklandırmanın kullanıldığı durumlarda genel olarak grup faktörleri için elde edilen korelasyon düşmekte ve hata istatistikleri artmaktadır. Bu duruma tek istisna, madde sayısının sabit ($k = 20$) olduğu durdurma kuralında W-rule madde seçim yönteminde görülmüştür. Söz konusu durumda, ağırlıklandırma kullanılması grup faktörlerinde elde edilen korelasyon ve hata istatistiklerinde neredeyse hiç bir değişime neden olmamıştır.
- Değişken madde sayısına bağlı sonlandırma kurallarının kullanıldığı simülasyon uygulamalarında standart hataya bağlı durdurma kuralında, D-rule madde seçim yöntemi ile yanıtlanan ortalama madde sayısı diğer yöntemlerle elde edilen değerlere yakın ve kimilerinden daha azken, θ değişimine bağlı sonlandırma kuralında D-rule madde seçim yönteminin kullanıldığı durumlarda yanıtlanan ortalama madde sayısı diğer yöntemlere göre yaklaşık %20 - 40 arasında daha fazladır.
- Simülasyon uygulaması sonucu tüm koşullarda, D-rule madde seçim yönteminin kullanıldığı durumlarda iki grup faktörü için elde edilen korelasyonlar ve hata istatistikleri birbirine oldukça yakinken, diğer tüm madde seçim koşullarının kullanıldığı durumlarda grup faktörleri için elde edilen istatistikler arasındaki farklar oldukça büyüktür. Söz konusu koşullarda θ_2 için elde edilen korelasyon değerleri θ_3 'e göre oldukça düşük, hata istatistikleri ise oldukça yüksektir.
- Genel olarak simülasyon sonucu tüm koşullarda, elde edilen korelasyon değerleri ile yanlılık (bias) haricindeki hata istatistikleri arasında ters yönde yüksek bir ilişki olduğu görülmüştür. Öyle ki; bir koşul için elde edilen korelasyon değeri diğer koşullara göre en yüksek değerken, standart hata ve RMSD ise en düşük değerleri vermiştir. Yanlılık değeri ise, genel olarak hiç

bir koşulda iyi istatistik değeri vermeyen KL madde seçim yönteminin kullanıldığı koşullarda diğer madde seçim yöntemlerinin kullanıldığı koşullara göre 0'a en yakın değerleri vermiştir.

- Madde seçiminde ağırlıklandırmanın kullanılabildiği yöntemlerden T-rule yöntemi, ağırlıklandırma yapılması durumunda yapılmadığı duruma göre genel yetenek için daha yüksek korelasyon ve daha düşük hata değerleri vermediği gibi, yanıtlanan madde sayısında da dramatik bir azalmaya neden olmamaktadır. Bununla birlikte grup faktörleri için elde edilen korelasyonlardaki düşüş ve hata değerlerindeki artış dikkat çekicidir. Bu nedenle, T-rule madde seçim yöntemiyle ağırlıklandırmanın kullanılmasının kayda değer bir avantaj yaratmayacağı söylenebilir.
- Simülasyon uygulamasında, standart hataya dayalı durdurma kuralının kullanıldığı koşullarda bireylerin yanıtladığı madde sayılarının belirlenen minimum ve maksimum madde sayıları olan 10 ile 60 arasında değiştiği, bununla beraber bireylerin büyük çoğunluğunun 10 ya da 10'a çok yakın sayıda madde yanıtlayarak sınavı tamamladıkları görülmüştür. Bunun tersine, θ değişimine dayalı sonlandırma kuralının kullanıldığı koşullarda ise yanıtlanan madde sayısı ranjı daha dar ve dağılım daha homojendir. Yanıtlanan madde sayısı en fazla 47 olmuştur. En dar ranjda ise yanıtlanan madde sayısı 10 – 31 arasında değişmiş olup, dağılımın hafif sağa çarpık olduğu görülmüştür.
- Standart hataya dayalı durdurma kuralının kullanıldığı simülasyon koşullarında; yetenek eskirimi olarak MAP kullanılması durumunda, EAP kullanıldığı duruma göre 60 madde yanıtlayan birey sayısı daha azdır.
- Simülasyon uygulamalarında sonlandırma kuralının standart hataya bağlı olduğu durumda; sadece genel yetenek için belirlenen standart hatanın altına düşüldüğü durumlarda test sonlanmış, grup faktörleri için herhangi bir değer belirlenmemiştir. θ değişimine dayalı sonlandırma kuralında ise, tüm faktörler için son iki yetenek kestirimi arasındaki farkın .5'in altına düşmesiyle test sonlanmıştır. Buna rağmen, grup faktörlerinin korelasyon ve hata değerleri göz önüne alındığında standart hataya dayalı durdurma kuralı için sadece

genel yeteneğe dayalı sonlandırma kıstasının kullanılmasının, grup faktörleri için elde edilen sonuçlara kayda değer bir etkisi olmadığı görülmüştür.

- Simülasyon sonuçlarına göre genel yetenek düzeyi için; standart hataya bağlı durdurma kuralında üst yetenek düzeyinin yaklaşık 1'in üzerine çıktığı durumlarda yanıtlanan madde sayısının arttığı, yanıtlanacak maksimum madde sayısı olan 60'ın sadece yetenek düzeyinin 2'nin üzerine çıktığında görüldüğü dikkat çekmektedir. Sabit maddeye bağlı durdurma kuralının kullanıldığı durumda ise, yine aynı yetenek düzeyinde standart hata değerleri giderek yükselmiştir. Bu durumun nedeninin, madde havuzunun yüksek yetenek düzeyi için yeterli bilgi vermemesi olduğu söylenebilir.
- Gerçek- zamanlı çok boyutlu BBT uygulaması, İngilizce yetenek düzeyleri benzer olan 99 öğrenciye yapılmıştır. Elde edilen sonuçlara göre gerçek- zamanlı çok boyutlu BBT uygulaması, ortalama yaklaşık 9 dakika sürmektedir. Kağıt-kalem formatında uygulanan sınav için her bir soruya bir dakika ayrıldığı düşünüldüğünde, BBT uygulaması ile kağıt-kalem uygulamasına göre yaklaşık %80 daha az zaman harcandığı söylenebilir.
- Gerçek-zamanlı uygulama sonuçlarına göre, yanıtlanan ortalama madde sayısı 12.3'tür. Kağıt-kalem formatındaki uygulamada 50 maddenin kullanıldığı düşünüldüğünde, çok boyutlu BBT uygulaması ile ortalama %75.4 daha az madde yanıtlandığı görülmektedir. Ayrıca sınavın uygulandığı öğrencilerden yaklaşık %75'i, 10 ila 12 arasında madde yanıtlamıştır.
- Gerçek-zamanlı uygulama sonucunda madde havuzunun %30'luk kısmını oluşturan 60 maddenin kullanıldığı görülmüştür. Söz konusu maddelerin yaklaşık %60'ının ise, 5 uygulamadan daha azında kullanıldığı bulgusuna ulaşılmıştır.
- Araştırmanın son bölümünde; gerçek-zamanlı BBT uygulamasına katılan bireylerin, bu sınavdan elde ettikleri yetenek kestirim sonuçlarıyla kağıt-kalem formatındaki YTS'nin gramer ve kelime becerisi bölümünden elde ettikleri toplam puan arasındaki korelasyon hesaplanmıştır. Elde edilen sonuçlara göre, bireylerin gerçek-zamanlı çok boyutlu BBT ile kestirilen genel yetenekleri ile kağıt-kalem formatında elde edilen toplam puanları arasında .77'lik bir korelasyon olduğu görülmüştür.

6.2. Öneriler

6.2.1. Gelecekte Yapılacak Çok Boyutlu BBT Araştırmalarına Yönelik Öneriler

Bu bölümde, araştırma sonuçlarında yola çıkarak yapılacak yeni araştırmalara dair önerilere yer verilmiştir.

- Bu çalışma, İngilizce yeterliğini ölçmeye yönelik hazırlanmış dört boyutlu bir sınavın iki boyutu üzerine kurgulanmıştır. Sonraki gerçek-zamanlı uygulamalarda, tüm boyutların ölçülmesine ve genel yetenek kestiriminin yapılmasına imkan verecek bir araştırma yürütülmesine ihtiyaç duyulduğu düşünülmektedir.
- Bu çalışma kapsamında, tüm maddelerin altında yatan ortak bir varyans kaynağına dayalı olarak genel yetenek kestirimi yapılırken çok boyutluluğun göz ardı edilmediği bifaktör model kullanılmıştır. Bifaktör modelin veri setine uygululuğunun sınanmasının nedeni, Şahin & Akyıldız (2016) tarafından yürütülen çalışmadır. Söz konusu çalışma YTS'nin tüm boyutları kullanılarak yürütülmüş olup bu çalışmada, veri setlerine uyumu sınanan tüm diğer faktör analitik modeller (tek boyutlu, çok boyutlu, ikinci dereceden) arasında sistematik olarak en iyi uyum değerlerini bifaktör modelin verdiği görülmüştür. Bunun yanında bifaktör modellerin, yüksek derece (high-order)/hiyerarşik modellerin alternatifi olduğu söylenebilir (Seo & Weiss, 2015). Yüksek derece MTK modellerinin kullanıldığı literatürdeki tek çok boyutlu BBT çalışması ise Huang, Chen & Wang (2012) tarafından yapılmıştır. Çalışma kapsamındaki boyutlarla, yüksek derece MTK modellerinin kullanılarak yapılacak bir çok boyutlu BBT çalışmasında elde edilecek sonuçların, bu çalışmanın sonuçlarıyla karşılaştırılmasının literatüre önemli bir katkı sunacağı düşünülmektedir.
- İngilizce yeterlik sınavlarında okuma becerisinin ölçülmesinde çoğunlukla aynı paragrafa bağlı birden fazla maddenin yer aldığı sorular kullanılmaktadır. Söz konusu paragrafa bağlı maddeler, “testlet” temelli MTK modellerine göre kalibre edilmektedir. Bu nedenle bu tür maddelerin de bulunacağı gerçek-zamanlı BBT uygulamaları, gelecekte üzerinde çalışılması gereken önemli bir konu olduğu düşünülmektedir.

- Bifaktör modeller, aynı becerinin farklı alt boyutlarından genel bir yetenek kestirmekle birlikte; “testlet” temelli MTK modellerinin kullanıldığı durumlarda, birden fazla madde için ortak bir yetenek kestirimi yapmak için de kullanılabilirler (De Mars, 2006). Buna bağlı olarak, gelecekte okuma becerisinin de ekleneceği çok boyutlu bir BBT uygulamasında paragrafa yönelik maddeler için de bifaktör modelin kullanılması, gerçek-zamanlı uygulamalar için bir alternatif oluşturacaktır.
- Bu araştırma kapsamında kullanılan bazı yöntemler, mirtCAT paketinin sınırlılıkları doğrultusunda şekillenmiştir. Paketin yaratıcısı, ihtiyaçlar ve istekler doğrultusunda çalışmalar yürütülerek geliştirileceğini belirtmektedir. Paketin, bu çalışmanın yürütüldüğü zaman aralığındaki özellikleri başlama noktası için belirli bir yetenek aralığından rastgele madde seçilmesi imkanını vermemektedir. Gelecekte paketin güncellenmesiyle bu özelliğin eklenmesi durumunda, söz konusu durumun maddelerin teşhir/kullanım oranları ve yetenek kestirimlerine etkisi üzerine çalışmalar yürütülebilir.
- Bu çalışma kapsamında yetenek kestirim yöntemi olarak Bayes yöntemlerden EAP ve MAP kullanılmıştır. Gelecekte yapılacak bifaktör modele dayalı simülasyon çalışmalarında ML yetenek kestirim yöntemiyle elde edilen korelasyon, hata değerleri ve yanıtlanan madde sayılarının Bayes yöntemlerle elde edilen değerlerle karşılaştırılması literatüre katkı sağlayacak, gerçek-zamanlı uygulamalara yol gösterici olacaktır.
- Bu çalışmada, İngilizce yeterliliğini belirlemeye yönelik bir sınav için bifaktör modele dayalı çok boyutlu bir BBT uygulaması yapılmıştır. Ülkemizde, gerek ortaöğretim gerekse üniversite seçme sınavlarında bifaktör modele dayalı çok boyutlu BBT uygulamalarına yönelik çalışmalara da yer verilmelidir. Söz konusu sınavların, bireyler için farklı alt testlerden bir genel yetenek kestirimi yaptığı düşünüldüğünde, bu sınavlara bifaktör modelin uygun olabileceği düşünülebilir.
- Ülkemizde, ulusal çapta İngilizce yeterliliğini test etmeye yönelik lisansüstü seviyede uzun yıllardır uygulanan YDS ve bu araştırmanın bitirildiği tarihte henüz ilki yapılmış olan YÖK-DİL yer almatadır. Bu sınavlar, İngilizce'nin farklı boyutlarını ölçen ve buna dayalı olarak toplam puan üzerinden sonuç

bildiren kağıt-kalem sınavlarıdır. Bu sınavların da, özellikle bifaktör modele dayalı olarak çok boyutlu BBT haline getirilmesine yönelik çalışmaların yararlı olabileceği düşünülmektedir.

6.2.2 Gerçek-Zamanlı Çok Boyutlu BBT Uygulamalarına Yönelik Öneriler

- Bu araştırma kapsamında, hipotetik olarak çok boyutlu olmasının yanında genel bir yetenek kestirimi yapma amacıyla uygulanan İngilizce yeterlik sınavı verileri bifaktör modele göre kalibre edilerek gerçek-zamanlı bir uygulama yapılmıştır. Hipotetik olarak çok boyutlu olup, benzer şekilde genel bir yetenek kestirimi yapma amacıyla olan yapıların ölçüldüğü başarı testleri için (fen becerisi, matematik vb.), bu model uygulanarak çok boyutlu gerçek-zamanlı BBT uygulamalarının eğitimdeki uygulamaları genişletilebilir.
- Araştırma sonunda, bifaktör modele uygun çok boyutlu bir BBT uygulaması yapılmış; bireylerin, kağıt-kalem testinden elde ettikleri toplam puanla gerçek-zamanlı uygulamadaki genel yetenek kestirimleri arasında yüksek ve anlamlı bir korelasyon bulunmuştur. Söz konusu kestirimlerin çok daha az madde ve dolayısıyla çok daha kısa bir zamanda yapıldığı göz önüne alındığında, İngilizce yeterlilik için üniversitelerde yapılan sınavların bu formatta uygulanabilirliği tartışılmalıdır.
- Bu çalışma kapsamında kullanılan madde havuzu 200 sorudan oluşmaktadır. Yapılacak bir uygulamada, soru havuzunun büyütülerek madde teşhir oranının da kontrol altına alınmasının sağlanması, sınav güvenliğini artıracaktır.
- Bifaktör modele dayalı gerçek-zamanlı çok boyutlu BBT uygulamasında, grup faktörlerin önemli görülmesi durumunda D-rule madde seçim yöntemi kullanılması önerilmektedir.
- Bifaktör modeller, genel yeteneği ölçmekle birlikte çok boyutluluğu da ihmal etmeyen yapılardır. Ancak bu faktör analitik modellerde; elde edilen bilgi öncelikle genel yetenek için kullanılırken, kalan bilgi grup faktörlerin kestirilmesine yöneliktir (Seo, 2012). Bu durum göz önüne alındığında, çok boyutlu yapılara yönelik uygulamalarda ancak genel yeteneğin dakik (pure)

kestirilmesinin grup faktörlerin kestirilmesine göre çok daha önemli olduğu durumlarda bu modelin tercih edilmesi önerilebilir.



KAYNAKÇA

- Akyıldız M. & Şahin, M. D. (2015). *Dimensionality and scalability of a large scale exam by means of item response theory*. Paper Presented at International Congress on Education for the Future: Issues and Challenges, May 13-15, 2015.
- Akyıldız M. & Şahin, M. D. (2016). *Anadolu Üniversitesi İngilizce yeterlilik sınavının bireyselleştirilmiş bilgisayarlı test olarak uygulanması*. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi. Akdeniz Üniversitesi-Antalya, 1-3 Eylül.
- Aybek, E. C. (2016). *Kendini Değerlendirme Envanteri'nin BOBUT olarak uygulanabilirliğinin araştırılması* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Babcock, B. & Weiss, D. J. (2012). Termination criteria in computerized adaptive tests: do variable-length CATs provide efficient and effective measurement? *Journal of Computerized Adaptive Testing*, 1(1), 1-18.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Bloxom, B., & Vale, C.D. (1987). Multidimensional adaptive testing: An approximate procedure for updating. In *meeting of the psychometric society*. Montreal, Canada, June.
- Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-459.
- Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, 12, 261-280.
- Boztunç-Öztürk, N. (2014). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı kontrol yöntemlerinin incelenmesi* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Bulut, O & Kan, A. (2012). Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eurasian Journal of Educational Research*, 49, 61-80.
- Chang H., Ying Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3), 213-229.
- Cai, L., Yang, J., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, 16, 221-248.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29.
- Chalmers, R. P. (2017). mirtCAT: Computerized adaptive testing with multidimensional item response theory. R package version 1.4.
- Chalmers, R. P. (2016). Generating adaptive and non-adaptive test interfaces for multidimensional item response theory applications. *Journal of Statistical Software*, 71(5), 1-38.

- Chernyshenko, O. S., Stark, S., & Chan, K. Y. (2001). Investigating the hierarchical factor structure of the fifth edition of the 16PF: An application of the Schmid-Leiman orthogonalization procedure. *Educational and Psychological Measurement*, 61, 290-302.
- Choi S. W., King D. (2014). *MAT: Multidimensional adaptive testing*. R package version 2.2.
- Choi, S. W., Grady, M. W., & Dodd, B. G. (2010). A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurement*, 71, 37–53.
- Colwell, N. M. (2013). Test anxiety, computer-Adaptive testing and the common core. *Journal of Education and Training Studies*, 1(2), 50-60.
- Cronbach, L. J. & Gleser, G. C. (1965). *Psychological test and personnel decisions* (2nd Ed). Urbana: University of Illinois Press.
- Cömert, M. (2008). *Bireye uyarlanmış bilgisayar destekli ölçme ve değerlendirme yazılımı geliştirilmesi* (Yayınlanmamış Yüksek Lisans Tezi). Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Çelen, Ü. & Aybek, E. C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen Puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 4(2), 64-75.
- De Ayala, R.J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- De Mars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement*, 43(2), 145-168.
- Eggen, T. (2007). Choices in CAT models in the context of educational testing. *Presented at the CAT Models and Monitoring Paper Session, June 7, 2007* (11.11.2016 tarihinde <http://publicdocs.iacat.org/cat2010/cat07eggen.pdf> adresinden erişilmiştir).
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Eroğlu, M. G. (2013). Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Fischer, G. H., & Molenaar, I. W. (1995). *Rasch models: Foundations, recent developments and applications*. New York: Springer.
- Frick, T. (1992). Computerized adaptive mastery tests as expert systems. *Journal of Educational Computing Research*, 8(2), 187-213
- Fritz, B. E. & Marszalek, J. M. (2010). Computerized adaptive testing, anxiety levels, and gender differences. *Social Psychology of Education. Volume 13* (3), 441–458.
- Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3), 357-374.

- Frey, A. & Nicki-Nils, S. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation*, 35, 89–94
- Gelbal, S. (2013). *Ölçme ve değerlendirme*. Eskişehir: Anadolu Üniversitesi Açıköğretim Yayınları.
- Gialluca, K. A., & Weiss, D. J. (1979). Efficiency of an adaptive inter-subset branching strategy in the measurement of classroom achievement (Research Report 79-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program. 07.01.2017 tarihinde <http://www.dtic.mil/get-tr-doc/pdf?AD=ADA080956> adresinden erişilmiştir.
- Gustafsson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407-434.
- Gustafsson, J., & Aberg-Bengtsson, L. (2010). *Unidimensionality and the interpretability of psychological instruments*. In S. E. Embretson (Ed.), *Measuring psychological constructs* (pp. 97-121). Washington, DC: American Psychological Association.
- Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., & Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, 31, 4-19.
- Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bifactor analysis. *Psychometrika*, 57, 423-436.
- Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., & ...Immekus, J. C. (2008). Using computerized adaptive testing to reduce the burden of zihinsel health assessment. *Psychiatric Services*, 59(4), 49-58.
- Gündeğer C. & Doğan, N. (2016). Bireyselleştirilmiş Bilgisayarlı Sınıflama Testlerinde Madde Havuzu Özelliklerinin Test Uzunluğu ve Sınıflama Doğruluğu Üzerindeki Etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* (Basım aşamasında).
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Academic Publishers.
- Hambleton, R. K., Zaal, J. N. & Pieters, J. P. M. (1991). *Computerized adaptive testing: Theory, applications and standards*. In R. K. Hambleton & J. N. Zaal (Eds). *Advances educational and psychological testing*. Springer Science+Business Media, LLC.
- Hart, D. L., Cook, K. F., Mioduski, J. E., Teal, C. R., & Crane, P. K. (2006). Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. *Journal of Clinical Epidemiology*, 59, 290–298.
- Holzinger, K. J., & Swineford, F. (1937). The bifactor method. *Psychometrika*, 2, 41-54.
- İşeri, A. İ. (2002). *Assessment of students' mathematics achievement through computer adaptive testing procedures* (Unpublished Doctoral Dissertation). Middle East Technical University, The Graduate School of Natural and Applied Sciences, Ankara.

- Huang, H., Chen, P & Wang, W. (2012). Computerized adaptive testing using a class of high-order item response theory. *Applied Psychological Measurement*, 36(8), 689-706.
- Kalender, İ. (2009). Başarı ve yetenek kestirimlerinde yeni bir yaklaşım: Bilgisayar Ortamında Bireyselleştirilmiş Testler (Computerized Adaptive Tests - CAT). *CITO*, 09.09.2016 tarihinde https://www.researchgate.net/publication/278955539_Basari_ve_Yetenek_Kestirimi_minde_Yeni_Bir_Yaklasim_Bilgisayar_Ortaminda_Bireysellestirilmis_Testler_adresinden_erisilmistir.
- Kalender, İ (2011). *Effects of different computerized adaptive testing strategies on recovery of ability* (Unpublished Doctoral Disertation) METU, The Graduate School of Natural and Applied Sciences, Ankara.
- Kaptan, F. (1993). *Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kağıt-kalem testi uygulamasının karşılaştırılması* (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Karasar, N. (2010). *Bilimsel araştırma yöntemi*. Ankara. Nobel Yayın Dağıtım.
- Kaskatı, O.T. (2011). *Rasch modelleri kullanarak romatoid artirit hastaları özürüllük değerlendirimi için bilgisayar uyarlamalı test yöntemi geliştirilmesi* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.
- Kezer, F. (2013). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.
- Kullback, S. & Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79–86.
- Larkin, K. C. & Weiss, D. J. (1975). *An empirical comparison of two-stage and pyramidal adaptive ability testing* (Research Report, 75-1). Minneapolis: Psychometrics Methods Program, Department of Psychology, University of Minnesota.
- Lee, Y H., Ip, E. H., & Fuh, C. D. (2002). *A strategy for controlling item exposure in multidimensional computerized adaptive testing*. 14.10.2016 tarihinde http://www3.stat.sinica.edu.tw/library/c_tec_rep/c-2002-11.pdf adresinden erişilmiştir.
- Leg, S. M. & Buhr, D. C. (1992). Computerized adaptive testing with different groups. *Educational Measurement: Issues and Practice*, 11(2), 23-27.
- Li, Y. H., & Schafer, W. D. (2005). Trait parameter recovery using multidimensional computerized adaptive testing in reading and mathematics. *Applied Psychological Measurement*, 29, 3-25.
- Lord, F. M. (1952). A theory of test scores. *Psychometric Monograph No. 7*. Psychometric Society.

- Lord, F. M. (1953a). An application of confidence intervals and ML to the estimation of examinee's ability. *Psychometrika*, 18, 57-75.
- Lord, F. M. (1953b). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13, 517-548.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of zihinsel test scores*. Oxford, England: Addison-Wesley.
- Lord, F. M. (1970). Some test theory for tailored testing. In W. H. Holtzman (Ed.), *Computer assisted instruction, testing, and guidance*. New York: Harper and Row.
- Lord, F. M. (1971a). Robbins-Monro procedures for tailored testing. *Educational and Psychological Measurement*, 31, 2-31.
- Lord, F. M. (1971b). The self-scoring flexilevel test. *Journal of Educational Measurement*, 8, 147-151.
- Luecht, R. M. (1996). Multidimensional computerized adaptive testing in a certification or licensure context. *Applied Psychological Measurement*, 20(4), 389-404.
- Magis D, Raïche G (2012). Random generation of response patterns under computerized adaptive testing with the R Package catR." *Journal of Statistical Software*, 48(8), 1-31.
- Maurelli, V., & Weiss, D. J. (1981). *Factors influencing the psychometric characteristics of an adaptive testing strategy for test batteries* (Research Rep. No. 81-4). 01.12.2016 tarihinde <http://iacat.org/sites/default/files/biblio/ma81-04.pdf> adresinden indirilmiştir.
- Mills, C. N., & Stocking, M. L. (1996). Practical issues in computerized adaptive testing. *Applied Psychological Measurement*, 9, 287-304.
- Mulder, J., & van der Linden, W. J. (2009). Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika*, 74, 273-296.
- Nydic S. (2014). *catlrt: An R package for simulating IRT-based computerized adaptive tests*. R package version 0.5-0.
- Nydic, S. & Weiss, D. J. (2009). A hybrid simulation procedure for developments of CATs. *Paper presented at the Item Pool Development Paper session at the 2009 GMAC Conference on Computerized Adaptive Testing*.
- 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 zihinsel testing. *Journal of the American Statistical Association*, 70, 351-356.
- Özdemir, B. (2015). *Madde Düzeyinde Boyutluluk Modellerinin Bilgisayar Ortamında Bireyselleştirilmiş Test Yöntemleri Üzerindeki Etkisinin İncelenmesi* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Özbaşı, D. (2014). Bilgisayar okuryazarlığı testinin BOBUT olarak uygulanabilirliğine ilişkin

bir araştırma (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Progar, S. & Sočan, G. An empirical comparison of Item Response Theory and Classical Test Theory. *Horizons of Psychology*, 17(3), 5-24.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL. <http://www.R-project.org>.

Reckase, M., D. (2009). *Multidimensional item response theory: Statistics for social and behavioral sciences*. New York, NY: Springer.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92, 544-559.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16, 19-31.

Rudner, L. M. (2002). *An examination of decision-theory adaptive testing procedures*. Paper presented at the annual meeting of the American Educational Research Association, April 1-5, 2002, New Orleans, LA. 18.02.2017 tarihinde https://www.researchgate.net/publication/255646477_An_Examination_of_Decision-Theory_Adaptive_Testing_Procedures adresinden erişilmiştir.

Rudner, L.M. and Guo, F.M. (2011). Computer adaptive testing for small scale programs and instructional systems. *Journal of Applied Testing Technology*, 12, Special Issue.

Sands, W. A., Waters, B. K. & McBride, J. R. (1997). *CATBOOK. Computerized adaptive testing: From inquiry to operation*. United States Army Research Institute for the Behavioral and Social Sciences.

Sarkar, D. (2011). Lattice: Trellis graphics with R. R package Version 0.20-30.

Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61, 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*. Dordrecht/Boston/London: Kluwer.

Segall, D. O. (2001). General ability measurement: An application of multidimensional item response theory. *Psychometrika*, 66, 79-97.

Segall, D. O. (2005). *Computerized adaptive testing*. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement*. New York: Academic Press.

Seo, D. G. (2012). *Application of the bifactor model to computerized adaptive testing*. Unpublished Doctoral Dissertation. University of Minnesota.

Seo, D. G. & Weiss, D. J. (2015). Best Design for Multidimensional Adaptive Testing With the Bifactor Model. *Educational and Psychological Measurement*, 75(6), 954-978.

- Sheng Y., Wikle C. K. (2007). Comparing multiunidimensional and unidimensional IRT models. *Educational & Psychological Measurement*, 67, 899–919.
- Stocking, M. L. (1994). *Three practical issues for modern adaptive testing item pools*. (ETS Research Report No. 93–2). Princeton, NJ: Educational Testing Service.
- Su, Y. (2016). A comparison of constrained item selection methods in multidimensional computerized adaptive testing. *Applied Psychological Measurement*, 40(5) 346–360.
- Sulak, S. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında kullanılan madde seçme yöntemlerinin karşılaştırılması* (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Şahin, M. D. & Akyıldız, M. (2016). Bifaktör modelin tek boyutlu, çok boyutlu ve ikinci dereceden modellerle karşılaştırılması. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi. Akdeniz Üniversitesi-Antalya, 1-3 Eylül.
- Tam, S. S. (1992). A comparison of methods for adaptive estimation of a multidimensional trait. Unpublished Doctoral Dissertation, Columbia University.
- Thompson, N. A. (2007). A practitioner's guide for variable-length computerized classification testing. *Practical Assessment, Research & Evaluation*, 12(1), 1-13.
- Thompson, N. A. & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*. 16(1). 1-9.
- 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. (Eds.). (2000). *Computerized adaptive testing: Theory and practice*. Dordrecht/Boston/London: Kluwer.
- van der Linden, W. & Glas, G. A. W. (2002). *Computerized adaptive testing: theory and practice*. Kluwer Academic Publishers. Dordrecht.
- Veldkamp, B. P., & van der Linden, W. J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika*, 67, 575-588.
- Wainer, H. W., Dorans, N. J., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg, L. & Thissen, D. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Erlbaum.
- Wainer, H. (2000). Rescuing computerized testing by breaking Zipf's law. *Journal of Educational and Behavioral Statistics*, 25, 203-224.
- Wang, C., Chang, H. & Boughton, K. A. (2012). Deriving stopping rules for multidimensional computerized adaptive testing. *Applied Psychological Measurement*, 37(2), 99–122.
- Wang, C., & Chen, H. (2004). Implementation and measurement efficiency of

- multidimensional computerized adaptive testing. *Applied Psychological Measurement*, 28, 295-316.
- Wang, S., & Wang, T. (2001). Precision of Warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*, 25, 317-331.
- Weiss, D. J. (1973). *The stratified adaptive computerized ability test* (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology.
- Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, 53, 774-789.
- Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37(2), 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), 1-27.
- Weiss, D. J., & Gibbons, R. D. (2007). Computerized adaptive testing with the bifactor model. *Paper presented at the New CAT Models session at the 2007 GMAC Conference on Computerized Adaptive Testing*. 12.10.2016 tarihinde <http://publicdocs.iacat.org/cat2010/cat07weiss&gibbons.pdf> adresinden erişilmiştir.
- 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.
- 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.
- Yao, L., Pommerich, M & Segall, D. O. (2014). Using multidimensional CAT to administer a short, yet precise, screening test. *Applied Psychological Measurement* 2014, 38(8) 614-631.
- Yoes, M. (1995). *An updated comparison of micro-computer based item parameter estimation procedures used with the 3-parameter IRT model*. Saint Paul, MN: Assessment Systems Corporation.
- Zaman, A., Kashmiri, A., Mubarak, M. & Ali, A. (2008). *Students ranking, based on their abilities on objective type test: comparison of CTT and IRT*. Paper presented at EDU-COM 2008 International Conference, Sustainability in Higher Education: Directions for Change. 19-21 November, 2008. 17.11.2016 tarihinde <http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1051&context=ceducum> adresinden erişilmiştir.

EKLER DİZİNİ

EK 1. ETİK KOMİSYON İZİN MUAFİYETİ FORMU

Form: 40

Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

22 / 01 / 2016

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

Tez Başlığı / Konusu:	GERÇEK ve ÜRETİLMİŞ VERİ SETLERİNDE ÇOK BOYUTLU BİREYSELLEŞTİRİLMİŞ BİLGİSAYARLI TEST UYGULAMA SONUÇLARININ KARŞILAŞTIRILMASI
----------------------------------	---

Yukarıda başlığı/konusu gösterilen tez çalışmam:

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

Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Murat Doğan ŞAHİN

Öğrenci Bilgileri

Adı Soyadı	Murat Doğan ŞAHİN
Öğrenci No	N11243756
Anabilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	☐ Yüksek Lisans ☒ Doktora ☐ Bütünleşik Dr.

Danışman Görüşü ve Onayı

Prof. Dr. Selahattin GELBAL

EK 2. ORİJİNALLİK RAPORU



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
DOKTORA TEZ ÇALIŞMASI ORİJİNALLİK RAPORU

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

Tarih: 23/5/2017

Tez Başlığı : GERÇEK ve ÜRETİLMİŞ VERİ SETLERİNDE ÇOK BOYUTLU BİREYSELLEŞTİRİLMİŞ BİLGİSAYARLI TEST UYGULAMA SONUÇLARININ KARŞILAŞTIRILMASI

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

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Endeksi	Gönderim Numarası
23/05/2017	131	223287	14/04 /2017	%2	817791628

Uygulanan filtreler:

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

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

Gereğini saygılarımla arz ederim.

Adı Soyadı: MURAT DOĞAN ŞAHİN
Öğrenci No: N11243756
Anabilim Dalı: EĞİTİM BİLİMLERİ
Programı: EĞİTİMDE ÖLÇME ve DEĞERLENDİRME
Statüsü: ☐ Y.Lisans ☒ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

(Prof. Dr. Selahattin GELBAL)



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

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

Date: 23/5/2017

Thesis Title : EXAMINATION OF MULTIDIMENSIONAL COMPUTERIZED ADAPTIVE TESTING APPLICATION RESULTS IN REAL AND GENERATED DATA SETS

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

Time Submitted	Page Count	Character Count	Date of Thesis Defence	Similarity Index	Submission ID
23/05/2017	131	223287	14/04 /2017	%2	817791628

Filtering options applied:

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

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

I respectfully submit this for approval.

Name Surname: MURAT DOĞAN ŞAHİN
Student No: N11243756
Department: EDUCATIONAL SCIENCES
Program: EDUCATIONAL MEASUREMENT and EVALUATION
Status: ☐ Masters ☒ Ph.D. ☐ Integrated Ph.D.

ADVISOR APPROVAL

APPROVED

(Prof. Dr. Selahattin GELBAL)

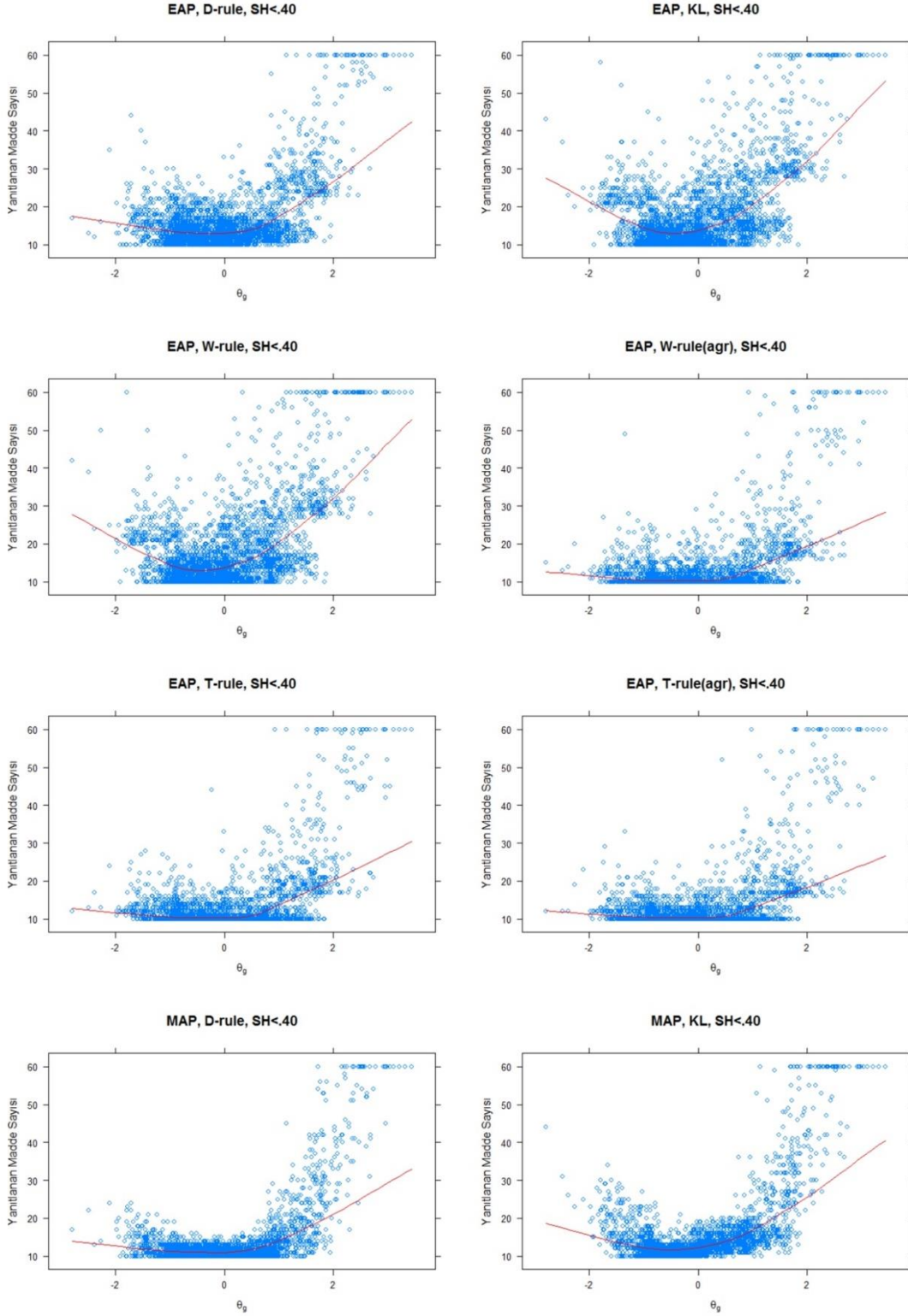
EK 3. ÇOK BOYUTLU BBT UYGULAMASI İÇİN KULLANILAN HAVUZDA YER ALAN MADDELERE AİT PARAMETRELER

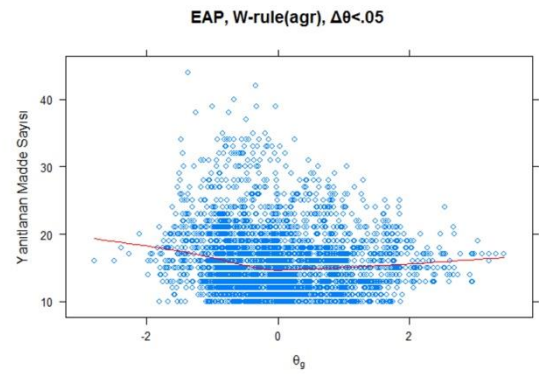
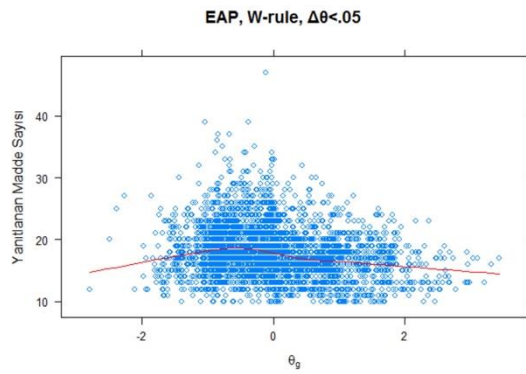
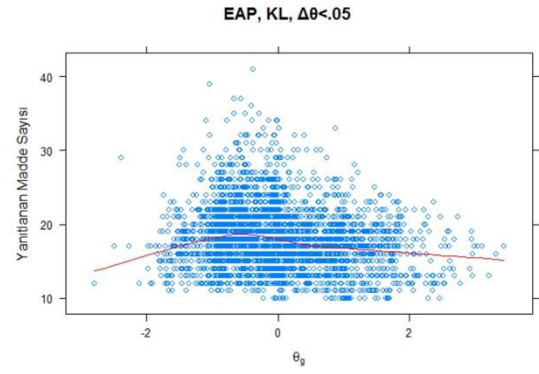
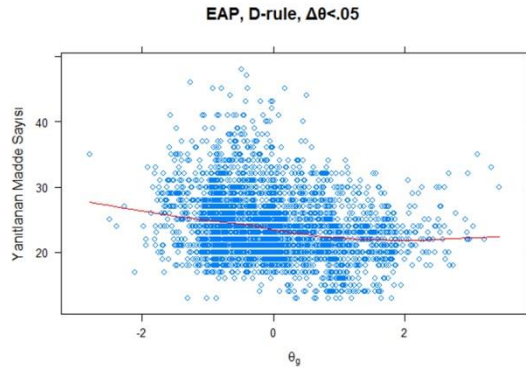
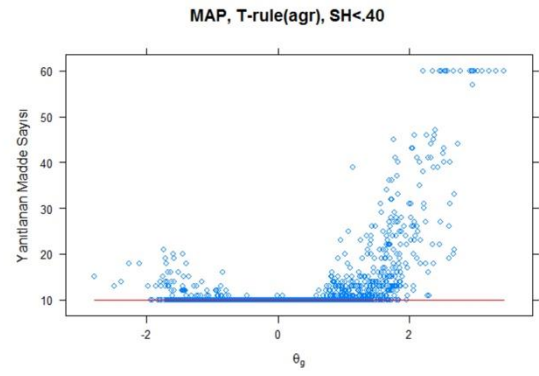
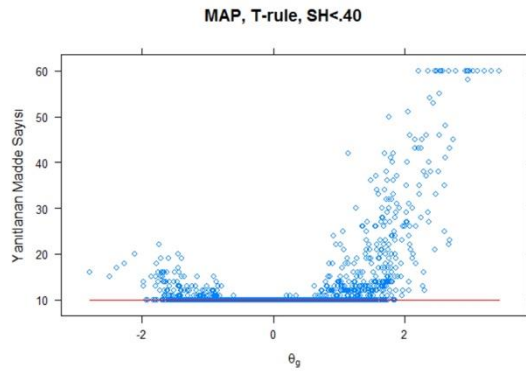
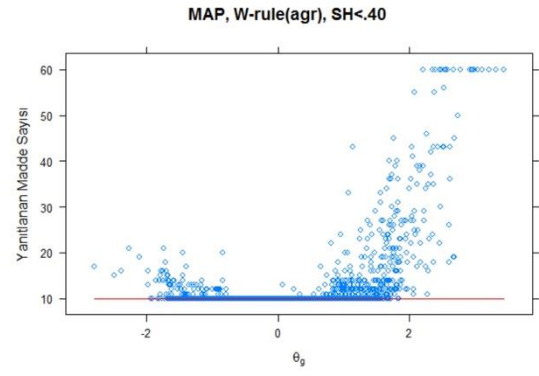
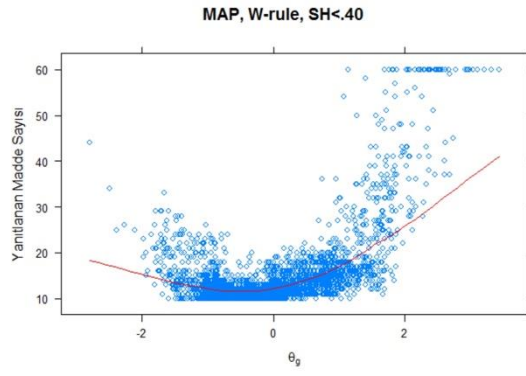
Soru No	a1	a2	a3	b	Soru No	a1	a2	a3	b
1	0.706	0.68	0	0.388	101	1.54	0.087	0	-1.216
2	0.67	0.09	0	0.121	102	0.569	0.448	0	-1.57
3	0.183	0.141	0	2.316	103	0.936	0.194	0	1.927
4	1.474	0.934	0	1.221	104	1.27	0.051	0	-0.539
5	0.512	0.127	0	0.921	105	0.787	0.06	0	0.146
6	1.694	-0.196	0	2.057	106	1.483	0.954	0	-0.272
7	0.913	0.335	0	0.317	107	0.966	-0.603	0	1.49
8	1.678	0.132	0	1.202	108	0.66	0.046	0	-0.3
9	1.69	0.136	0	0.628	109	0.362	0.167	0	-0.206
10	1.965	0.733	0	2.502	110	0.313	0.42	0	0.401
11	1.597	1.021	0	1.357	111	0.927	0.262	0	1.491
12	1.511	-1.173	0	2.282	112	2.371	0.737	0	-0.566
13	0.842	-0.33	0	1.328	113	0.016	0.902	0	0.483
14	0.919	0.019	0	1.236	114	1.131	0.975	0	-1.107
15	0.997	0.211	0	1.16	115	0.892	0.447	0	0.027
16	1.015	-0.32	0	1.262	116	1.208	-0.422	0	1.048
17	0.941	0.003	0	1.376	117	1.328	0.232	0	-0.013
18	0.348	0.28	0	-0.095	118	0.907	0.465	0	1.014
19	1.566	0.166	0	1.092	119	1.307	-0.164	0	-0.735
20	1.11	0.613	0	0.035	120	1.27	0.336	0	-0.353
21	1.706	0.273	0	0.787	121	2.011	0.215	0	-1.045
22	1.154	0.369	0	0.956	122	0.78	-0.287	0	1.1
23	2.111	-0.026	0	-0.128	123	2.175	0.662	0	-1.771
24	0.358	0.065	0	-0.662	124	0.436	0.034	0	-1.53
25	0.637	-0.251	0	0.226	125	1.075	0.398	0	0.844
26	1.335	0	-0.01	2.792	126	0.565	0	0.255	-0.362
27	1.452	0	-0.097	0.84	127	1.228	0	0.628	2.454
28	0.679	0	0.313	-0.668	128	1.415	0	0.198	0.185
29	1.214	0	0.541	-0.127	129	1.121	0	0.369	0.173
30	1.422	0	0.207	0.819	130	1.101	0	0.564	-0.329
31	0.346	0	0.364	1.211	131	1.565	0	0.652	-0.802
32	0.82	0	0.104	0.837	132	1.383	0	1.28	0.994
33	1.788	0	0.355	0.927	133	1.288	0	0.608	0.464
34	2.168	0	2.587	2.046	134	0.465	0	0.532	1.219
35	1.936	0	-0.023	0.813	135	1.158	0	0.596	0.55
36	2.638	0	0.568	2.516	136	1.292	0	0.819	0.701
37	1.482	0	0.231	0.755	137	1.442	0	0.642	0.859
38	0.681	0	0.476	0.615	138	1.228	0	0.548	-0.456
39	1.221	0	0.867	2.595	139	0.846	0	0.456	-0.33
40	0.355	0	0.363	-0.514	140	2.055	0	0.943	1.659
41	0.719	0	0.633	-0.38	141	1.347	0	-0.022	-1.416
42	0.069	0	0.53	-0.823	142	0.163	0	-0.151	-2.057
43	1.875	0	0.635	-0.078	143	0.578	0	0.22	-0.861
44	1.507	0	1.353	3.696	144	0.865	0	0.637	0.12
45	1.982	0	0.161	1.577	145	0.43	0	0.142	-2.107
46	0.899	0	0.042	1.783	146	0.75	0	0.3	-0.736
47	2.015	0	0.128	1.298	147	0.855	0	0.377	0.678
48	0.994	0	0.057	-1.102	148	1.16	0	0.479	-1.348
49	1.77	0	0.001	2.37	149	1.36	0	1.132	-0.466
50	0.277	0	0.001	2.162	150	0.665	0	0.292	-1.477
51	0.326	0.253	0	-0.786	151	0.972	0.443	0	-0.609
52	0.045	0.21	0	1.556	152	1.131	0.363	0	1.786
53	0.637	0.265	0	1.695	153	1.094	0.493	0	2.544
54	0.312	0.653	0	2.62	154	0.138	-0.091	0	-1.865
55	0.386	0.591	0	1.316	155	0.529	0.261	0	0.102
56	0.538	0.344	0	1.505	156	0.607	0.051	0	2.357
57	0.36	0.151	0	-0.532	157	0.76	0.37	0	2.237
58	0.621	0.24	0	0.705	158	0.263	0.372	0	1.121
59	0.08	0.274	0	-0.853	159	0.685	0.699	0	1.206
60	0.317	0.461	0	1.442	160	0.361	0.207	0	-1.244

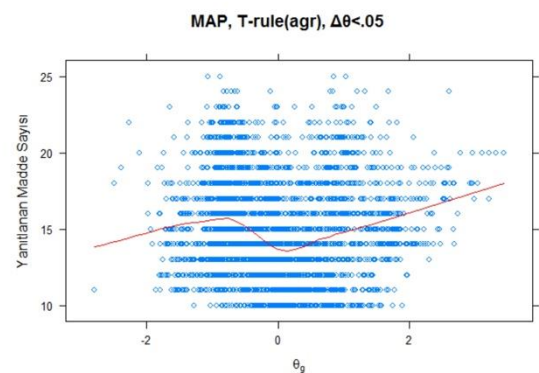
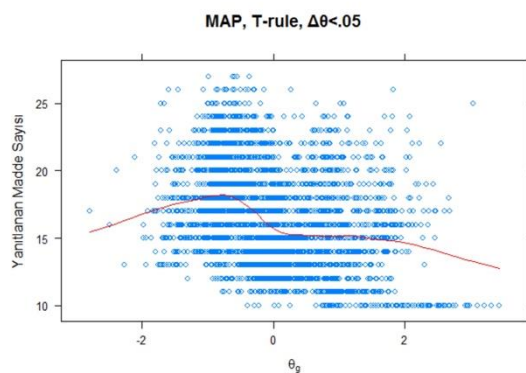
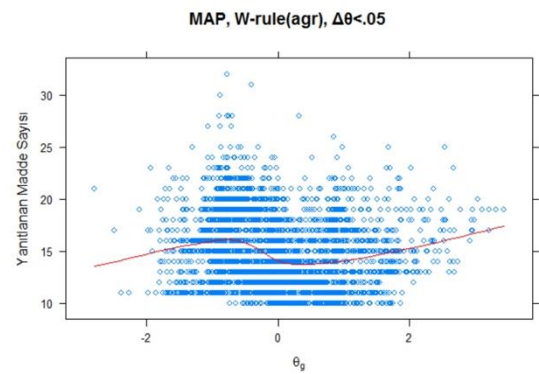
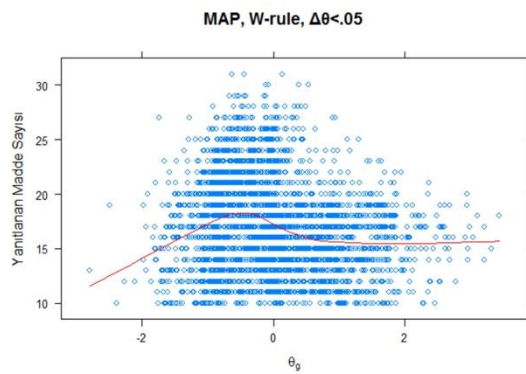
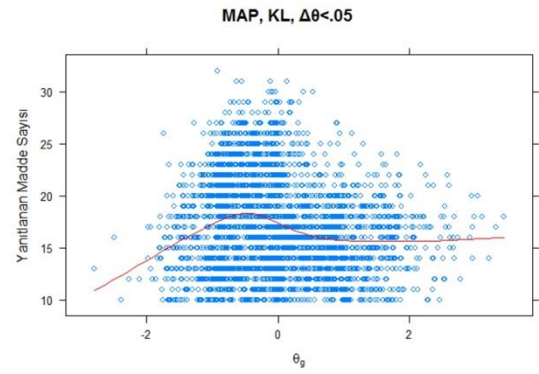
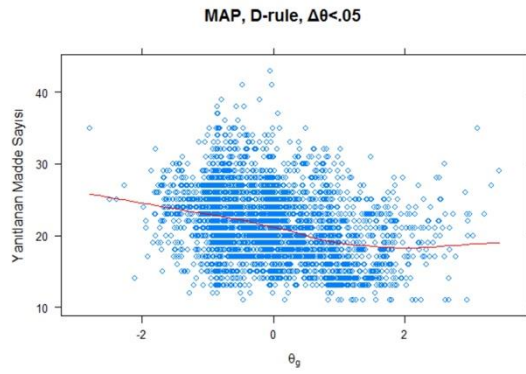
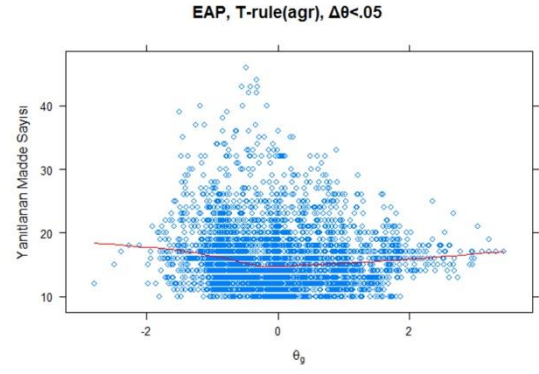
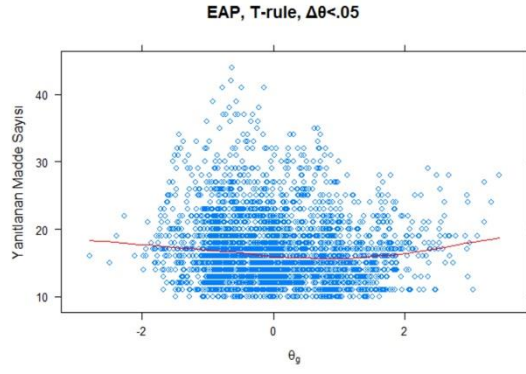
61	0.311	-0.024	0	-1.252	161	0.361	0.307	0	0.063
62	0.725	-0.019	0	-2.607	162	0.691	0.889	0	1.375
63	0.153	0.718	0	2.627	163	0.376	-0.055	0	-1.633
64	0.26	0.047	0	-1.595	164	0.563	0.305	0	1.896
65	0.271	0.478	0	-0.018	165	0.86	0.383	0	2.297
66	0.392	1.003	0	3.108	166	0.359	0.282	0	-0.609
67	0.046	1.048	0	0.932	167	0.493	-0.052	0	-1.085
68	0.862	0.094	0	-0.126	168	0.969	0.309	0	1.263
69	0.999	-0.031	0	-0.776	169	0.223	0.282	0	-0.62
70	0.376	0.941	0	3.337	170	0.508	0.174	0	-2.454
71	0.074	0.34	0	0.478	171	0.373	0.284	0	0.052
72	0.562	0.643	0	3.041	172	0.905	0.678	0	-0.006
73	0.957	-0.261	0	-2.291	173	0.595	0.499	0	0.847
74	0.037	-0.04	0	-1.927	174	0.558	0.184	0	-0.13
75	0.175	-0.065	0	0.134	175	0.604	0.592	0	-0.952
76	0.29	0	0.147	1.078	176	0.506	0	-0.019	0.396
77	0.312	0	0.228	-0.923	177	0.738	0	-0.19	-0.163
78	0.234	0	0.242	0.465	178	1.174	0	0.21	1.803
79	0.926	0	0.097	1.87	179	0.653	0	0.05	-0.25
80	0.909	0	0.364	-1.814	180	0.758	0	-0.109	-1.134
81	0.615	0	0.335	-0.606	181	1.3	0	0.097	1.876
82	1.856	0	-1.99	-0.389	182	1.31	0	0.385	1.158
83	0.674	0	-0.038	0.536	183	0.433	0	0.09	0.903
84	0.54	0	0.092	0.802	184	0.835	0	0.222	1.263
85	0.376	0	0.354	-0.531	185	1.264	0	0.432	1.737
86	1.125	0	0.045	0.726	186	0.471	0	0.128	0.551
87	1.054	0	0.31	-0.405	187	0.597	0	-0.042	-0.003
88	0.073	0	0.038	-1.062	188	0.354	0	0.075	-1.341
89	0.219	0	0.182	-1.034	189	1.341	0	1.019	0.021
90	1.225	0	-0.072	-0.803	190	0.471	0	0.34	-1.077
91	0.571	0	0.381	-0.776	191	0.318	0	-0.13	-0.757
92	0.628	0	0.41	-1.164	192	0.74	0	0.108	-1.437
93	0.852	0	0.149	0.769	193	0.928	0	0.126	-2.628
94	0.467	0	-0.051	0.568	194	0.66	0	-0.088	-0.864
95	0.742	0	0.112	-0.281	195	0.679	0	-0.112	-1.356
96	0.313	0	0.229	-0.544	196	0.438	0	0.574	-1.291
97	0.826	0	0	1.001	197	0.83	0	0.115	-1.323
98	0.539	0	0.226	-0.485	198	0.95	0	0.67	1.831
99	0.454	0	0.252	-0.55	199	0.892	0	0.48	0.797
100	1.086	0	0.559	0.785	200	1.17	0	0.504	2.008

EK 4. SİMÜLASYONLAR SONUCU HER BİR KOŞULDA TÜM BOYUTLAR İÇİN ELDE EDİLEN GRAFİKLER

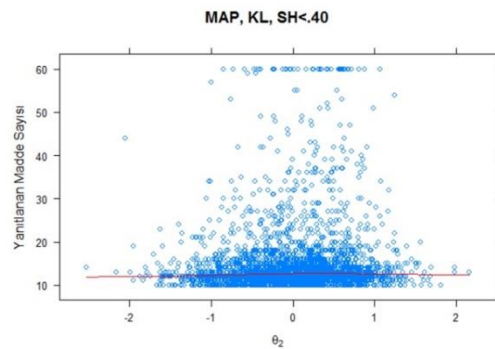
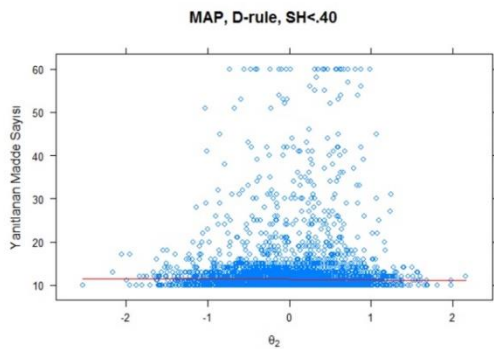
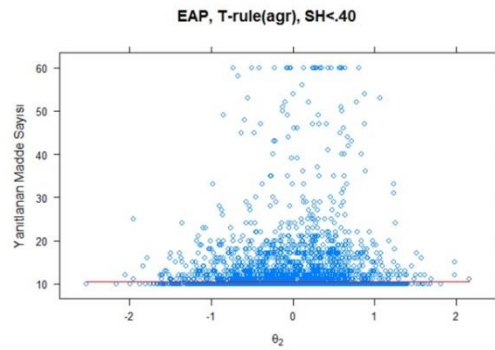
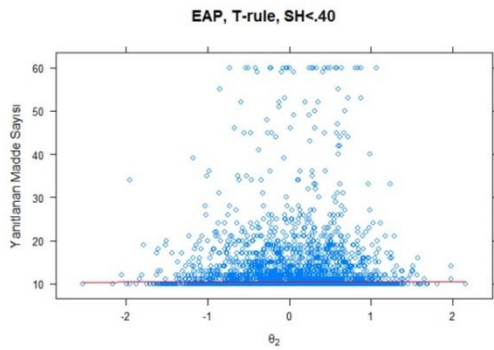
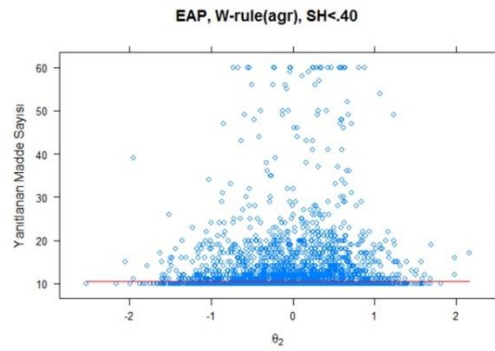
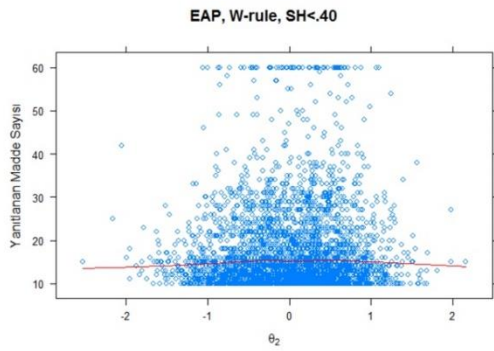
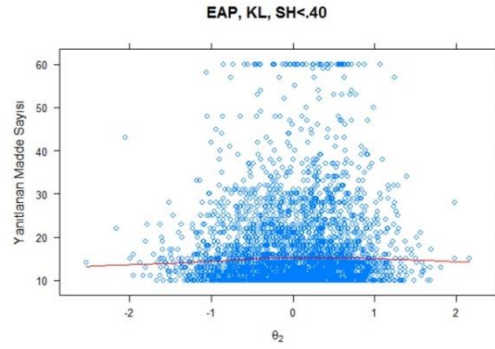
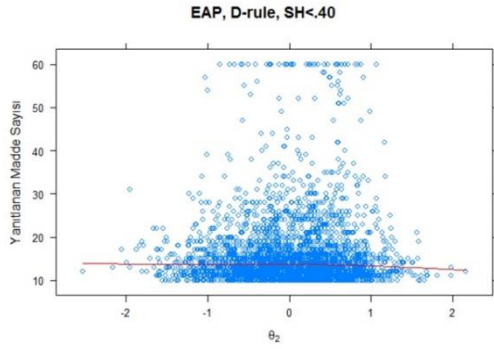
EK 4.1 GENEL YETENEK İÇİN ELDE EDİLEN YANITLANAN MADDE SAYISI – THETA GRAFİKLERİ

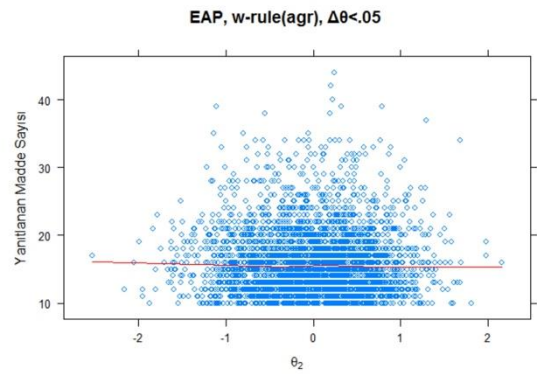
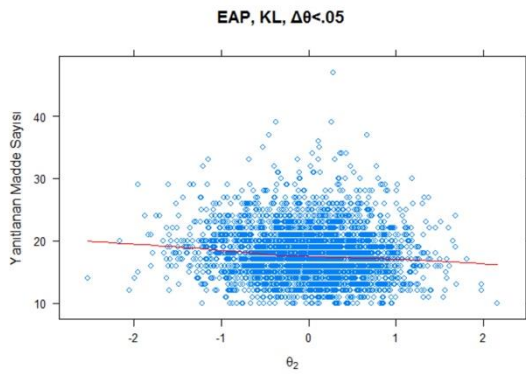
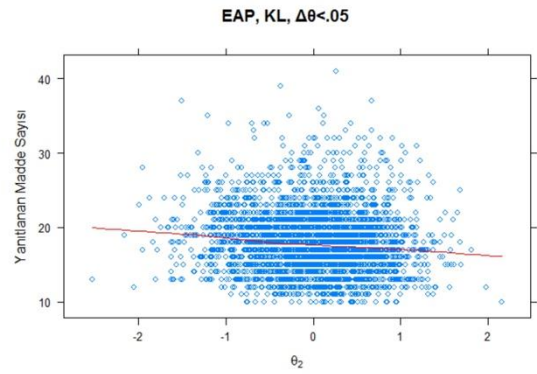
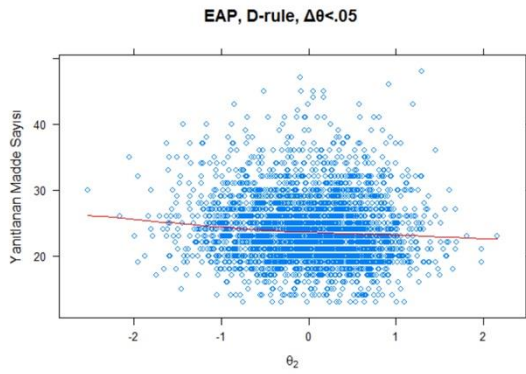
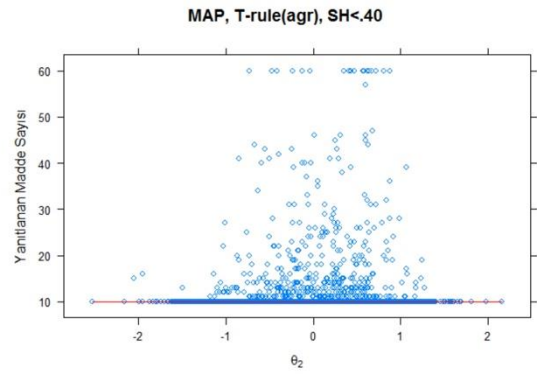
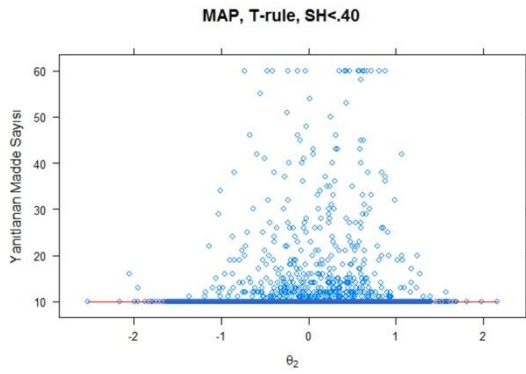
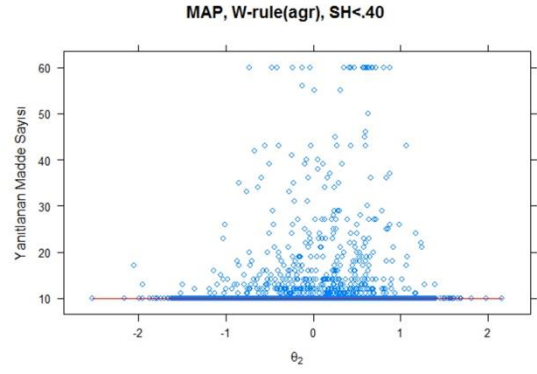
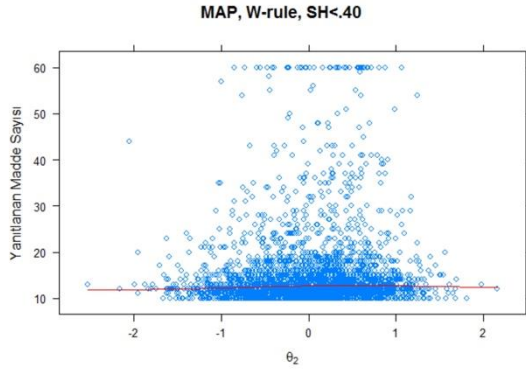


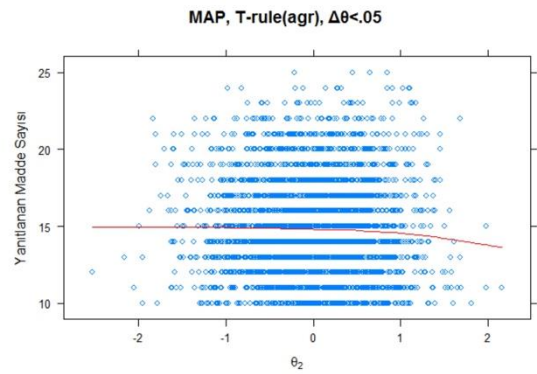
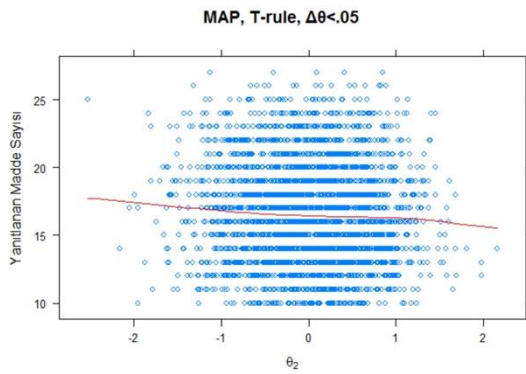
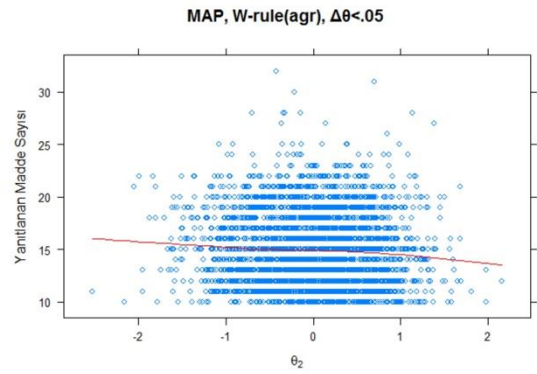
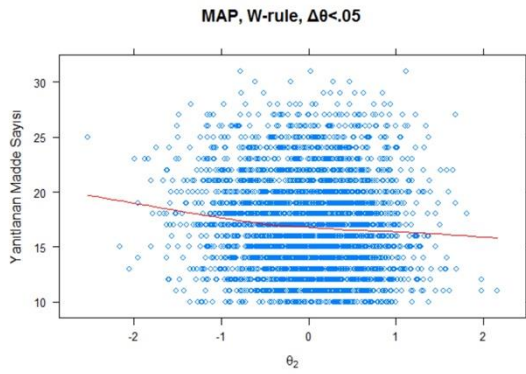
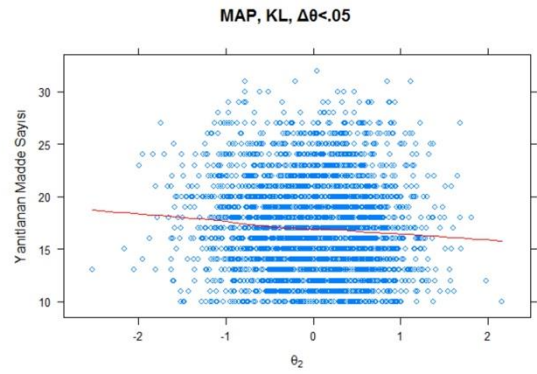
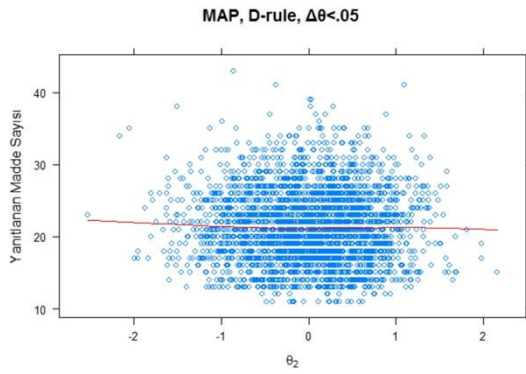
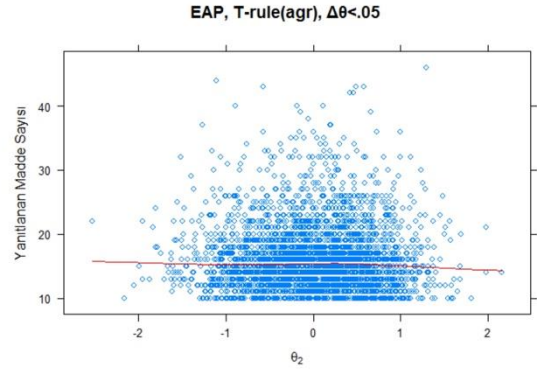
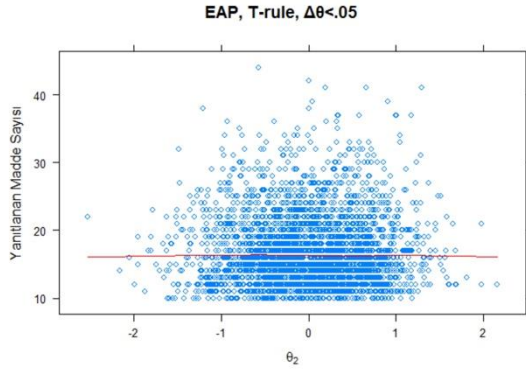




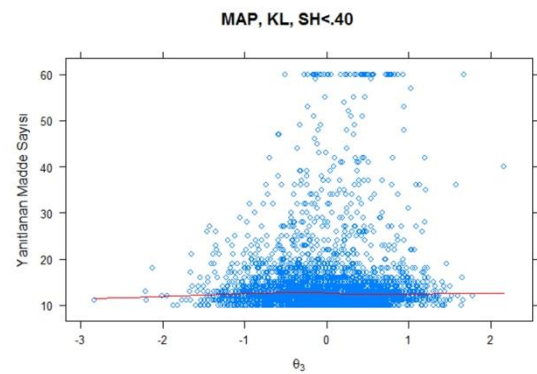
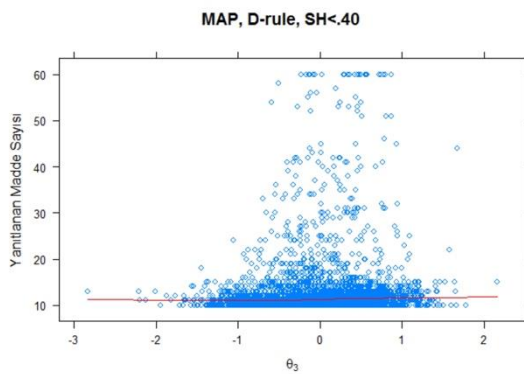
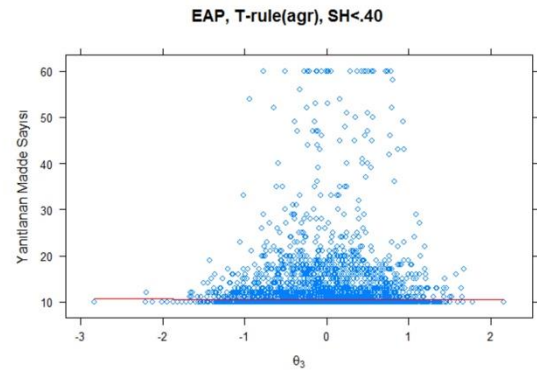
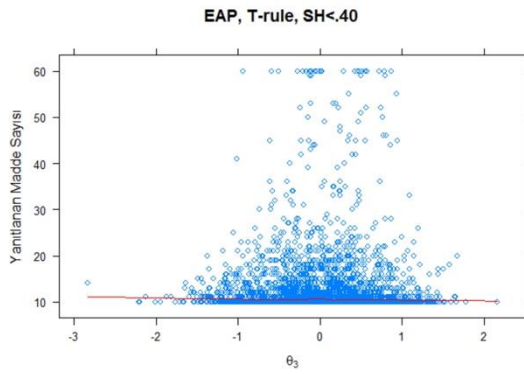
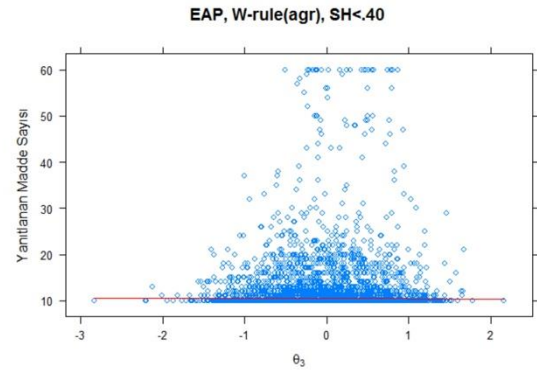
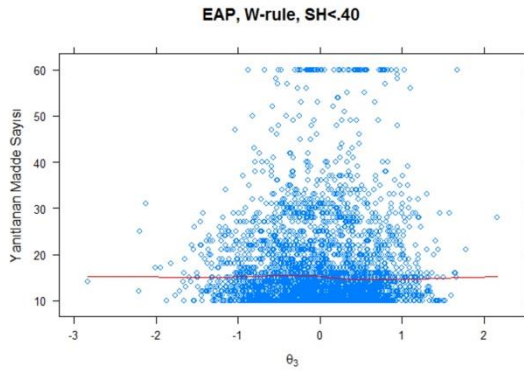
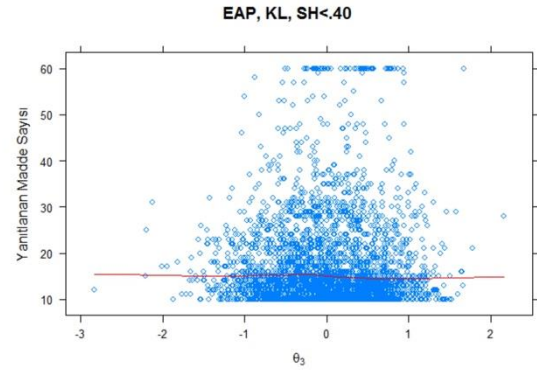
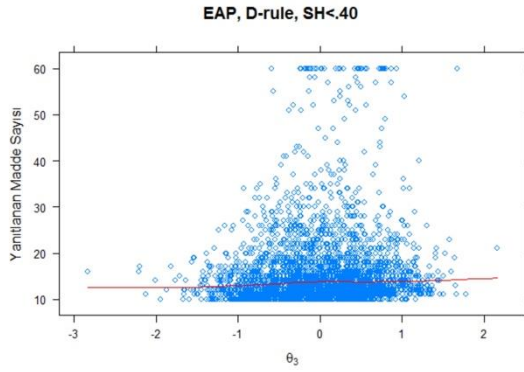
EK 4.2 GRAMER BECERİSİ İÇİN ELDE EDİLEN YANITLANAN MADDE SAYISI – THETA GRAFİKLERİ

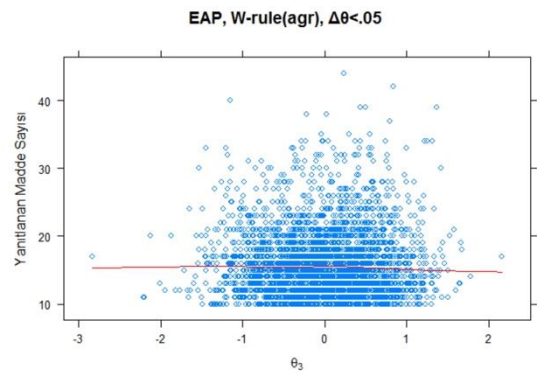
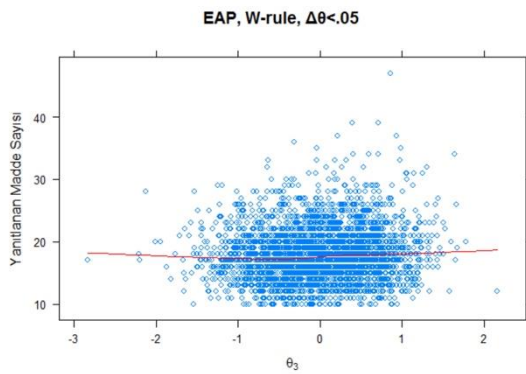
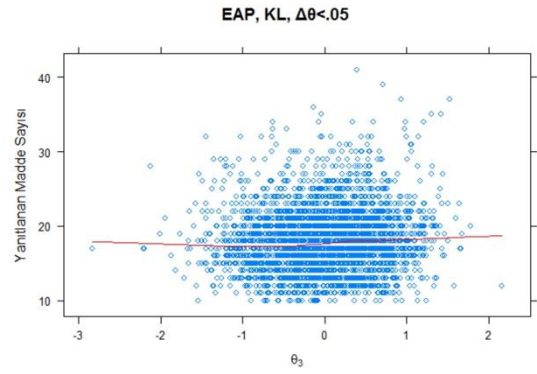
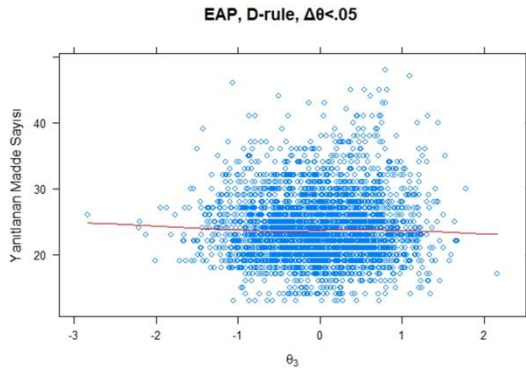
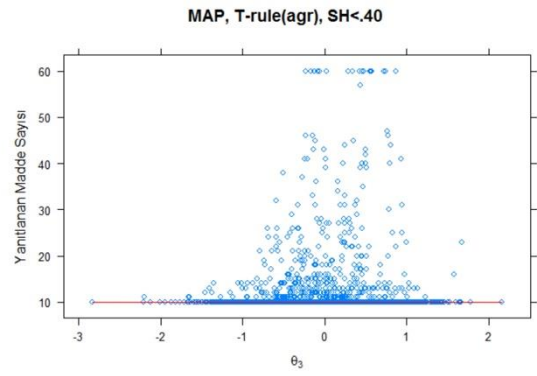
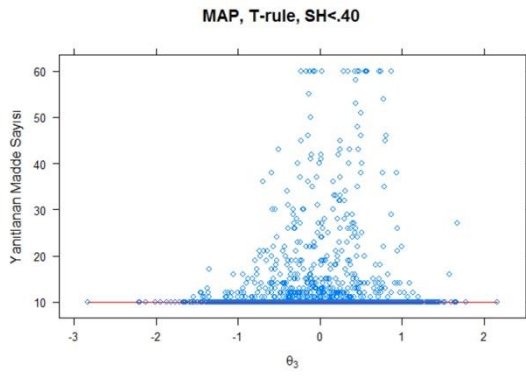
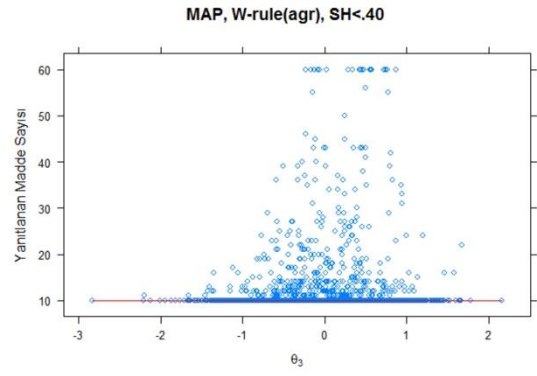
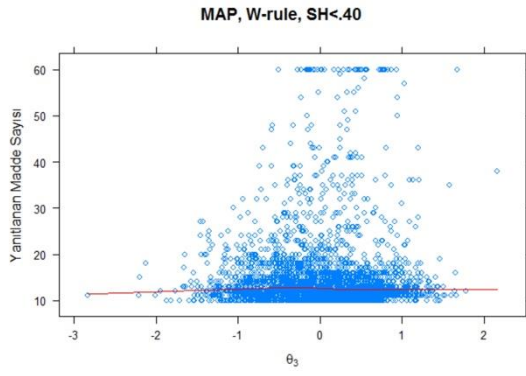


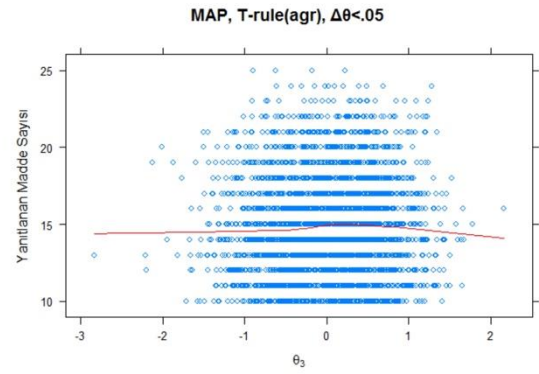
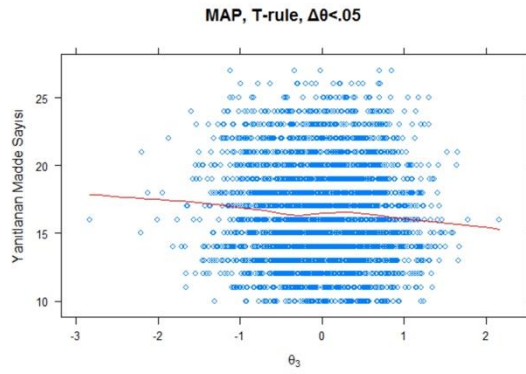
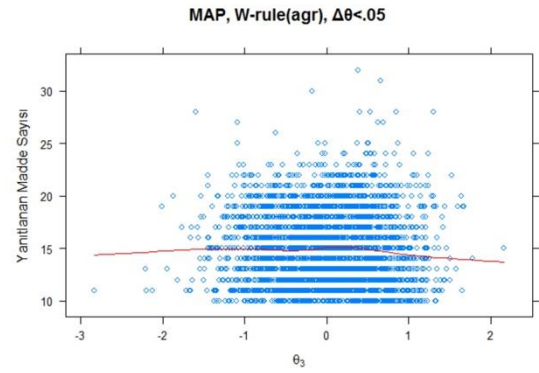
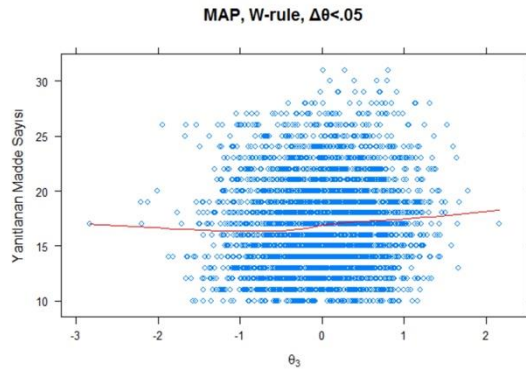
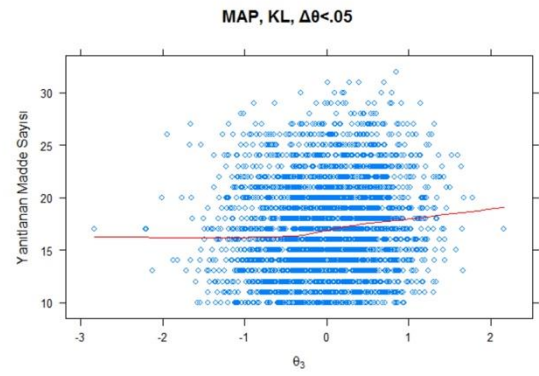
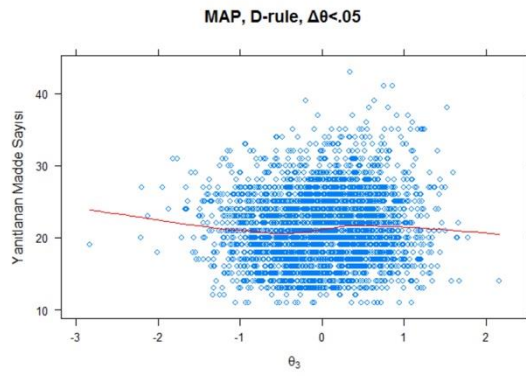
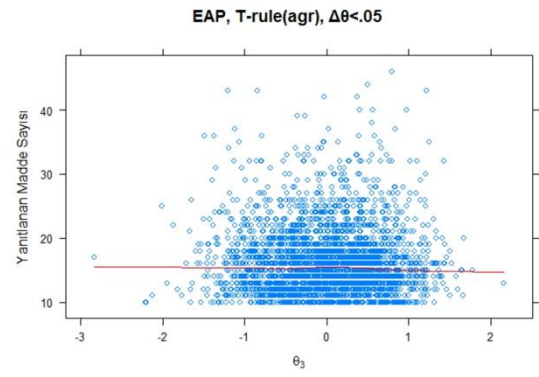
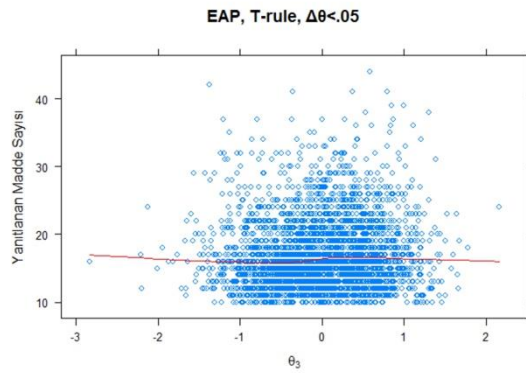




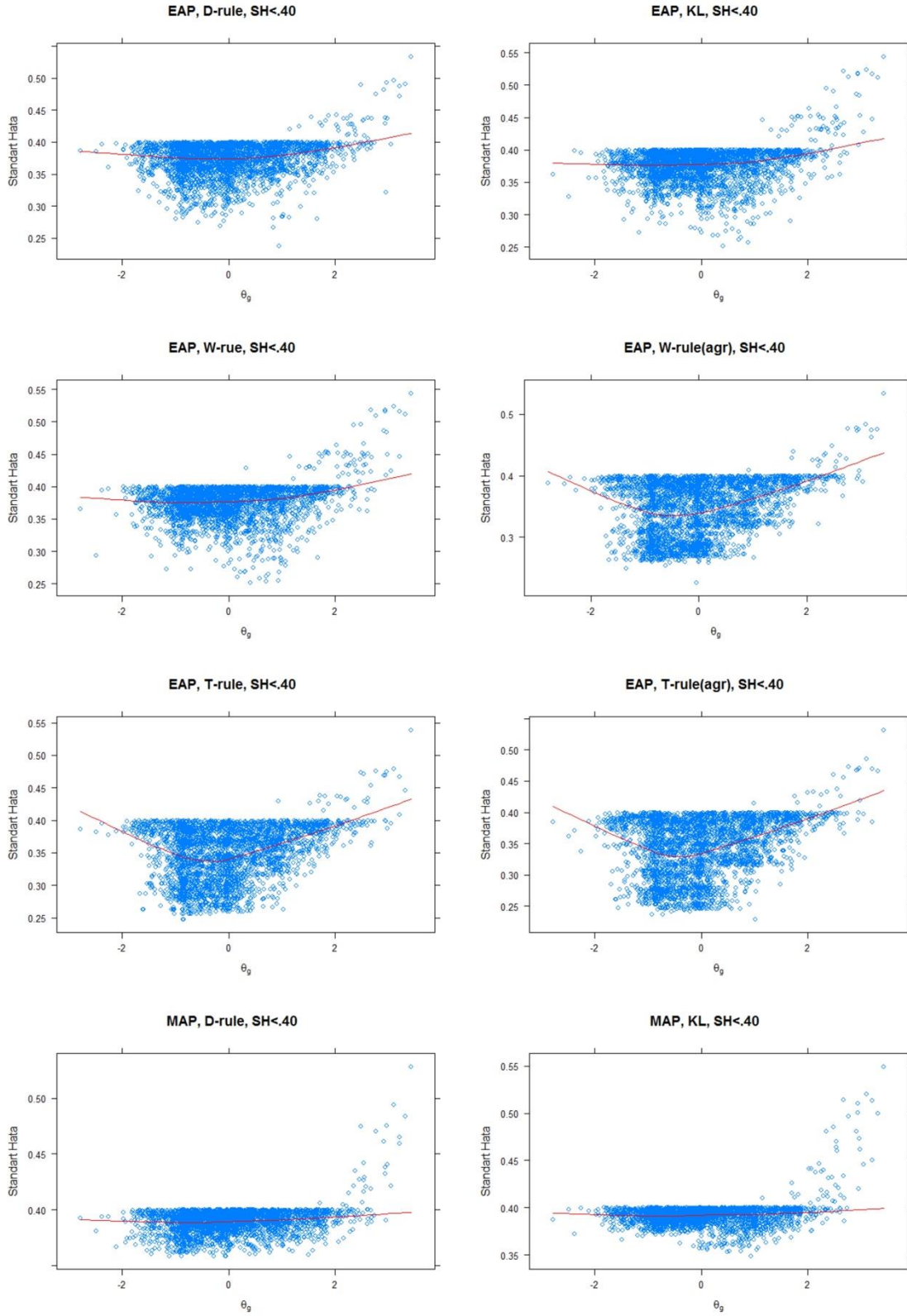
EK 4.3 KELİME BİLGİSİ BECERİSİ İÇİN ELDE EDİLEN YANITLANAN MADDE SAYISI – THETA GRAFİKLERİ

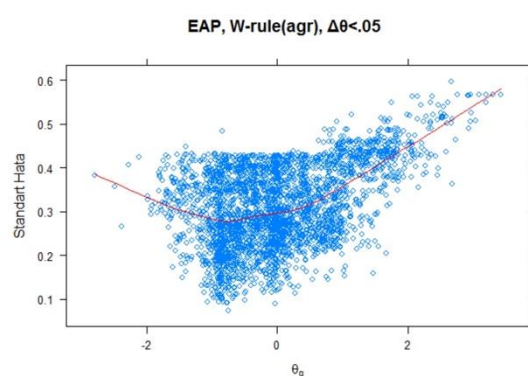
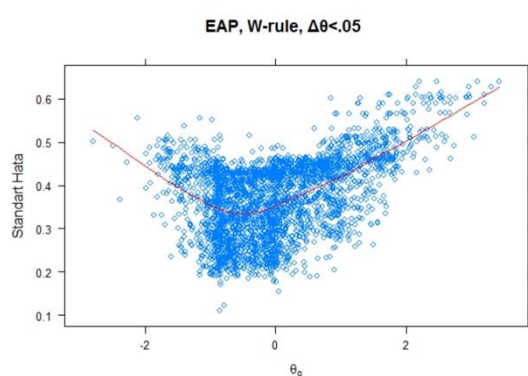
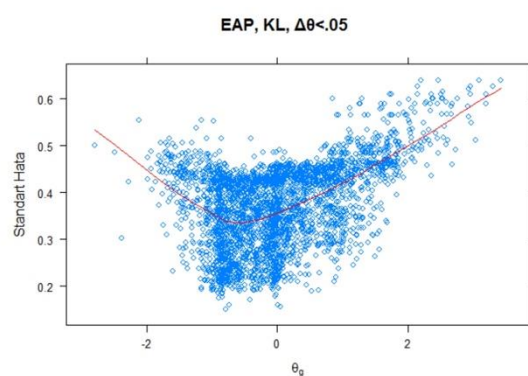
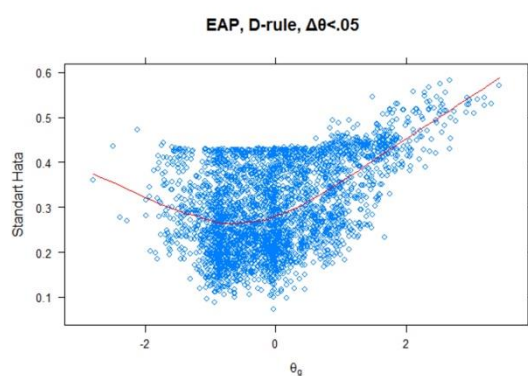
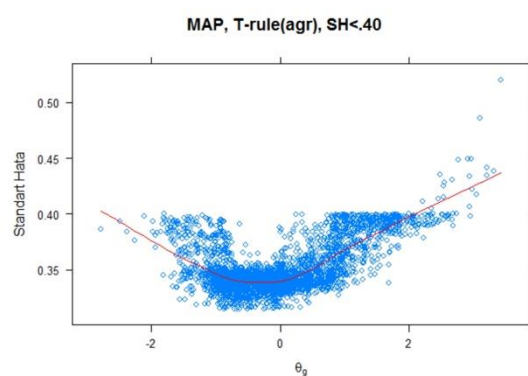
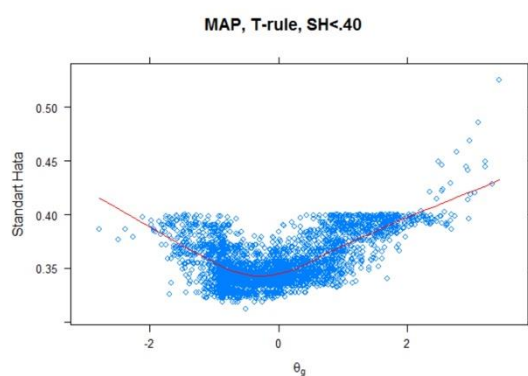
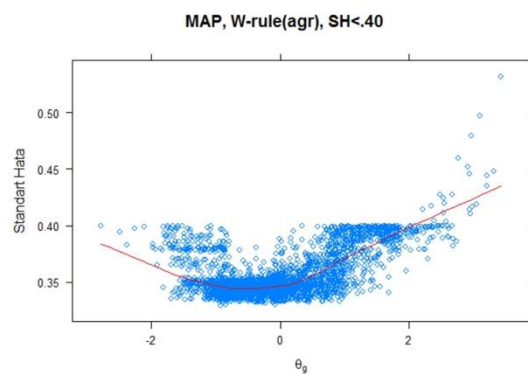
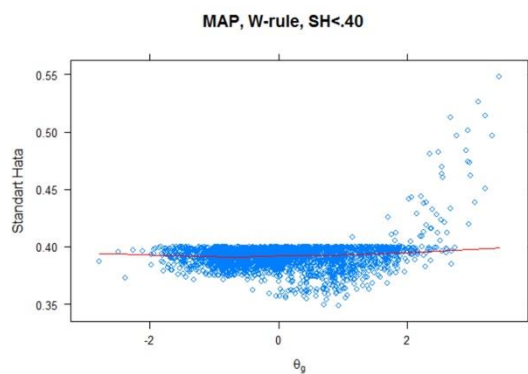


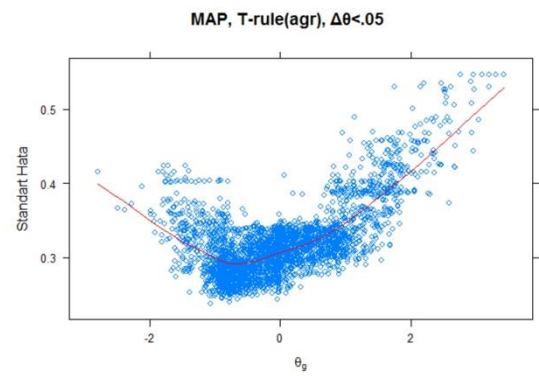
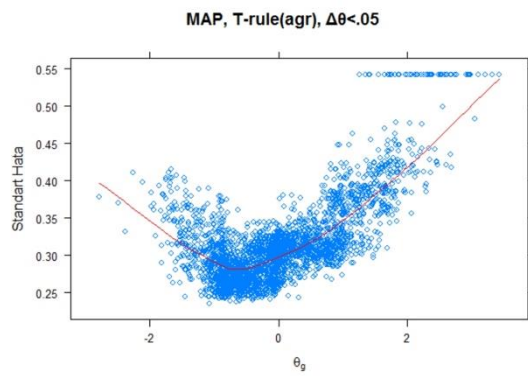
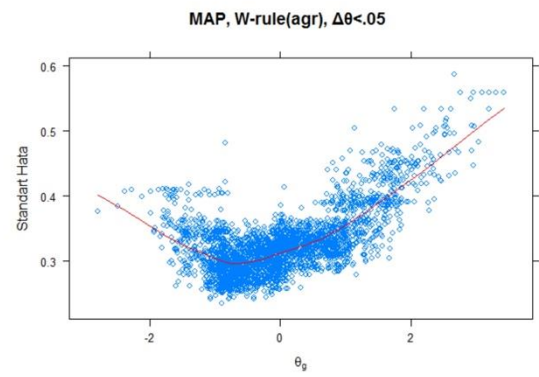
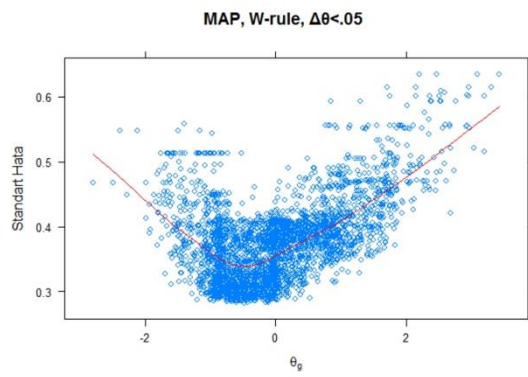
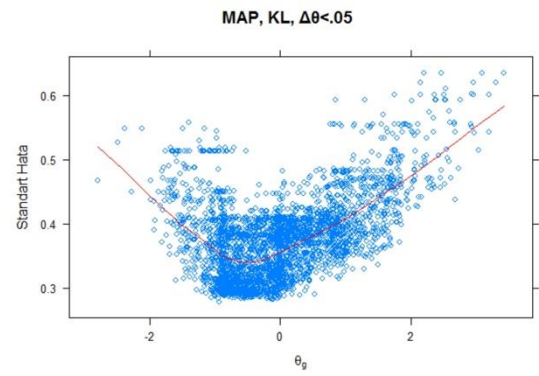
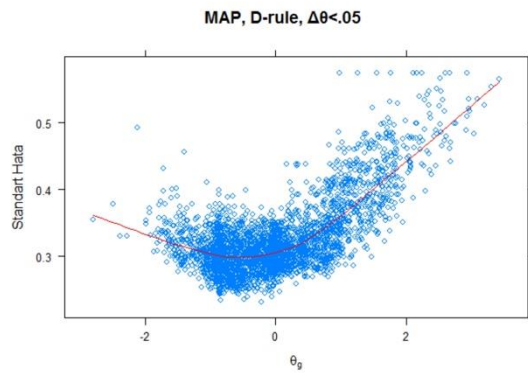
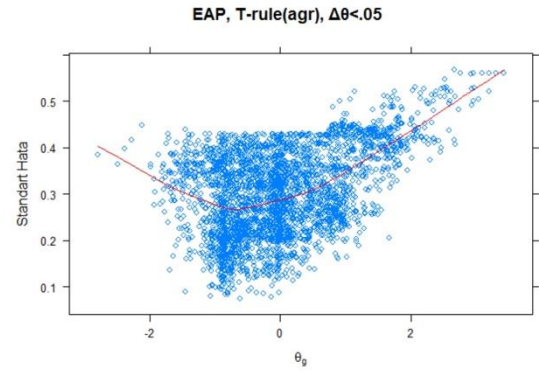
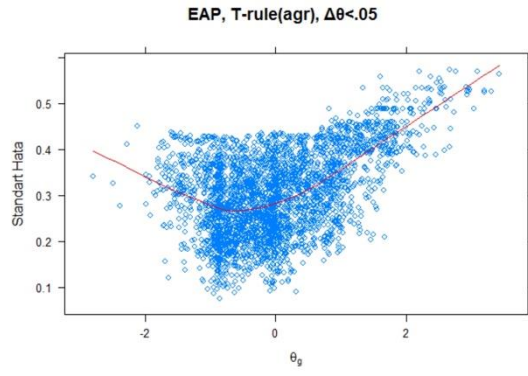


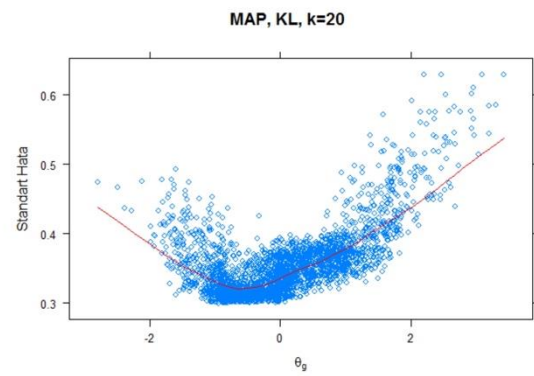
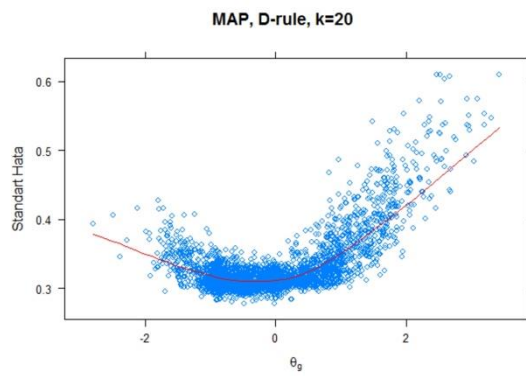
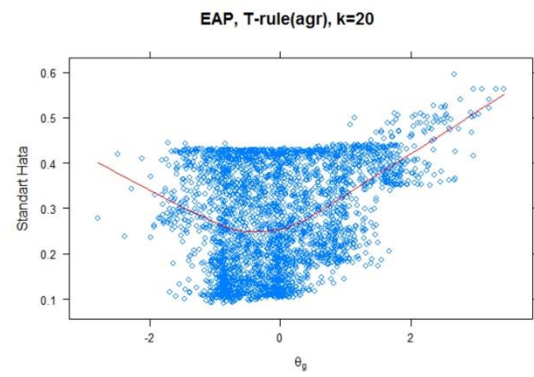
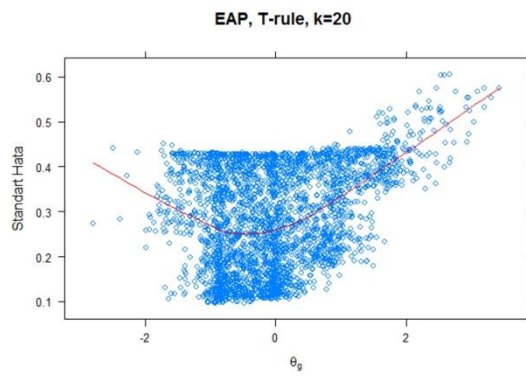
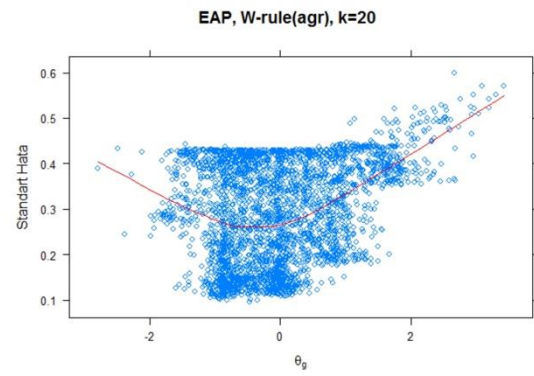
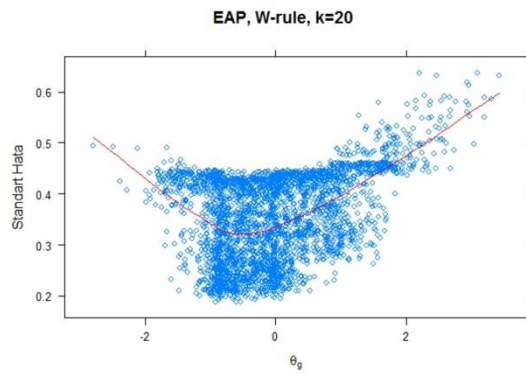
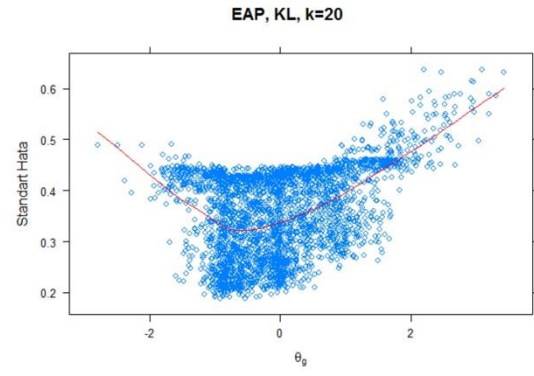
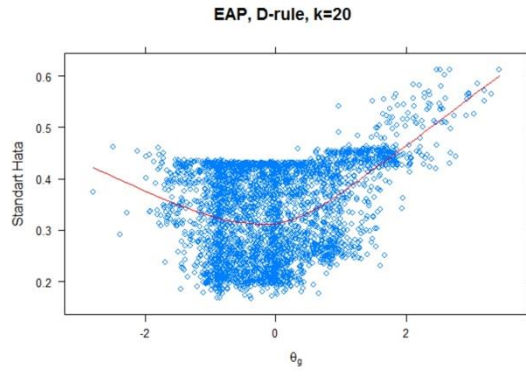


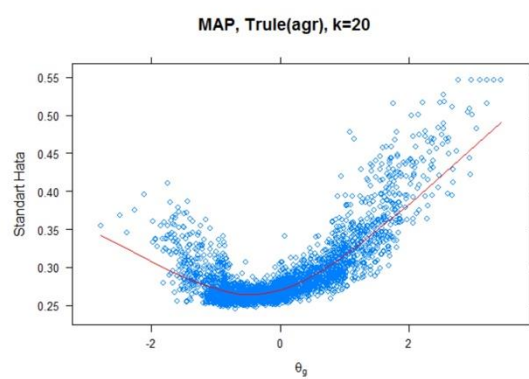
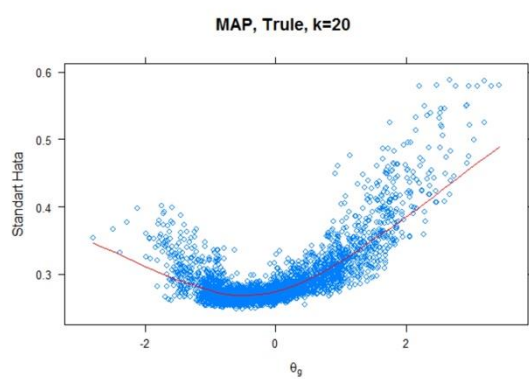
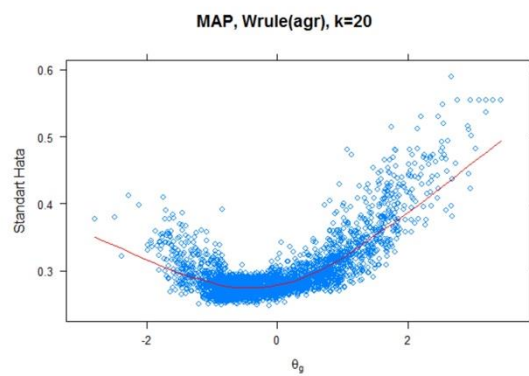
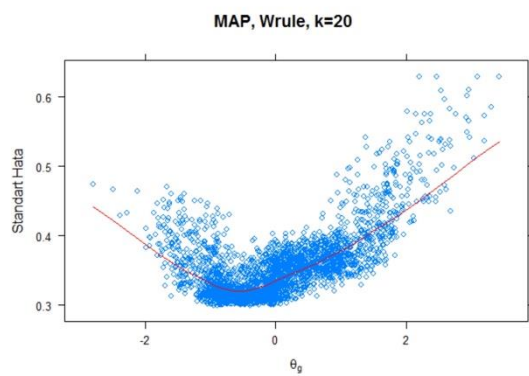
EK 4.4 GENEL YETENEK İÇİN ELDE EDİLEN STANDART HATA – THETA GRAFİKLERİ



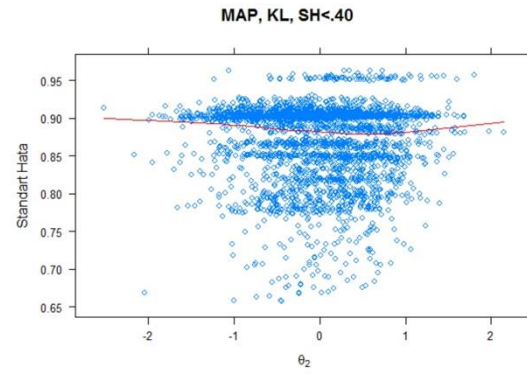
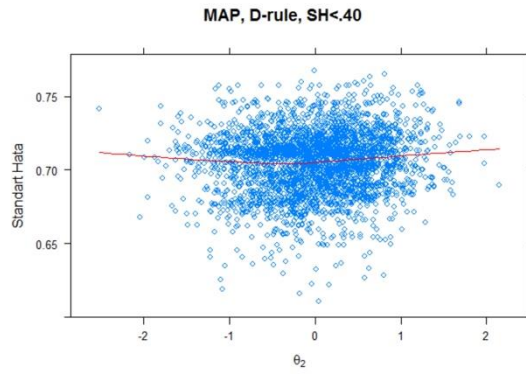
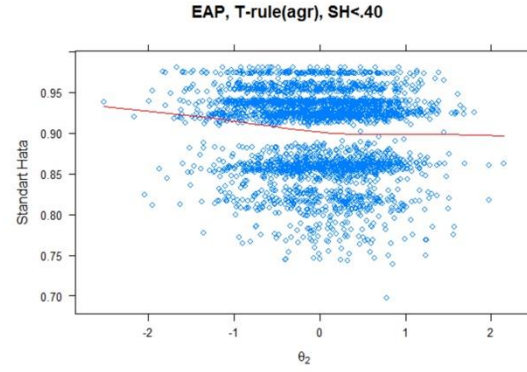
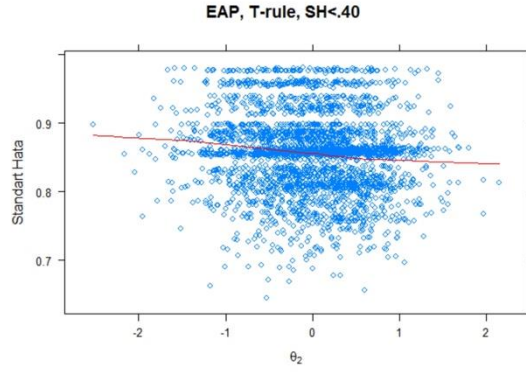
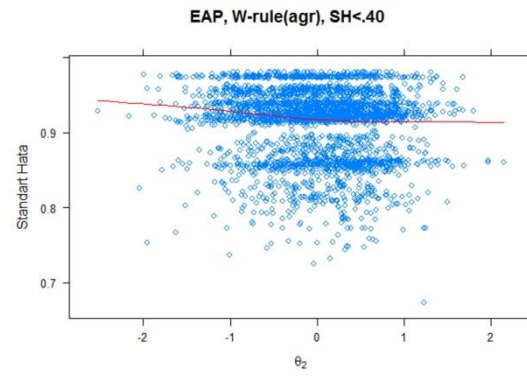
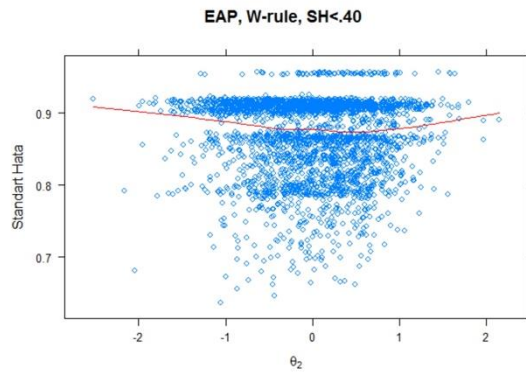
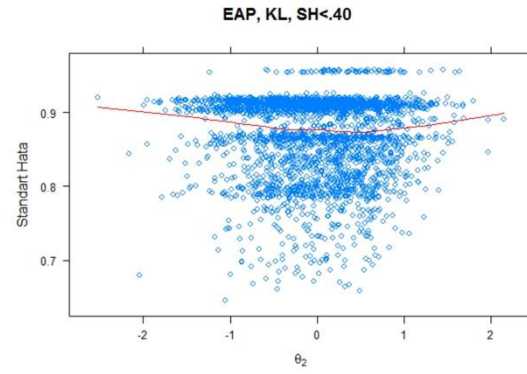
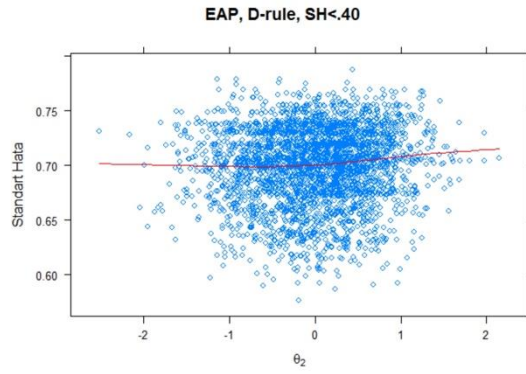


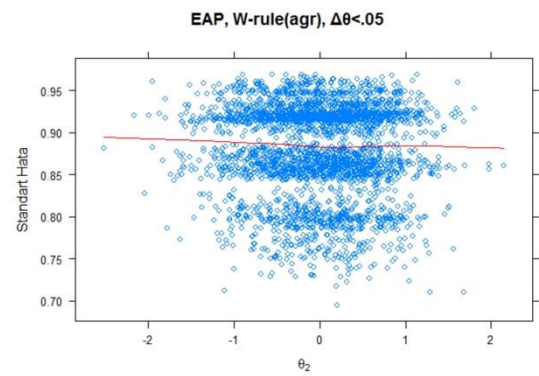
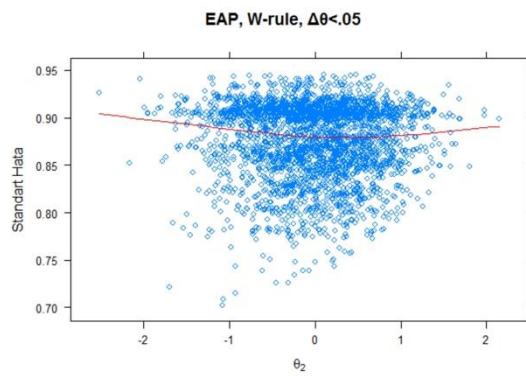
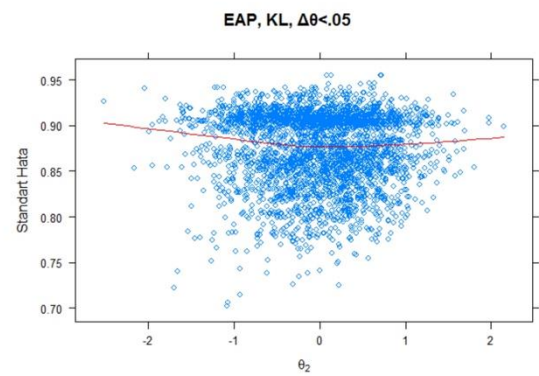
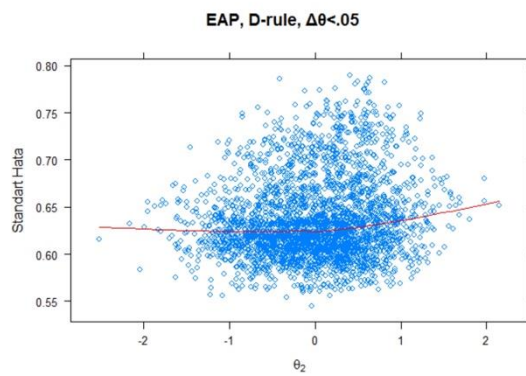
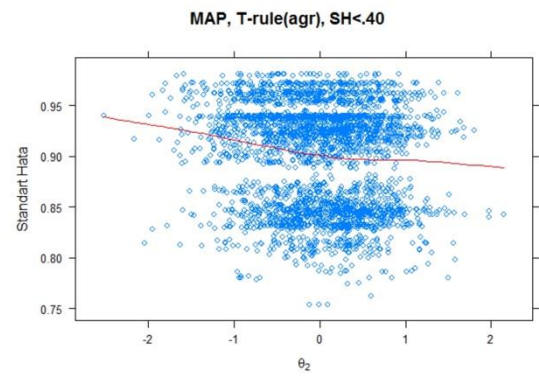
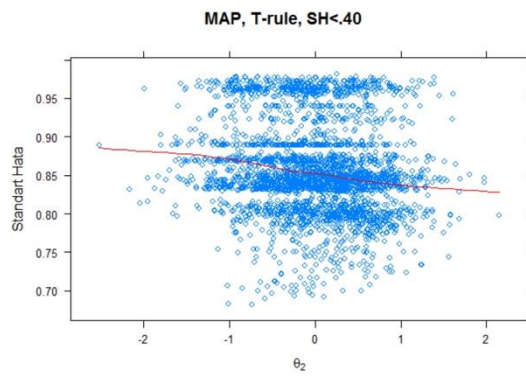
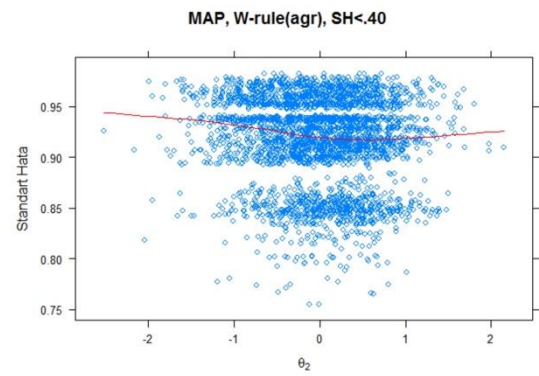
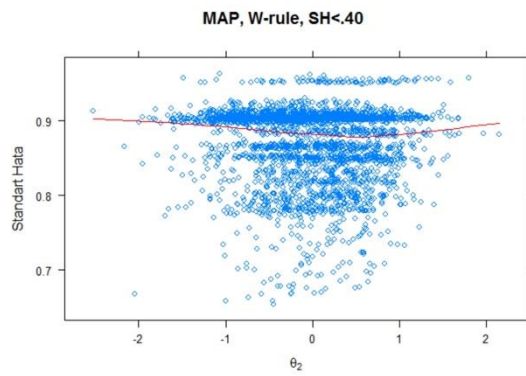


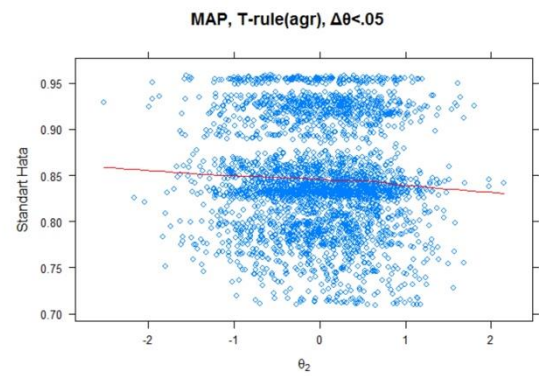
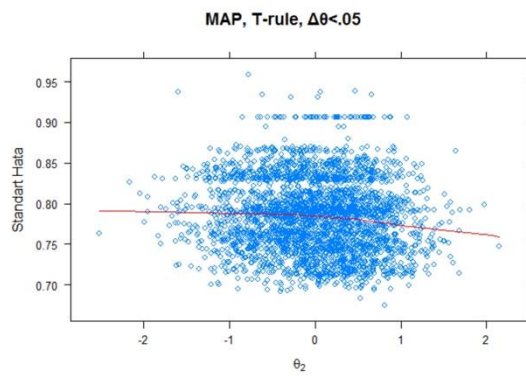
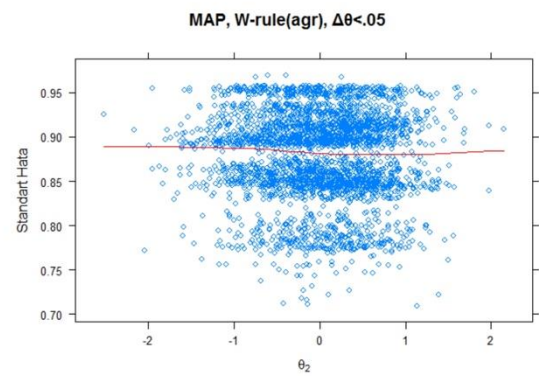
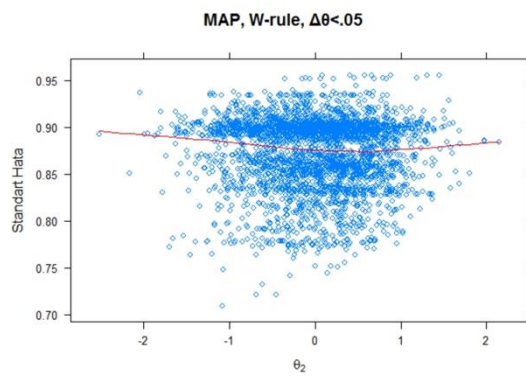
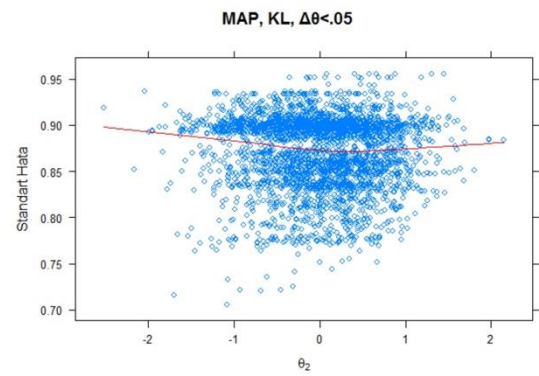
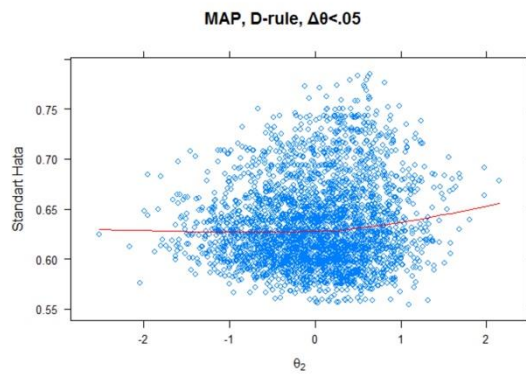
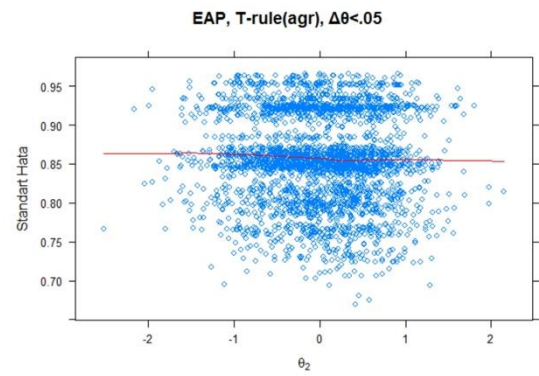
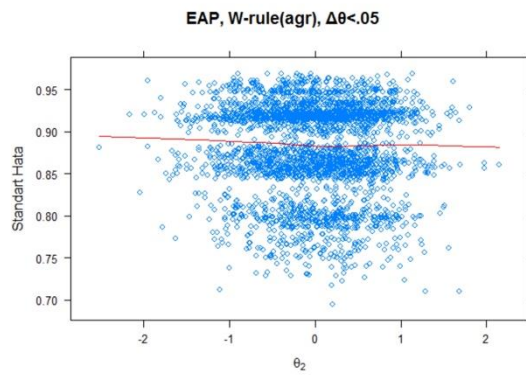


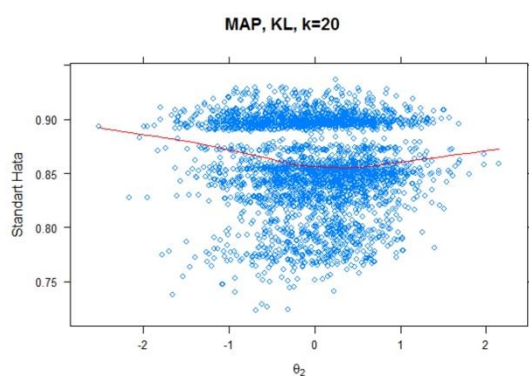
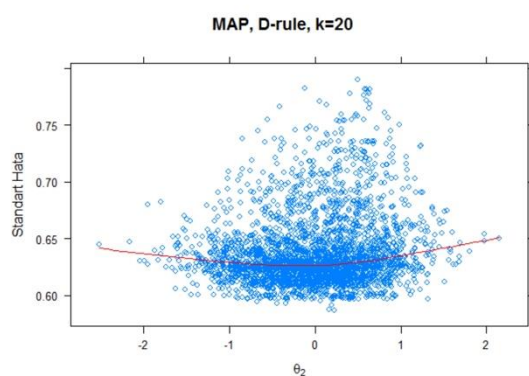
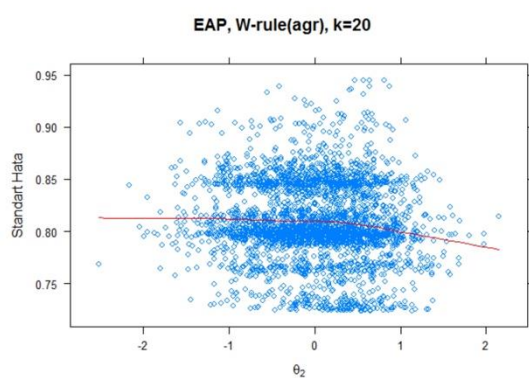
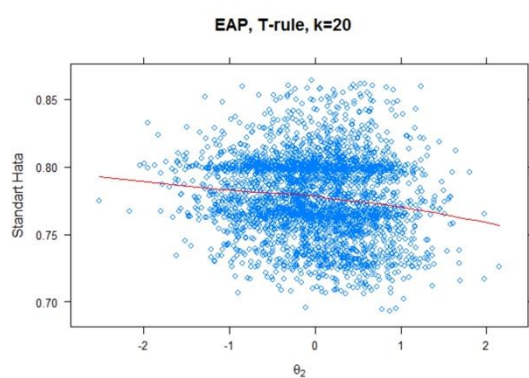
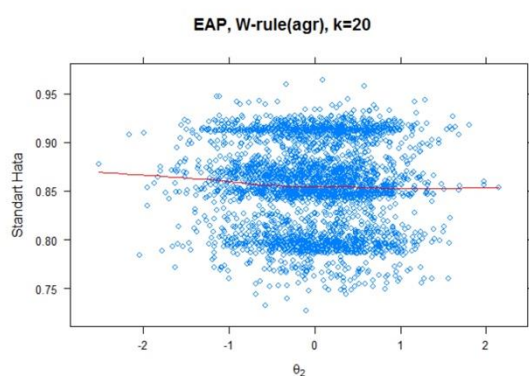
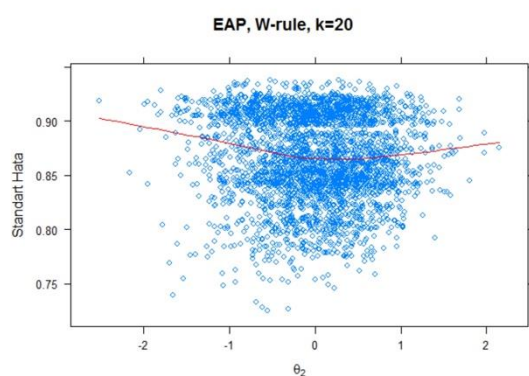
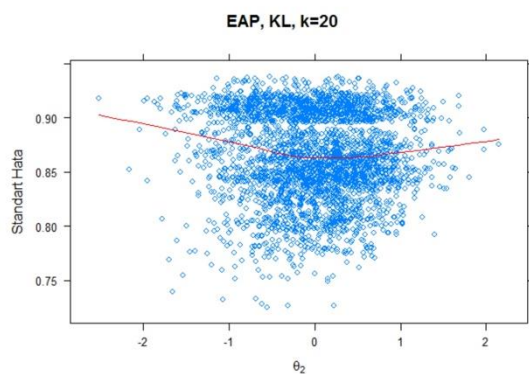
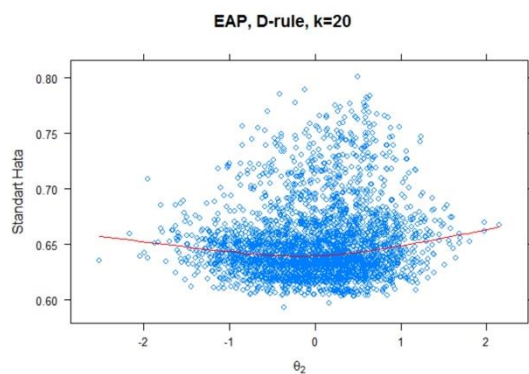


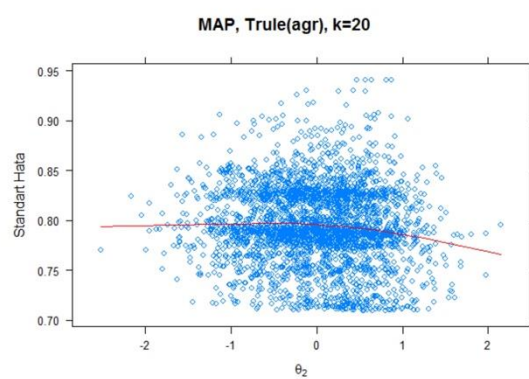
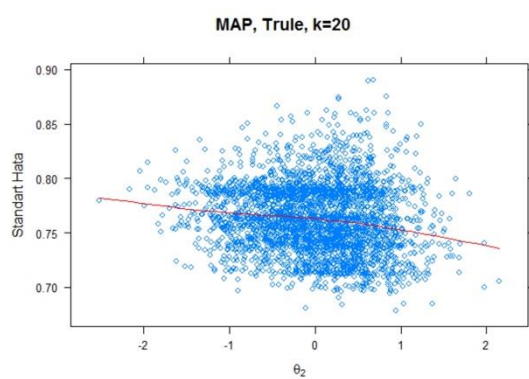
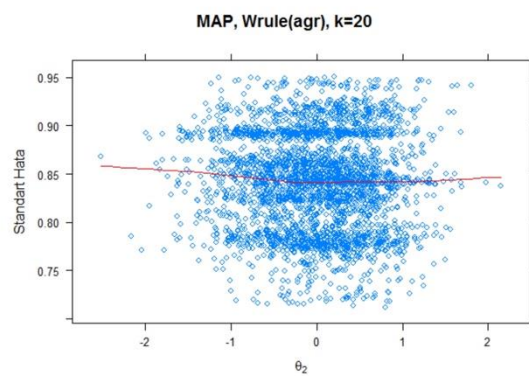
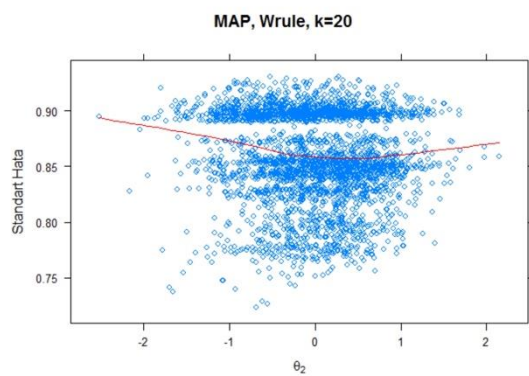
EK 4.5 GRAMER İÇİN ELDE EDİLEN STANDART HATA – THETA GRAFİKLERİ



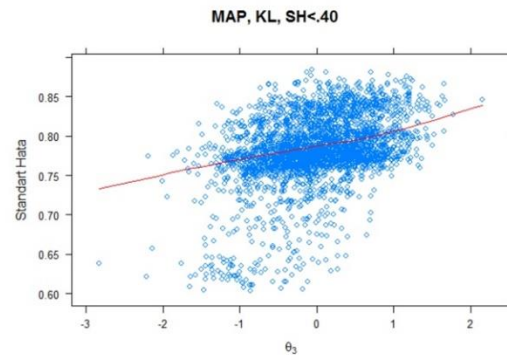
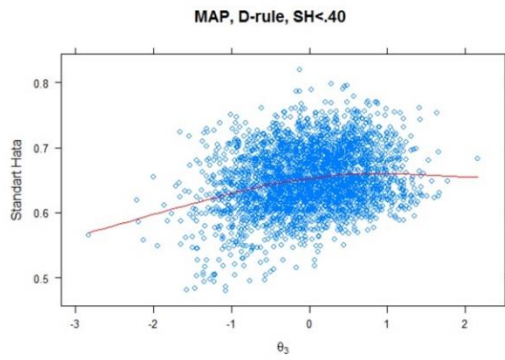
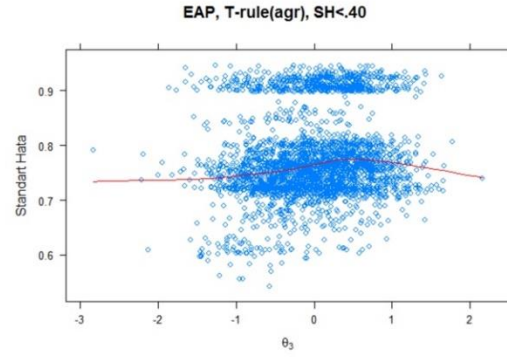
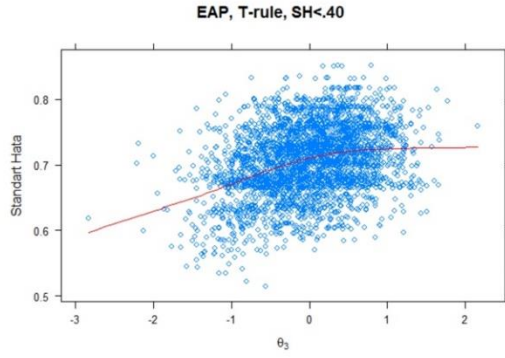
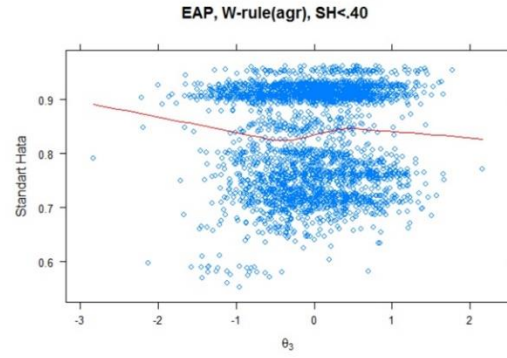
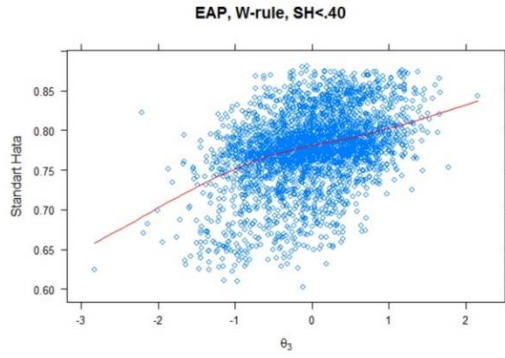
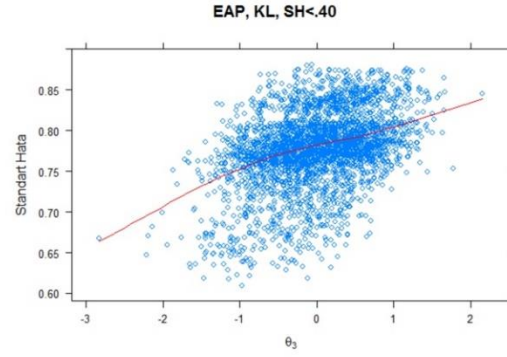
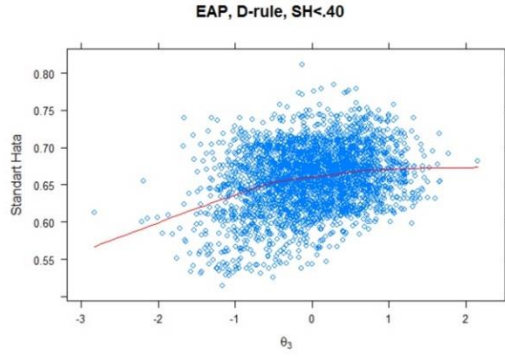


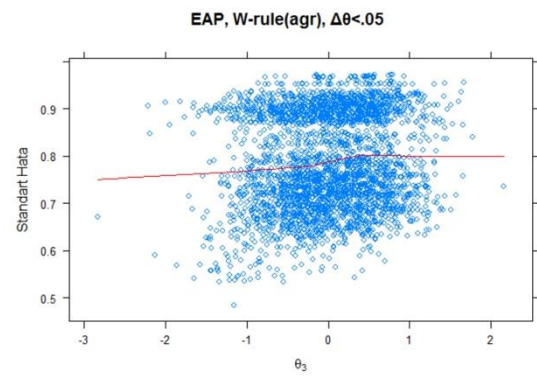
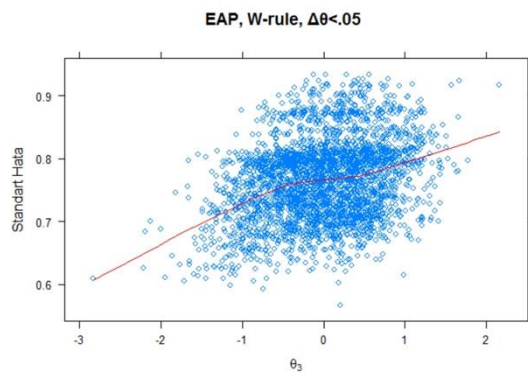
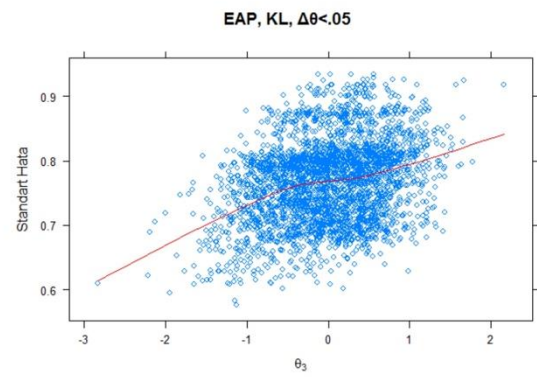
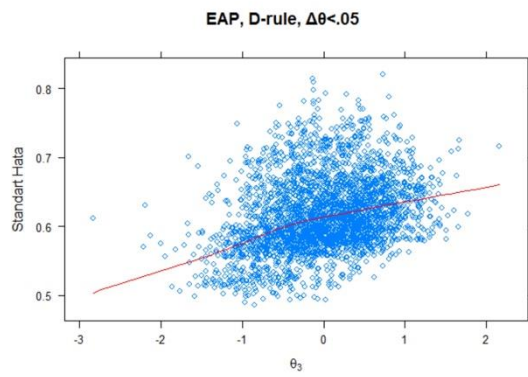
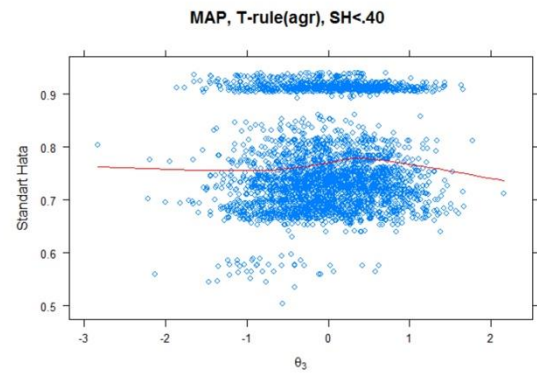
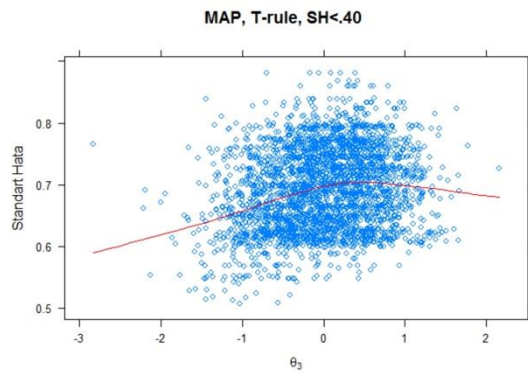
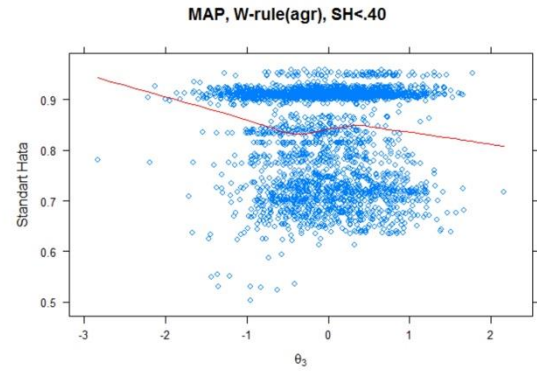
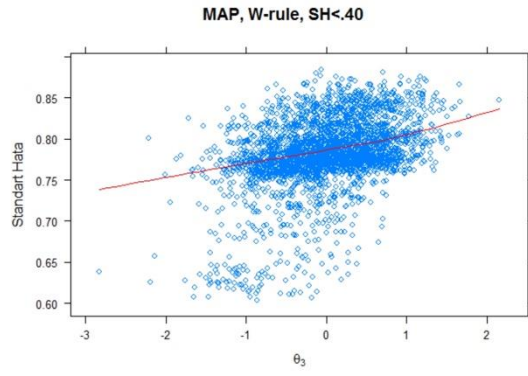


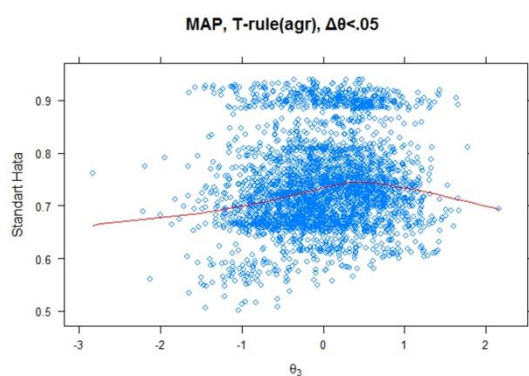
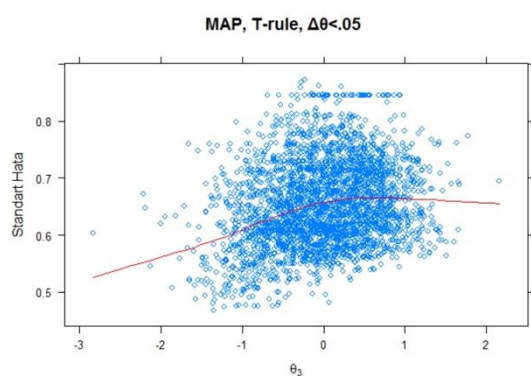
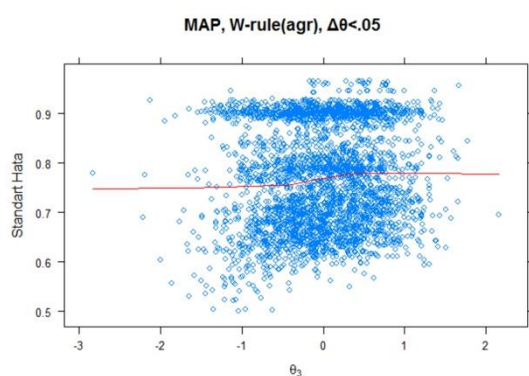
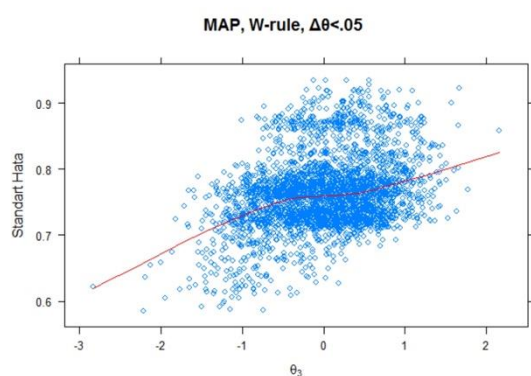
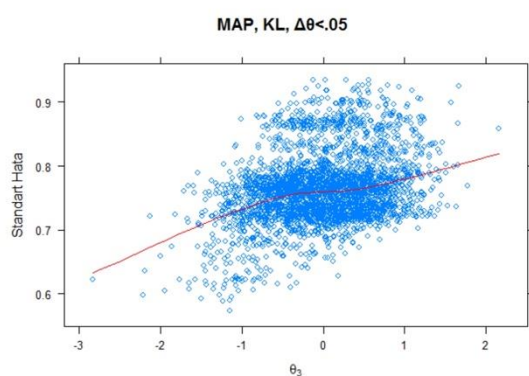
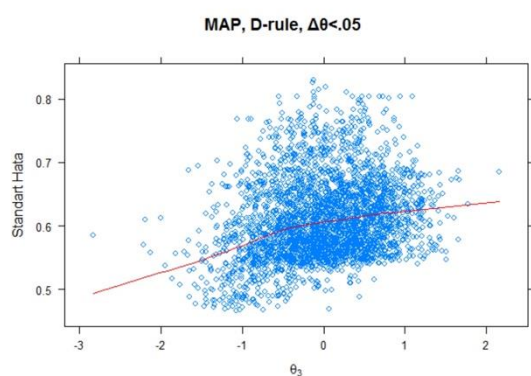
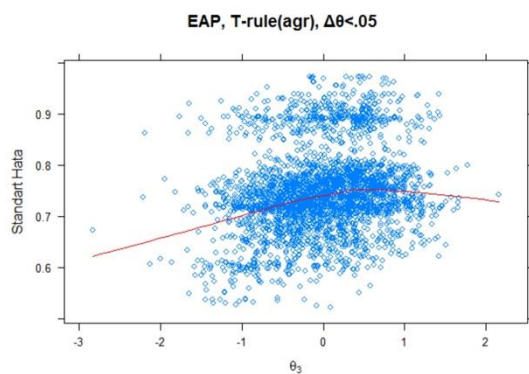
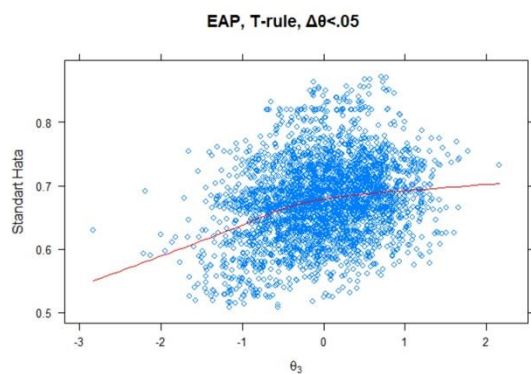


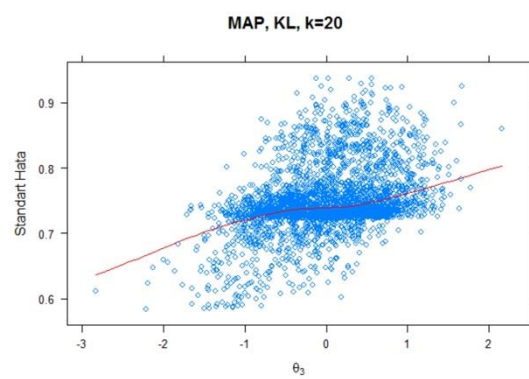
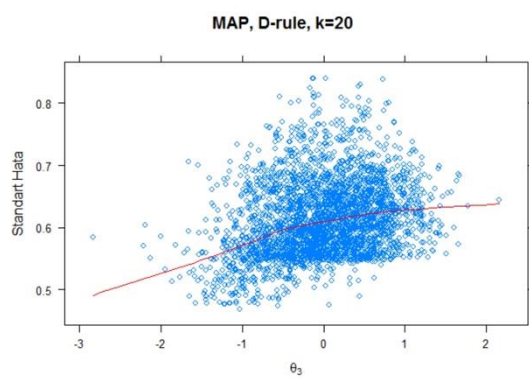
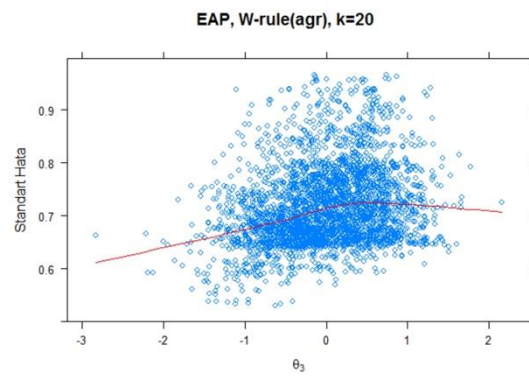
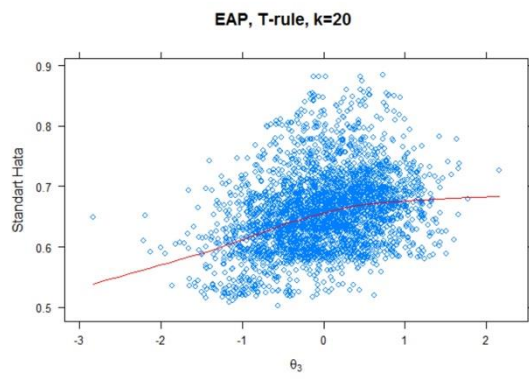
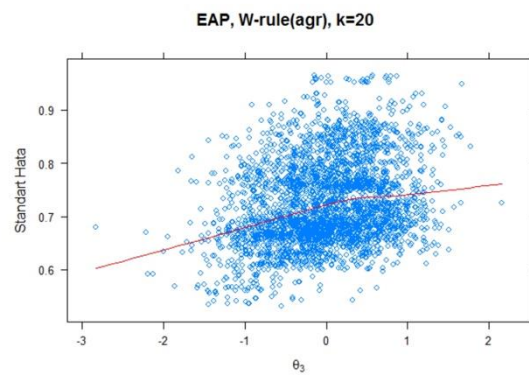
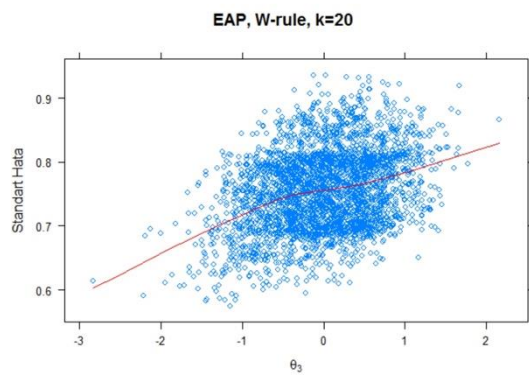
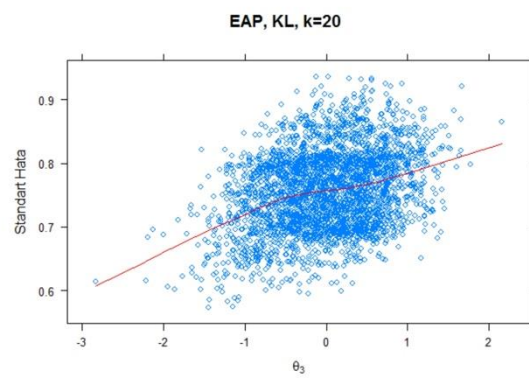
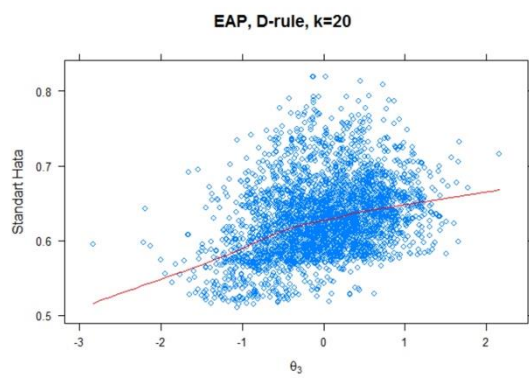


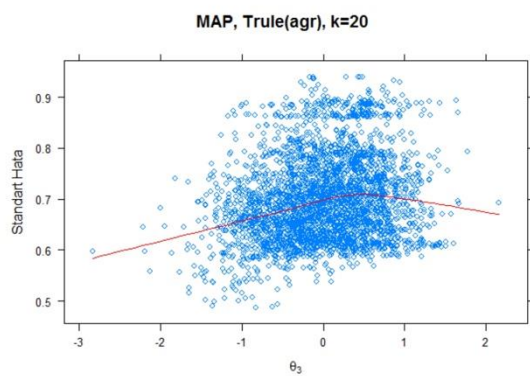
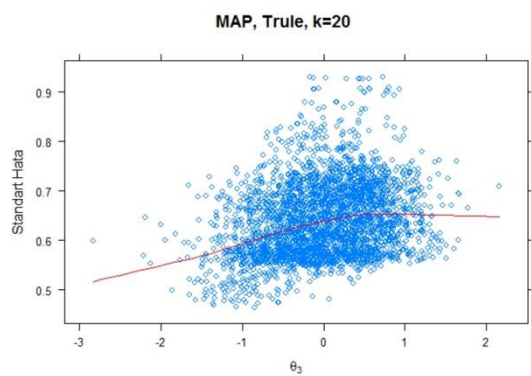
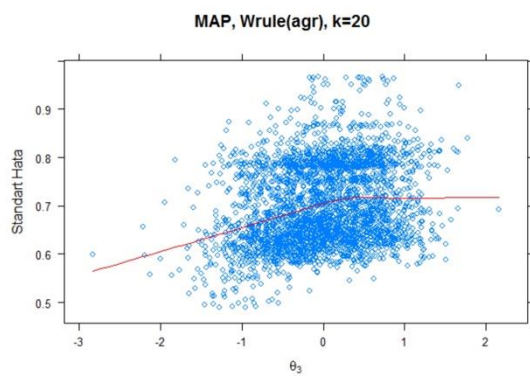
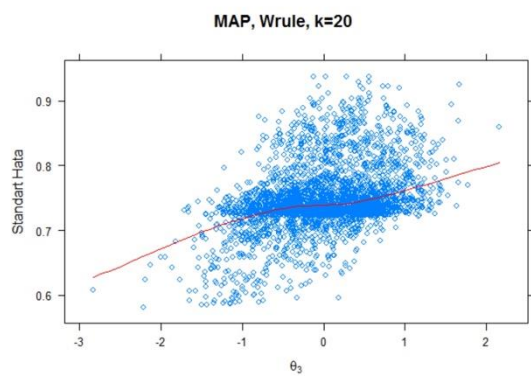
EK 4.6 KELİME BİLGİSİ İÇİN ELDE EDİLEN STANDART HATA – THETA GRAFİKLERİ



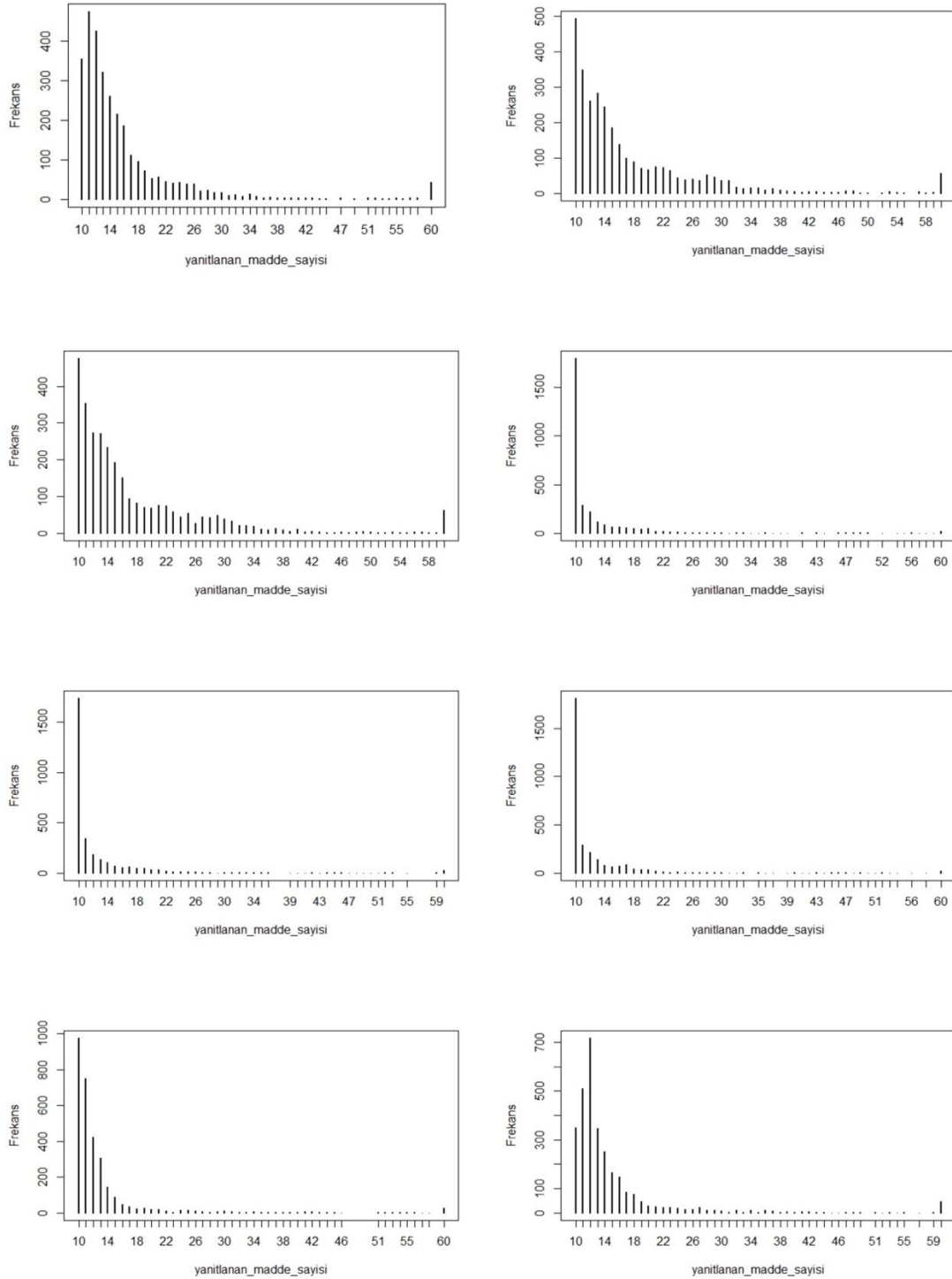




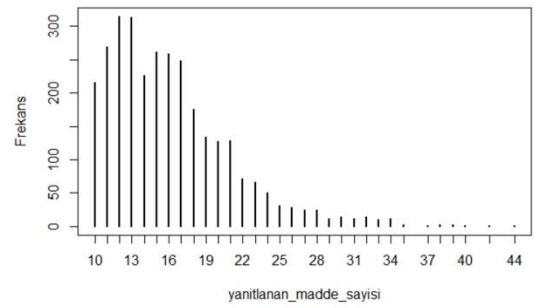
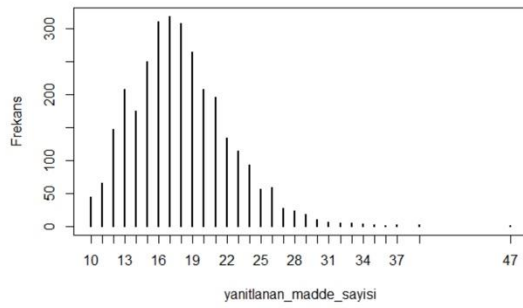
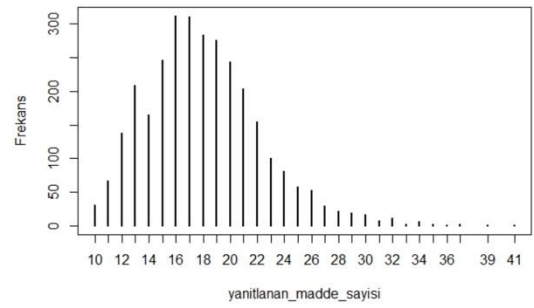
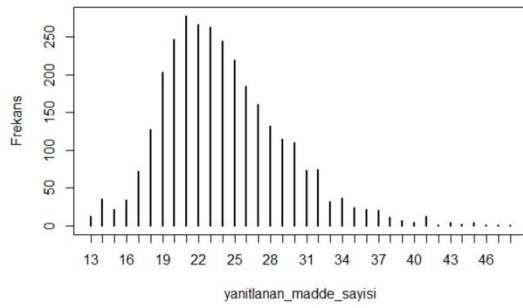
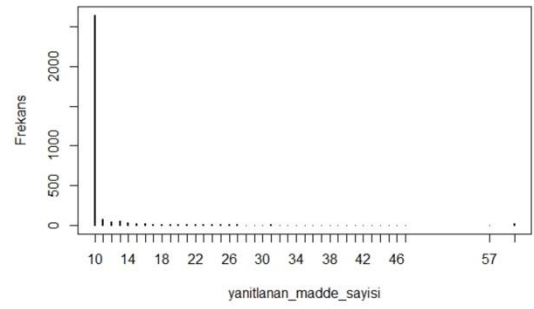
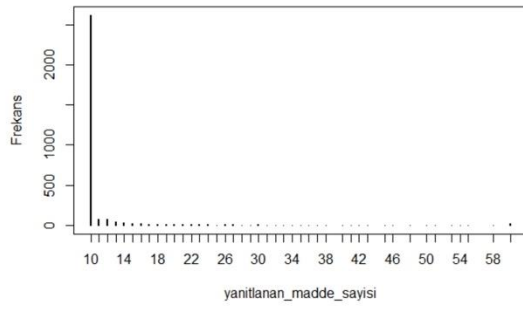
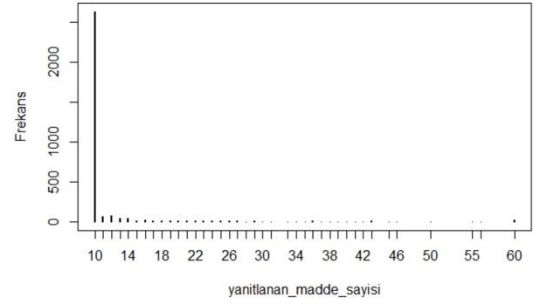
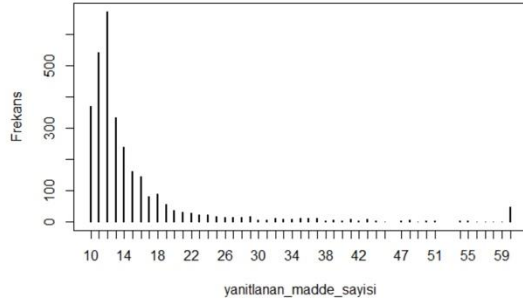


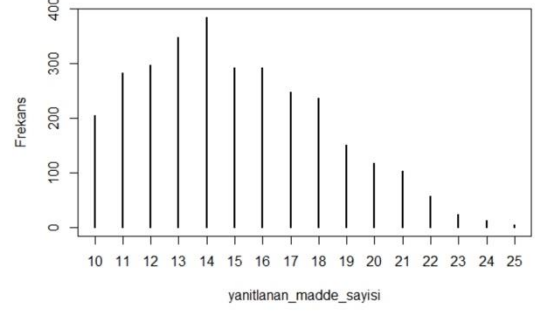
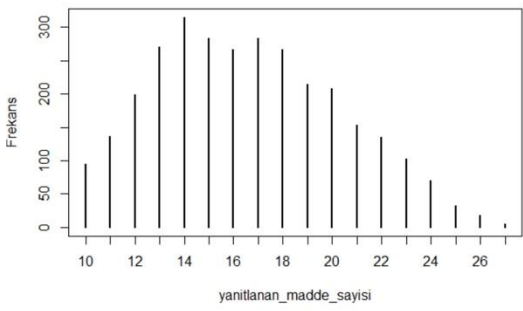
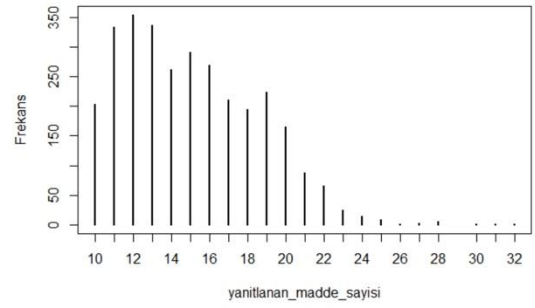
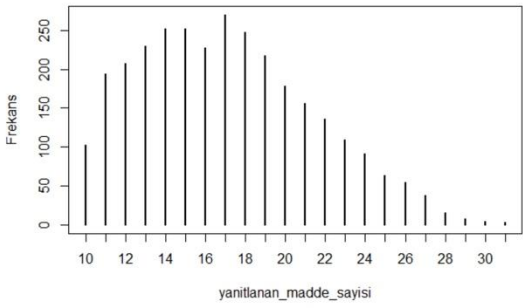
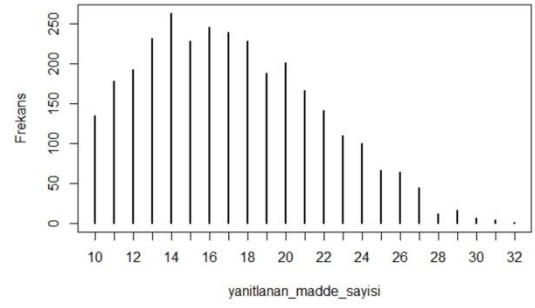
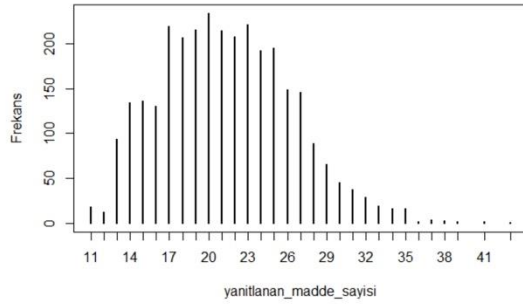
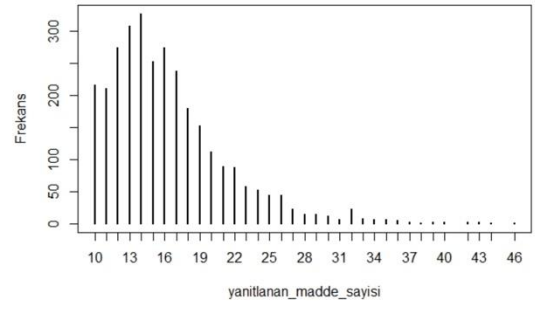
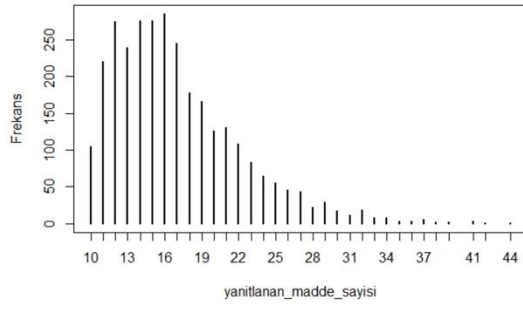


EK 4.7 DEĞİŞKEN MADDE SAYISINA DAYALI SONLANDIRMA KURALLARININ KULLANILDIĞI KOŞULLARDA YANITLANAN MADDE SAYILARININ FREKANS DAĞILIM GRAFİKLERİ³



³ Bu bölümde yer alan grafiklerin ilk yarısı standart hataya dayalı, ikinci yarısı ise theta değişimine dayalı sonlandırma kuralları için elde edilmiştir. Her bir yarıdaki grafikler (tablolarda olduğu gibi) sırasıyla D-rule, KL, W-rule, agr. W-rule, T-rule ve agr. T-rule madde seçim yöntemlerinin kullanıldığı koşullara aittir.





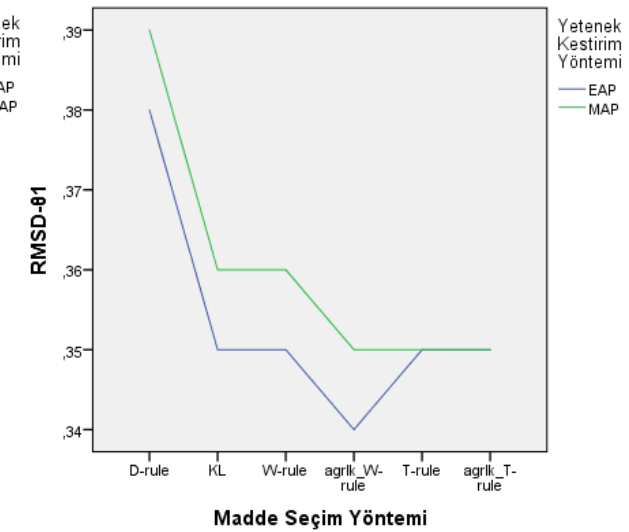
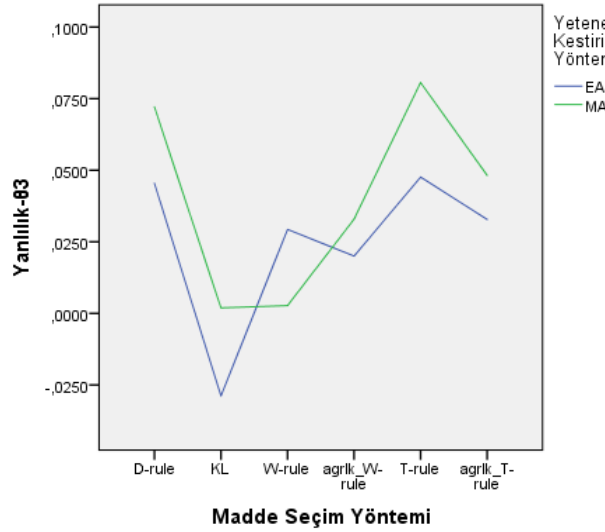
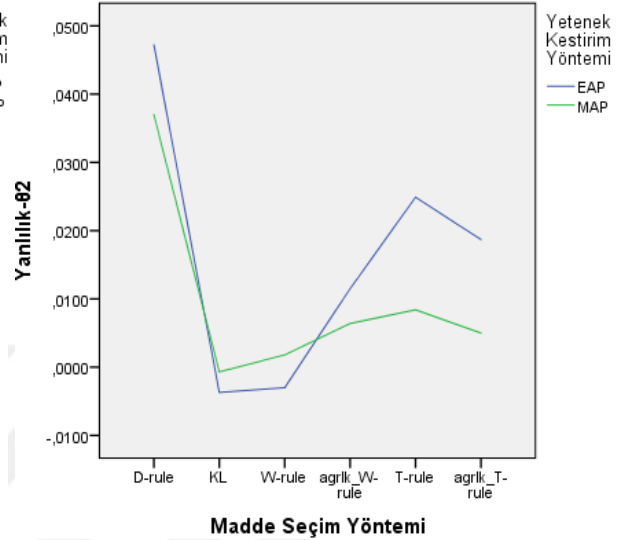
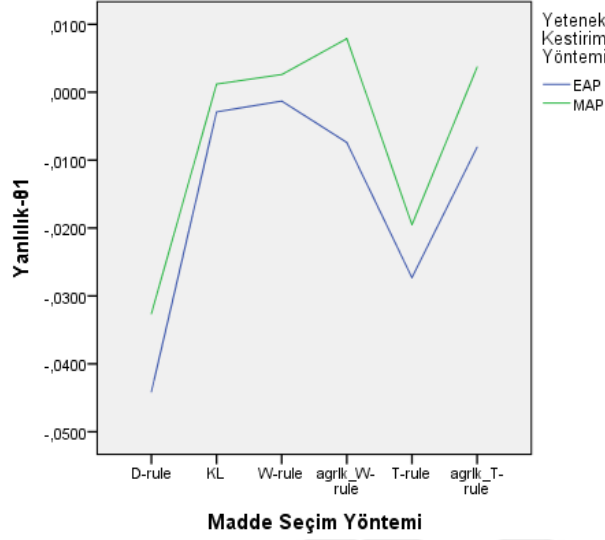
EK 5. GERÇEK-ZAMANLI ÇOK BOYUTLU BBT UYGULAMASINDA KULLANILAN R KODLARI

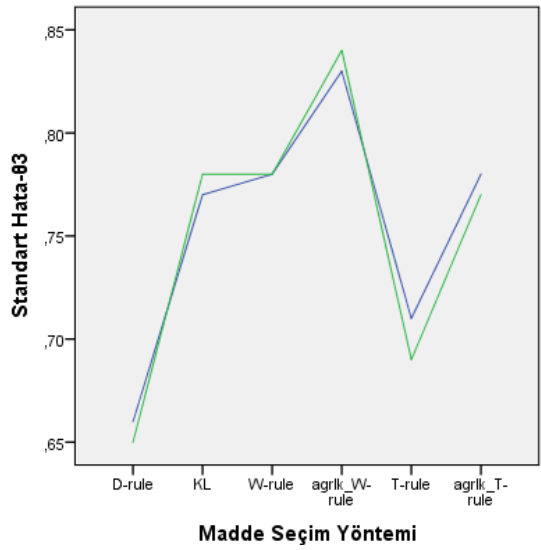
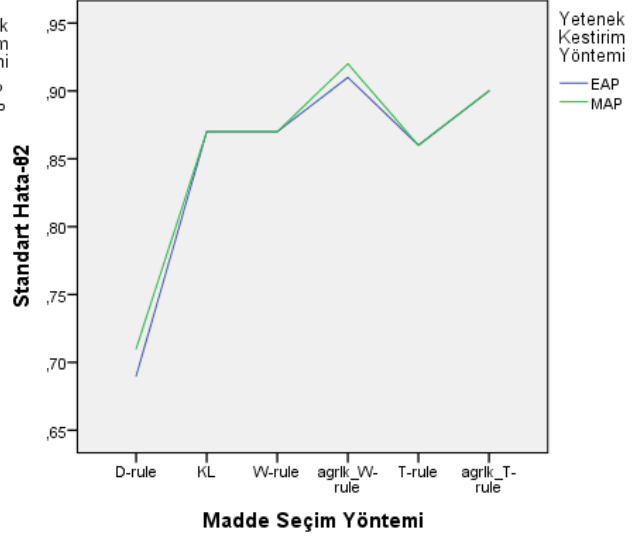
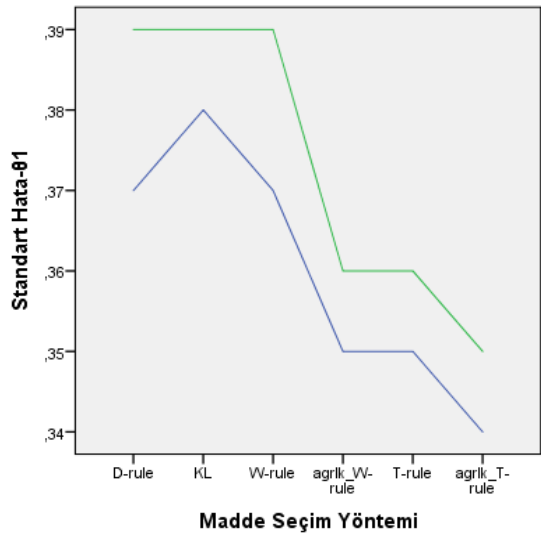
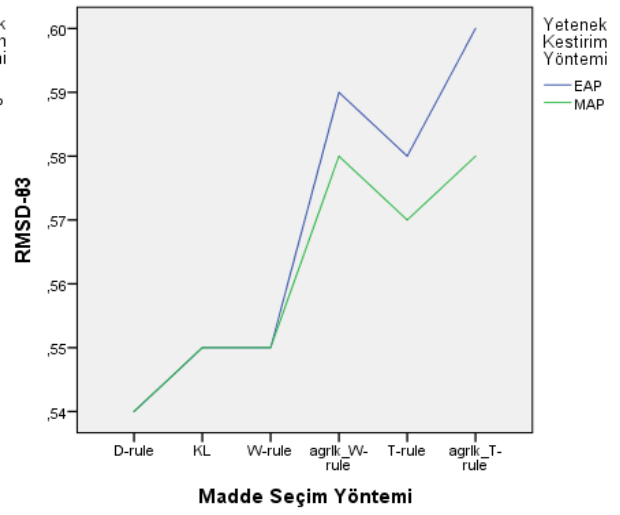
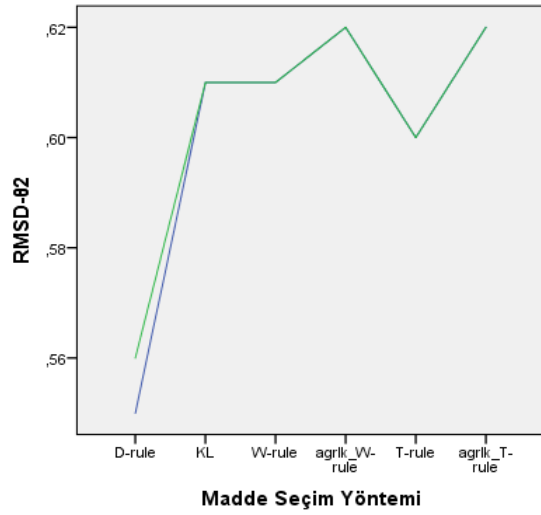
```
options(stringsAsFactors = FALSE, shiny.reactlog=TRUE, useFancyQuotes=FALSE)
sorular <- read.csv("yts.csv", header = TRUE, ";")
set.seed(1234)
nitems <- nrow(sorular)
itemnames <- paste0("Soru ", 1:nitems)
questions <- answers <- character(nitems)
choices <- matrix(' ', nitems, 4)
a1 <- as.matrix(sorular$a1)
a2 <- as.matrix(sorular$a2)
a3 <- as.matrix(sorular$a3)
d <- as.matrix(sorular$b)
for (i in 1:nrow(sorular)) {
  questions[i] <- sorular$soru[i]
  answers[i] <- sorular[i, sorular[,6]]
  choices[i,] <- c(sorular$s1[i], sorular$s2[i], sorular$s3[i], sorular$s4[i])
  a1[i] = sorular$a1[i]
  a2[i] = sorular$a2[i]
  a3[i] = sorular$a3[i]
  d[i] = sorular$b[i]
}
pars <- data.frame(a1=a1, a2=a2, a3=a3, d=d)
mod <- generate.mirt_object(pars, '2PL')
df <- data.frame(Questions=questions, Answer=answers, Option=choices, Type='radio')
head(df)
spacing <- floor(d - min(d)) + 1
pdf(NULL)
title <- "Bireyselleştirilmiş İngilizce Sınavı"
authors <- "Murat Doğan Şahin"
firstpage <- list(h2("İngilizce Sınavı"), h5("Lütfen tüm sorulara cevap veriniz. ",
      "Sorulara geri dönme şansınız olmayacağından lütfen iyi okuyunuz. ",
      "Sınav sonunda puanınız size bildirilecektir."))
demographics <- list(textInput(inputId = "Adınız", label = "Lütfen adınızı ve soyadınızı yazın.",
      value = ""), selectInput(inputId = "Cinsiyet", label = "Lütfen cinsiyetinizi seçin.",
      choices = c("", "Erkek", "Kadın"), selected = ""),
      textInput(inputId="Bölüm", label="Hazırlıktan sonra hangi bölümde öğrenim
      göreceksiniz?", value=""))
lastpage <- function(person){
  list(h2("Testimizi tamamladığınız için teşekkür ederiz."),
    h4('Puanınız -3 ile 3 arasında hesaplanmaktadır, 0 ortalama puan demektir.'),
    h4(sprintf("Sizin Puanınız: %.3f, Puanınızın Standart Hatası = %.3f", person$thetas[1],
      person$thetas_SE_history[nrow(person$thetas_SE_history), 1])),
    h3("Lütfen Next'e tıklayarak sınavınızı bitiriniz."))
}
shinyGUI_list <- list(title = title, authors = authors, demographics =
demographics, demographics_inputIDs = c("Adınız", "Bölüm"), firstpage = firstpage, lastpage =
lastpage)
design <- list(min_SEM = c(.40, Inf, Inf), min_items= 10, max_items = 60)

result <- mirtCAT_preamble(df, mod, method = 'MAP', criteria = 'Drule', start_item = 117, shinyGUI
= shinyGUI_list, design = design)
person <- getPerson()
summary(person)
createShinyGUI()
```

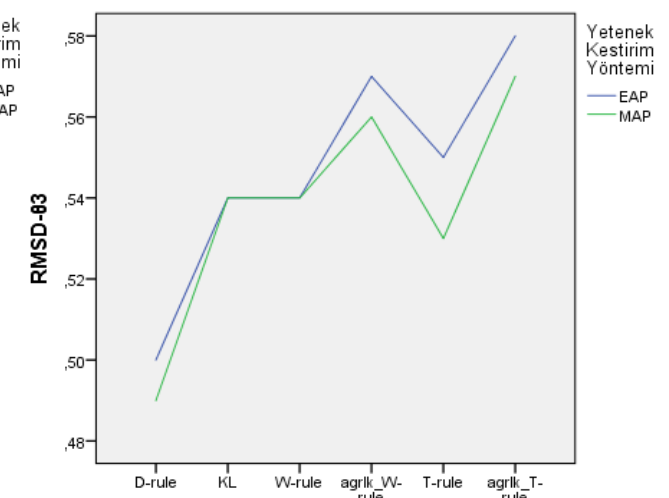
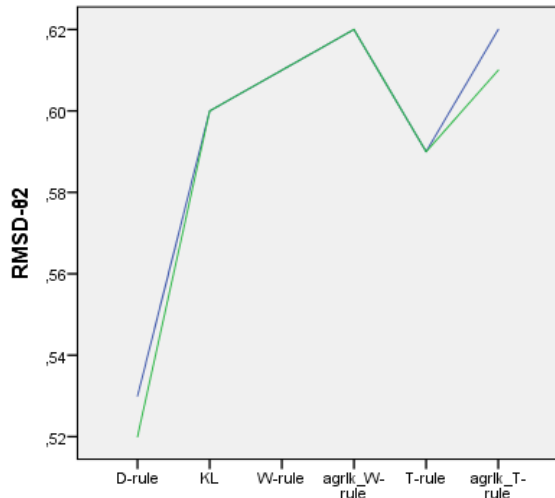
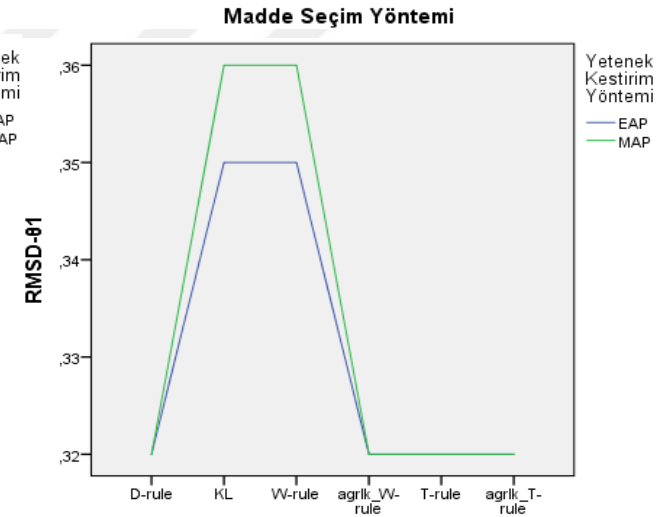
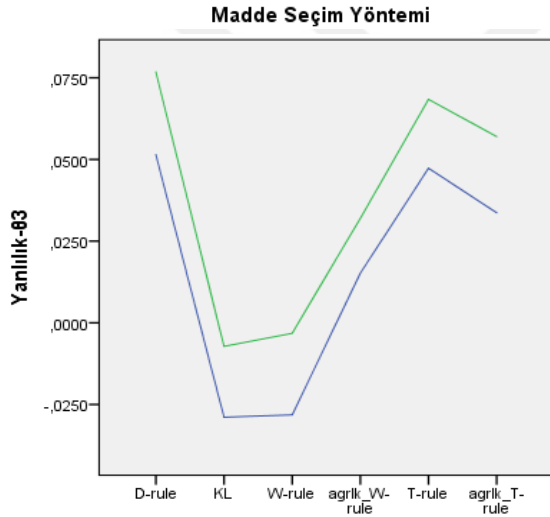
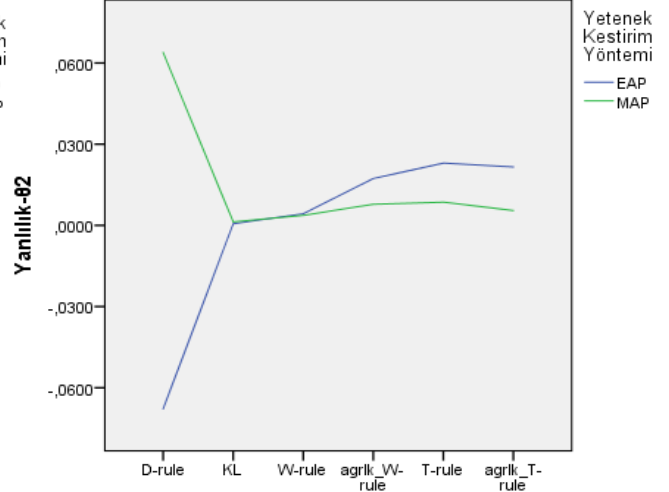
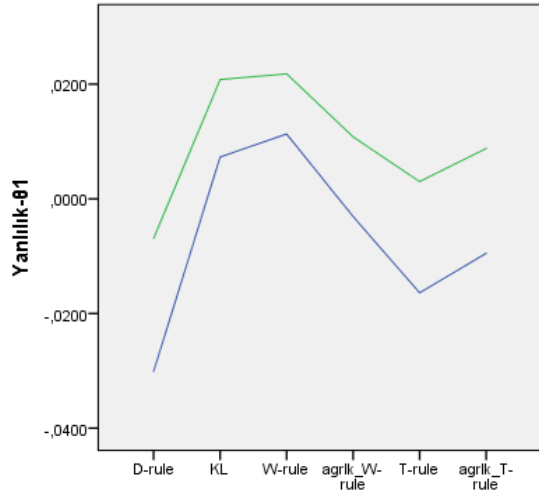
EK 6. FARKLI SONLANDIRMA KURALLARININ KULLANILDIĞI KOŞULLAR İÇİN ELDE EDİLEN İSTATİSTİKLERE AİT GRAFİKLER

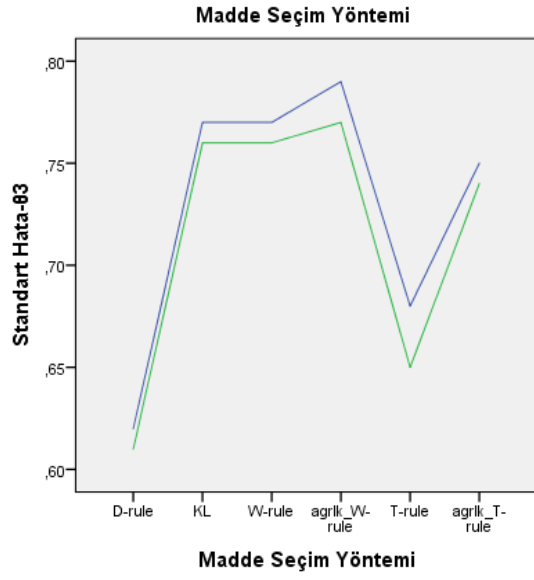
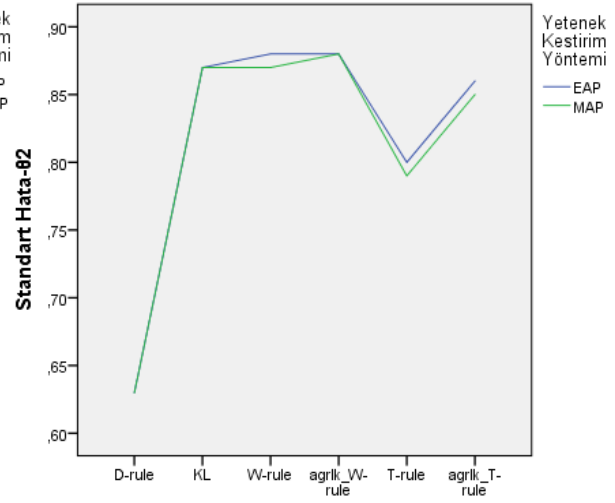
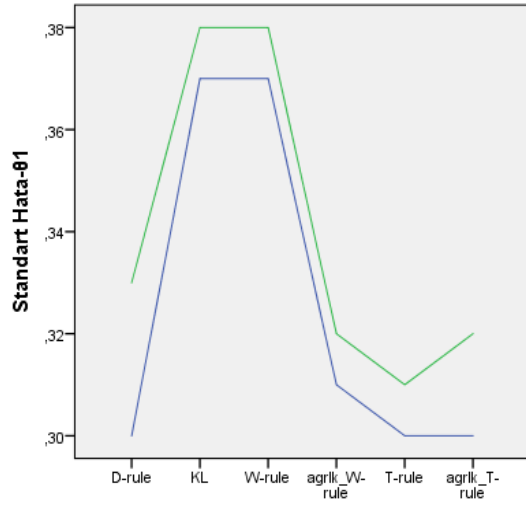
EK-6.1 STANDART HATAYA DAYALI SONLANDIRMA KURALININ KULLANILDIĞI KOŞULLAR İÇİN ELDE EDİLEN GRAFİKLER



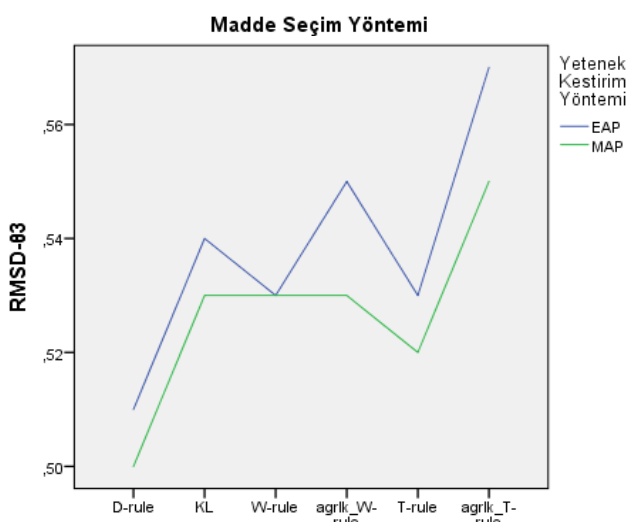
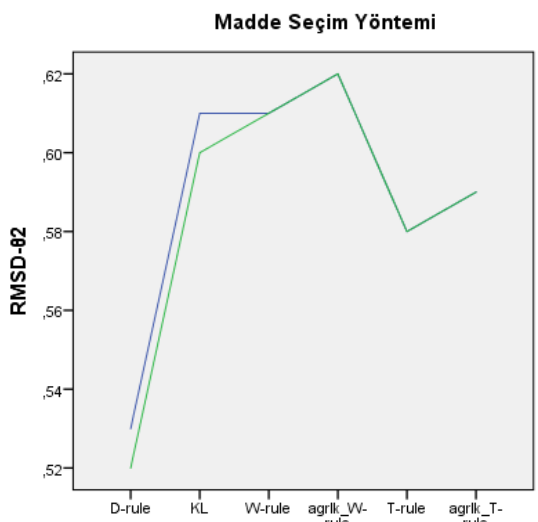
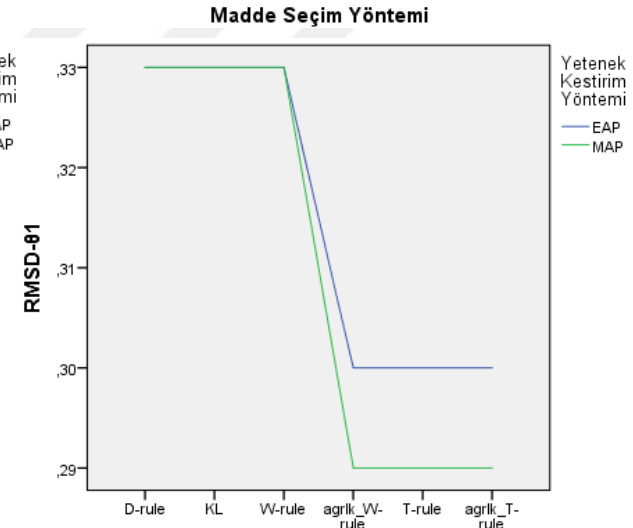
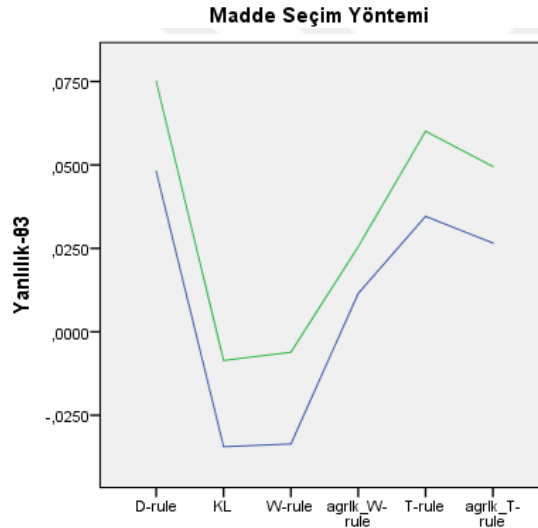
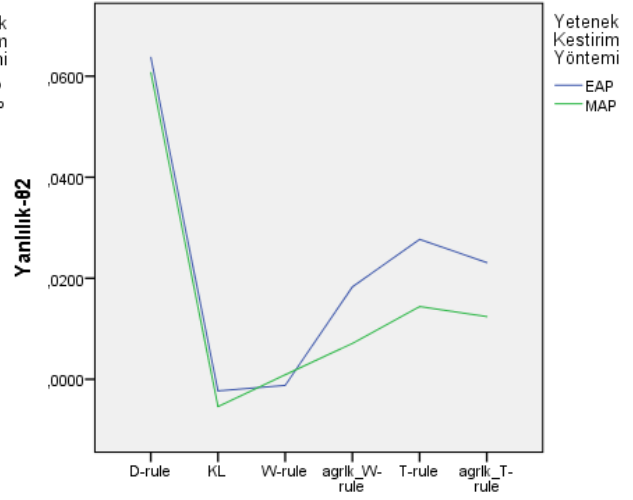
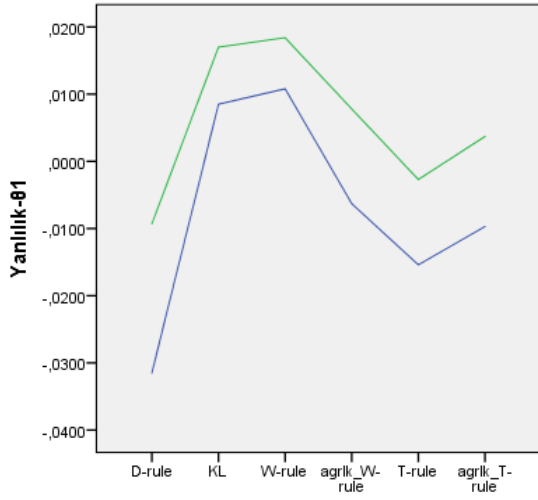


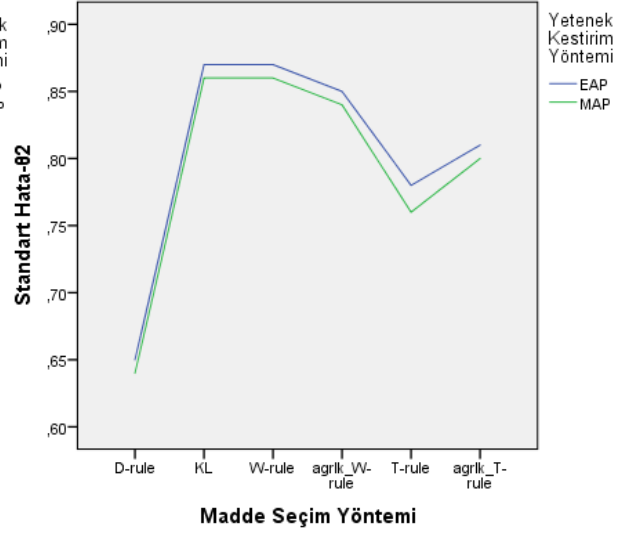
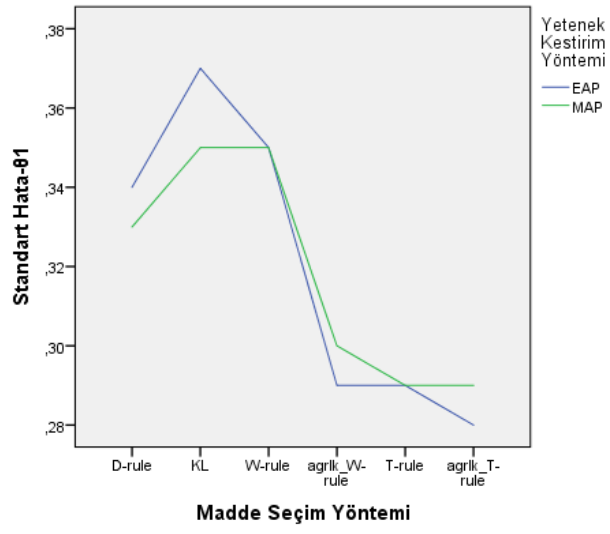
EK-6.2 THETA DEĞİŞİMİNE DAYALI DAYALI SONLANDIRMA KURALININ KULLANILDIĞI KOŞULLAR İÇİN ELDE EDİLEN GRAFİKLER





EK-6.3 SABİT MADDE SAYISINA DAYALI SONLANDIRMA KURALININ KULLANILDIĞI KOŞULLAR İÇİN ELDE EDİLEN GRAFİKLER





ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı	Murat Doğan ŞAHİN
Doğum Yeri	KÜTAHYA
Doğum Tarihi	08.08.1984

Eğitim Durumu

Lise	KÜTAHYA ANADOLU ÖĞRETMEN LİSESİ	2002
Lisans	HACETTEPE ÜNİVERSİTESİ FEN BİLGİSİ ÖĞRETMENLİĞİ	2006
Yüksek Lisans	HACETTEPE ÜNİVERSİTESİ EĞİTİMDE ÖLÇME ve DEĞERLENDİRME ABD	2012
Yabancı Dil	İngilizce: Okuma (iyi), Yazma (İyi), Konuşma (İyi)	

İş Deneyimi

Stajlar		
Projeler		
Çalıştığı Kurumlar	Özel Başaran Yıldız İ.Ö.O. Özel Germiyan Koleji Anadolu Üniversitesi	2006-2011 2011-2012 2012-Halen

Akademik Çalışmalar

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

- Şahin, M. D. & Anıl, D. (2012). **7. Sınıf Öğrencilerinin SBS Fen ve Teknoloji Alt Test Başarılarını Etkileyen Bazı Faktörler.** *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi Özel Sayı 2: 162-170.*
- Koçdar, S., Karadağ, N., Şahin, M. D. & Karadeniz, A. (2017). **Uzaktan Eğitimde Çoktan Seçmeli Soruların Güçlük ve Ayırt Edicilik Değerlerinin Soru Türlerine Göre İncelenmesi.** *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi 32(1), 168-184.*
- Koçdar, S., Karadağ, N. & Şahin, M. D. (2016). **Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Gognitive Levels in an Open and Distance Learning Context.** *The Turkish Online Journal of Educational Technology, 15(4), 16-24*
- Şahin, M. D. & Doruk, Ö. (2015). **Invariance of Factor Structure in Different Booklets of a Large Scale Test.** *E-Learn 2015, Kona-Hawaii.*
- Akyıldız, M & Şahin, M. D. (2015). **Dimensionality and Scalability of a Large Scale Exam by Means of Item Response Theory.** *International Congress on Education for the Future:Issues and Challenges,2015-Ankara.*
- Akyıldız, M & Şahin, M. D. (2016). **Anadolu Üniversitesi İngilizce Yeterlilik Sınavının Bireyselleştirilmiş Bilgisayarlı Test Olarak Uygulanması.** *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, 2016, Antalya.*
- Şahin, M. D. & Akyıldız, M. (2016). **Yönlendirmeye ve Biçimlendirmeye Dayalı Çok Aşamalı Bilgisayarlı Testlerde Yetenek Kestirimlerinin Karşılaştırılması.** *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, 2016, Antalya.*

Şahin, M. D. & Akyıldız, M. (2016). **Bifaktör Yapının Tek boyutlu, Çok boyutlu ve İkinci Dereceden Doğrulamalı Yapılarla Karşılaştırılması.** *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, 2016, Antalya.*

Seminer ve Çalıştaylar

Yapısal Eşitlik Modellemesi Çalıştayı (Eğitimci) – Ekonomi Yaz Seminerleri, Pamukkale Üniversitesi, Denizli (2016)

Ölçek Geliştirme Çalıştayı (Eğitimci) – Ekonomi Yaz Seminerleri, Pamukkale Üniversitesi, Denizli (2016)

Sertifikalar

R'ye Giriş – IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, 2014

İletişim

e-Posta Adresi	mdsahin@anadolu.edu.tr muratdogansahin@gmail.com
Jüri Tarihi	14.04.2017