

**BİREYSELLEŐTİRİLMİŐ BİLGİSAYARLI TEST  
UYGULAMALARINDA FARKLI SONLANDIRMA  
KURALLARININ ÖLÇME KESİNLİĐİ VE TEST UZUNLUĐU  
AÇISINDAN KARŐILAŐTIRILMASI**

**COMPARISON OF DIFFERENT TEST TERMINATION  
RULES IN TERMS OF MEASUREMENT PRECISION AND  
TEST LENGTH IN COMPUTERIZED ADAPTIVE TESTING**

**Melek Gölőah EROĐLU**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

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

Öngördüğü

Doktora Tezi

olarak hazırlanmıştır.

2013

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

Melek G¼lřah EROđLU'nun hazırladıđı "Bireyselleřtirilmiř Bilgisayarlı Test Uygulamalarında Farklı Sonlandırma Kurallarının ¼lme Kesinliđi Ve Test Uzunluđu Açıısından Karřılařtırılması" bařlıklı bu alıřma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde ¼lme ve Deđerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiřtir.



Bařkan

Prof. Dr., řener B¼Y¼K¼ZT¼RK




¼ye (Danıřman)

Prof. Dr., H¼lya KELECIOđLU



¼ye

Prof. Dr., Mehtap AKAN



¼ye

Do. Dr., Nuri DOđAN



¼ye

Do. Dr., Burcu ATAR

ONAY

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim-¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından ..... /...../..... tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca ...../...../..... tarihinde kabul edilmiřtir.

Prof. Dr. Berrin AKMAN  
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

# **BİREYSELLEŐTİRİLMİŐ BİLGİSAYARLI TEST UYGULAMALARINDA FARKLI SONLANDIRMA KURALLARININ ÖLÇME KESİNLİĐİ VE TEST UZUNLUĐU AÇISINDAN KARŐILAŐTIRILMASI**

**Melek Gülőah EROĐLU**

## **ÖZ**

Son yıllarda bilgi teknolojilerinde yaőanan hızlı deėişim ve dönüşümler, bireylerden talep edilen yetenek ve beceri türleri ile seviyelerini etkilemiştir. Bunun sonucu olarak eğitim sistemlerinde de bir takım deėişikliklere gidilmiştir. Bu deėişiklikler eğitimin önemli bir kısmını oluőturan deėerlendirme sürecine de yansımıştır. Bu çerçevede, eğitimde kullanılan geleneksel testlerin yanında bireyselleőtirilmiş bilgisayarlı test uygulamalarının kullanımı artmaktadır. Bireyselleőtirilmiş testlerde geleneksel testlerden farklı olarak test algoritması söz konusudur. Test algoritması teste başlama, devam etme ve testi sonlandırma olmak üzere 3 bölümden oluőmaktadır. Bu çalıőmanın amacı bireyselleőtirilmiş bilgisayarlı test (BBT) uygulamalarında farklı sonlandırma kurallarının kullanılmasının ölçme kesinliėine ve test uzunluėuna etkisini incelemek ve birbirleri ile karőılaőtırmaktır.

Araőtırma simülasyon çalıőması olarak yürütülmüőtür. 1000 kiőiyeye ait yetenek parametresi deėerleri -3 ve +3 deėerleri arasında deėiőecek ve tek biçimli daėıllacak şekilde oluőturulmuőtur. Madde havuzu için; Madde Tepki Kuramında yer alan 3 parametrelili lojistik model kullanılarak madde parametre deėerleri oluőturulmuőtur. Madde havuzu oluőturulurken a parametresi [0,50;2,00]; b parametresi [-3,00;+3,00] ve c parametresi ise [0,05;0,20] aralıėında belirlenmiştir.

Araőtırma kapsamında sabit uzunluk, standart hata, standart hata-en az madde, theta yakınsama ve theta yakınsama-en az madde olmak üzere 5 farklı sonlandırma kuralı kullanılmıőtır. Her bir sonlandırma kuralında farklı koőullar söz konusu olup toplam 12 koőul birbiriyle karőılaőtırılmıőtır. Ayrıca sonlandırma kurallarının karőılaőtırılmasında BBT'de test algoritmasında önemli yere sahip olan farklı madde havuzu büyüklükleri (250 ve 500 madde), yetenek kestirim yöntemleri (Maksimum Likelihood Estimation ve Expected a Posteriori) ve başlama kuralları ( $b=0$  ve  $-1 < b < 1$ ) seçilmiştir. Her bir BBT uygulamasında ölçme kesinliėi için

RMSE, yanlılık ve uyum deęerleri hesaplanmış ve test uzunlukları elde edilip, birbirleriyle karşılaştırılmıştır.

Araştırmanın sonucunda, genel olarak 20 madde sabit uzunluk, 0,220 standart hata ve 0,02 theta yakınsama sonlandırma koşullarında RMSE, yanlılık deęerlerinin düşük elde edildięi ancak uyum katsayılarının önemli oranda etkilenmedięi belirlenmiştir

Ayrıca en az madde koşulunun eklenmesi ile bazı sonlandırma koşullarında ölçme kesinlięi açısından daha iyi sonuçlar vermiştir. Ortalama test uzunluklarına bakıldığında RMSE deęerleri ile ters yönde deęiştii bulunmuştur. Aynı sonlandırma koşullarında madde havuzu büyüklüğünün artması ile ölçme kesinlięi için elde edilen RMSE ve yanlılık deęerlerinin genel olarak daha düşük elde edildięi bulunmuştur. Teste başlama kurallarının etkisinin incelenmesinde ise çok önemli bir farklılık elde edilmemiştir. Yetenek kestirim yöntemi olarak Expected A Posteriori yönteminin kullanılmasının RMSE ve yanlılık deęerlerinde düşmeye neden olduęu belirlenmiştir.

**Anahtar sözcükler:** BBT, Sonlandırma Kuralları, Sabit Uzunluk Sonlandırma Kuralı, Deęişen Uzunluk Sonlandırma Kuralı, RMSE, Yanlılık, Uyum, Test Uzunluğu

**Danışman:** Prof. Dr. Hülya KELECİOĞLU, Hacettepe Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Deęerlendirme Bilim Dalı

# **COMPARISON OF DIFFERENT TEST TERMINATION RULES IN TERMS OF MEASUREMENT PRECISION AND TEST LENGTH IN COMPUTERIZED ADAPTIVE TESTING**

**Melek Gülşah EROĞLU**

## **ABSTRACT**

Fast changes and transformations observed in information technologies have been influencing the type and level of the skills and abilities demanded from individuals. As a result of this, there have been changes in the education systems as well. These changes are also pronounced in the educational evaluation processes that compose an important part of education. Within this framework, in addition to the classical test techniques, computer adaptive testing applications are increasingly preferred. In computer adaptive testing, there exists a test algorithm different than the classical tests. The test algorithm consists of three parts which are Starting, Resuming and Termination. The aim of this study is to measure the effect of different termination rules on measurement precision and test length.

The research was implemented as a simulation study. Skill parametric values that take a value between +3 and -3 and that are uniformly distributed have been formed for 1000 people. For the item pool, item parameter values have been formed by using the 3 parameter logistic model of Item Response Theory. While forming the Item Pool, the intervals for the parameters are defined as such: a parameter [0.50;2.00]; b parameter [-3.00;+3.00] and c parameter [0.05;0.20].

5 different termination rules have been used for the study which are: fixed length, standard error, standard error-least item, theta convergence and theta convergence-least item. Different conditions are in place in each termination rule and a total of 12 conditions are compared. Additionally, in comparing termination rules, different item pools (250 and 500), ability estimation methods (Maximum Likelihood Estimation and Expected a Posteriori), starting rules ( $b=0$  ve  $-1 < b < 1$ ) have been selected since these are critical in the algorithms of Computer Adaptive Testing. RMSE, bias and fidelity values were calculated for the measurement

precision and test lengths were obtained and compared for each of the CAT implementations.

As a result, for the 20 item fixed length, 0,220 standard error and 0,02 theta convergence termination conditions RMSE and bias values are small but fidelity factors are not significantly affected. And with the addition of the least item factor, better results were achieved in some of the termination conditions in terms of measurement precision. The test length is observed to be negatively correlated with the RMSE values. In the same termination conditions, with the increases in item pool generally smaller RMSE and bias values were for measurement precision were achieved. Not a significant change was observed in the evaluation of the effect of the starting rules. The preference of Expected A Posteriori method for the ability estimation is observed to cause a drop in values for RMSE and bias values.

**Keywords:** Computer Adaptive Testing, (CAT), Termination Rules, Fixed Length Termination Rule, Variable Length Termination Rule, RMSE, Bias, Fidelity, Test Length

**Advisor:** Prof. Dr. Hülya KELECİOĞLU, Hacettepe University, Department of Educational Sciences Program in Measurement and Evaluation in Education

## 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.

Melek Gülşah EROĞLU

## TEŞEKKÜR

Akademik ve özel hayatımda önemli bir yeri olan çok değerli danışmanım Sayın Prof. Dr. Hülya Keleciođlu'na tez çalışmam boyunca değerli katkılarını esirgemeyerek güler yüzü ve enerjisiyle her zaman yanımda olduđu için teşekkürü bir borç bilirim. Beni her zaman destekleyen, öğrencilerine yaklaşım ve tutumu ile takdir ettiđim, iş ve akademik anlamda çok değerli bilgiler edindiđim Sayın Prof. Dr. Şener Büyüköztürk'e şükranlarımı sunarım. Ayrıca tez sürecimde özellikle değerli yorumlarıyla çalışmama katkılarını sunan Sayın Prof. Dr. Mehtap Çakan, Sayın Doç. Dr. Burcu Atar ve Sayın Doç. Dr. Nuri Dođan ile lisansüstü eğitimim boyunca her zaman desteđini hissettiđim değerli hocam Sayın Prof. Dr. Selahattin Gelbal'a çok teşekkür ederim.

Ders dönemi ve tez çalışma sürecim dahil hayatımın birçok aşamasında sürekli yanımda olan, birlikte çok şey paylaştıđım değerli dostlarım Uzm. Nagihan Boztunç Öztürk ve Öğr. Gör. Gülşen Taşdelen Teker'e hayatımda her zaman yanımda olmaları dileđiyle sevgilerimi sunarım. Ayrıca manevi desteđini ortaya koyan sevgili Arş. Gör. Kübra Atalay Kabasakal ve Arş. Gör. Elif Sezer'e; sıcacık gülüşüyle yaşamıma renk katan Arş. Gör. Yıldız Yıldırım'a ve burada ismini sayamadıđım Gazi Üniversitesi ve Hacettepe Üniversitesi'ndeki değerli hocalarım ve meslektaşlarıma çok teşekkür ederim.

Son olarak her zaman yanımda olan ve başaracağıma inanıp beni yüreklendiren canım anneme; uzun süredir yanımda olmasalarda varlıklarını hep yanımda hissettiđim ve emeklerini asla unutamadıđım babaannem ve dedeme sonsuz teşekkürlerimi sunarım.



## İÇİNDEKİLER

ÖZ .....	iii
ABSTRACT .....	v
ETİK BEYANNAMESİ .....	vii
TEŞEKKÜR .....	viii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ .....	xi
ŞEKİLLER DİZİNİ.....	xii
SİMGELER VE KISALTMALAR DİZİNİ .....	xiii
1. GİRİŞ.....	1
1.1. Problem Durumu .....	1
1.1.1. Madde Havuzu Geliştirme .....	10
1.1.2. Test Algoritması.....	12
1.2. Problem Cümlesi.....	23
1.2.1. Alt Problemler .....	23
1.3. Sınırlılıklar .....	24
1.4. Araştırmanın Amacı ve Önemi .....	24
2. İLGİLİ ARAŞTIRMALAR.....	26
2.1. İlgili Araştırmalar Özet.....	34
3. YÖNTEM .....	35
3.1. Araştırmanın Türü .....	35
3.2. Verilerin Türetilmesi .....	35
3.2.1. Yetenek Parametrelerinin Elde Edilmesi.....	35
3.2.2. Madde Parametrelerinin Elde Edilmesi .....	36
3.3. BBT Koşulları .....	38
3.4. Verilerin Analizi .....	41
4. BULGULAR VE YORUM .....	43
4.1. Bulgular .....	43
4.1.1. Birinci Alt Probleme İlişkin Bulgular .....	43
4.1.2. İkinci Alt Probleme İlişkin Bulgular .....	47
4.1.3. Üçüncü Alt Probleme İlişkin Bulgular.....	51
4.1.4. Dördüncü Alt Probleme İlişkin Bulgular .....	55

4.1.5. Beşinci Alt Probleme İlişkin Bulgular .....	59
4.1.6. Altıncı Alt Probleme İlişkin Bulgular .....	63
4.1.7. Yedinci Alt Probleme İlişkin Bulgular .....	67
4.1.8. Sekizinci Alt Probleme İlişkin Bulgular .....	71
4.2. Tartışma ve Yorum.....	75
5. SONUÇ VE ÖNERİLER .....	78
5.1. Sonuçlar .....	78
5.1.1. Madde Havuzu Büyüklüğünün 250 veya 500 Madde Olarak Belirlenmesi .....	78
5.1.2. Teste Başlama Kuralının $b=0$ veya $-1 < b < 1$ Belirlenmesi .....	80
5.1.3. Yetenek Kestirim Yöntemi Olarak MLE veya EAP Belirlenmesi .....	81
5.2. Öneriler .....	82
KAYNAKÇA.....	85
EKLER DİZİNİ .....	89
EK-1: Bireylerin Yetenek Parametrelerine Göre Dağılım Grafiği .....	90
EK-2: 250 Maddeden Oluşan Havuz Cevap Fonksiyonu .....	91
EK-3: 500 Maddeden Oluşan Havuz Cevap Fonksiyonu .....	92
EK-4: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 1 .....	93
EK-5: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 2 .....	94
EK-6: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 3 .....	95
EK-7: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 4 .....	96
EK-8: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 5 .....	97
EK-9: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 6 .....	98
EK-10: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 7 ...	99
EK-11: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 8 .	100
EK-12: Her Bir Koşula Ait Elde Edilen Ortalama RMSE değerleri .....	101
EK-13: Her Bir Koşula Ait Elde Edilen Ortalama BIAS değerleri .....	102
EK-14: Her Bir Koşula Ait Ortalama Uyum Katsayıları .....	103
EK-15: Her Bir Koşula Ait Elde Edilen Ortalama Test Uzunlukları .....	104
ÖZGEÇMİŞ .....	105

## TABLolar DİZİNİ

Tablo 3.1. Araştırma Kapsamında Kullanılan Sonlandırma Koşulları .....	40
Tablo 3.2. Araştırmada Kapsamında Ele Alınan Değişkenler ve Koşulları.....	40
Tablo 4.1. Madde Havuzu Büyüklüğü 250, Başlama kuralı $b=0$ ve Yetenek Kestirimi MLE Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri.....	43
Tablo 4.2. Madde Havuzu Büyüklüğü 250, Başlama kuralı $-1 < b < 1$ ve Yetenek Kestirimi MLE Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri.....	47
Tablo 4.3. Madde Havuzu Büyüklüğü 500, Başlama kuralı $b=0$ ve Yetenek Kestirimi MLE Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri.....	51
Tablo 4.4. Madde Havuzu Büyüklüğü 500, Başlama kuralı $-1 < b < 1$ ve Yetenek Kestirimi MLE Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri.....	55
Tablo 4.5. Madde Havuzu Büyüklüğü 250, Başlama kuralı $b=0$ ve Yetenek Kestirimi EAP Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri.....	59
Tablo 4.6. Madde Havuzu Büyüklüğü 250, Başlama kuralı $-1 < b < 1$ ve Yetenek Kestirimi EAP Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri.....	63
Tablo 4.7. Madde Havuzu Büyüklüğü 500, Başlama kuralı $b=0$ ve Yetenek Kestirimi EAP Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri.....	67
Tablo 4.8. Madde Havuzu Büyüklüğü 500, Başlama kuralı $-1 < b < 1$ ve Yetenek Kestirimi EAP Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri.....	71

## ŞEKİLLER DİZİNİ

Şekil 1. 1. Bireyselleştirilmiş Bilgisayarlı Test Algoritmasının Şeması .....	13
Şekil 3. 1. Bireylerin Yetenek Parametrelerine Göre Dağılımı .....	36
Şekil 3. 2. 250 Maddelik Havuza Ait Havuz Bilgi Fonksiyonu .....	37
Şekil 3. 3. 500 Maddelik Havuza Ait Havuz Bilgi Fonksiyonu .....	37
Şekil 4. 1. Birinci Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki.....	45
Şekil 4. 2. Birinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu .....	46
Şekil 4. 3. İkinci Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki .....	49
Şekil 4. 4. İkinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu.....	50
Şekil 4. 5. Üçüncü Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki .....	53
Şekil 4. 6. Üçüncü Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu .....	54
Şekil 4. 7. Dördüncü Alt Probleme Ait Test Uzunluğu ve RMSE Arasındaki İlişki .....	57
Şekil 4. 8. Dördüncü Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu.....	58
Şekil 4. 9. Beşinci Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki .....	61
Şekil 4. 10. Beşinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu.....	62
Şekil 4. 11. Altıncı Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki .....	65
Şekil 4. 12. Altıncı Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu .....	66
Şekil 4. 13. Yedinci Alt Probleme Ait Test Uzunluğu ve RMSE Arasındaki İlişki .....	69
Şekil 4. 14. Yedinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu .....	70
Şekil 4. 15. Sekizinci Alt Probleme Ait Test Uzunluğu ve RMSE Arasındaki İlişki.....	73
Şekil 4. 16. Sekizinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu ...	74

## SİMGELER VE KISALTMALAR DİZİNİ

**BBT:** Bireyselleştirilmiş Bilgisayarlı Test

**KTK:** Klasik Test Kuramı

**MTK:** Madde Tepki Kuramı

**3PLM:** 3 Parametrelili Lojistik Model

**RMSE:** Standart Hata Ortalamalarının Karekökü

**BIAS:** Yanlılık

**MLE:** Maksimum Likelihood Estimation

**MAP:** Maksimum a Posteriori

**EAP:** Expected a Posteriori

**MFI:** Fisher'ın En Çok Bilgi Yöntemi

**SE( $\theta$ ):** Theta yetenek düzeyinin standart hatası

# 1. GİRİŞ

Bu bölümde; araştırmanın temelini oluşturan problem durumu, problem cümlesi alt problemler, sınırlılıklar ve araştırmanın amacı ve önemi ile ilgili bilgiler yer almaktadır.

## 1.1. Problem Durumu

Bilim ve teknolojiadaki gelişmeler bireylerden beklenen nitelikleri hızla değiştirmektedir. Bu niteliklere temel olabilecek bilgi ve becerilerin kazandırılması eğitim sisteminin başarılı bir şekilde yürütülmesi ile mümkün olmaktadır. Eğitim sisteminde girdi, süreç ve çıktıların etkililiğinin belirlenmesi iyi bir ölçme ve değerlendirme ile mümkün olmaktadır. Ölçme ve değerlendirme amacıyla geliştirilen testler genellikle, testle ölçülen özellikler ile bir grup testi alan adayın beklenen performans düzeyini karşılaştırmayı amaçlamaktadır. Belirli bir grup adaya çok kolay bir test uygulamak adayların dikkatsizce davranıp hatalar yapmasına veya testin içinde bir şaşırtmaca olduğunu düşünerek kasıtlı olarak yanlış cevabı tercih etmelerine neden olabilir. Diğer taraftan adaylara çok zor maddelerden oluşan bir test uygulandığında ise adayların tahmin etme yoluna başvurmalarına veya benzer başka davranışlara yönelmelerine neden olabilmektedir (Linacre, 2000). Her iki durumda bilgi verici olmayan sonuçlar elde edilmesi kaçınılmazdır.

Geleneksel testlerin hem bu sınırlılığının üstesinden gelme hem de iç tutarlılık anlamında güvenilirliği artırmak için, testte tipik olarak orta güçlükte yani  $p = 0,5$  zorluk derecesindeki maddeler seçilir. Bu maddeler orta yetenek düzeyindeki bireyler için en iyi ölçümü vermekte iken, ortalamanın çok altında ve üstünde yer alan bireyler açısından yeterince bilgi vermemektedir. Örneğin dördüncü sınıflar için geliştirilmiş bir aritmetik testi ikinci sınıflar için çok zor, altıncı sınıflar için ise çok kolay olacaktır. Bununla birlikte dördüncü sınıf öğrencileri arasından ikinci sınıf seviyesinin altında veya altıncı sınıf seviyesinin üstünde yetenek düzeyine sahip öğrencilerde olabilecektir. Klasik testin seviyesinden sapan öğrenciler için test çok az bilgi sağlayacaktır. Düşük yetenek düzeyindeki öğrenciler hemen hemen tüm maddeleri yanlış cevaplarken; yüksek seviyedeki öğrenciler ise doğru cevaplayacaktır. Bu öğrenci grupları için elde edilen sonuçlar öğrencileri

birbirinden ayırt etmeye imkan tanımayacaktır. Bu durum sadece başarı testlerinde değil psikolojik özelliklerinde ölçülmesinde de geçerlidir (Weiss, 2004).

Geleneksel testlerle ilgili belirtilen sınırlılıklarından dolayı bireye biçimlendirilmiş (tailored tests) veya bireye uyarlanmış (adaptive testing) testlere ihtiyaç duyulmuştur (Lord, 1970; Wood, 1973; Weiss, 1983; Hambleton and Swaminathan 1985). Bireyselleştirilmiş testlerde, bireyin ölçülmek istenen yeteneğine uygun olarak test maddelerinin güçlüğü'nün ayarlanması esastır. Test maddelerinin güçlüğü'nün bireyin yeteneğine göre ayarlanıyor olması ile bireyin yeteneğinin çok üstünde ya da çok altında soru almasının da önüne geçilebilmektedir.

Bireyselleştirilmiş testler bir test algoritması takip edilerek uygulanır. Test algoritması ise 3 bölüm ile tanımlanır (i) *Başlama*: Bireye sunulacak ilk madde nedir? (ii) *Teste Devam Etme*: Her bir maddeye verilen cevaptan sonra hangi madde sorulacaktır? (iii) *Sonlandırma*: Test ne zaman bitirilecektir? Bireysel testler özellikle başlama ve sonlandırma aşamalarında daha karmaşık bireye uyarlanmış test algoritmalarını kullanırlar. Bireye uyarlanmış test algoritmalarında; teste başlama ve sonlandırma aşamaları bireyin özellikleri ve/ veya madde cevapları dikkate alınarak oluşturulmaktadır.

Bireyselleştirilmiş testler ile ilgili literatür incelendiğinde, 1905 yılından itibaren önemli ilerlemeler sağlandığı görülmektedir. Bu alanda çalışan ilk araştırmacılar arasında olan Alfred Binet'in hazırladığı bireyselleştirilmiş testte yer alan sorular zorluk derecelerine göre sıralanmıştır. Binet bireyin yeteneğini tahmin ederek uygun soru zorluğu ile testi başlatmış, testi alan birey soruyu doğru cevaplamışsa daha zor bir soru ile eğer yanlış cevaplamışsa daha kolay bir soru ile testi devam ettirmiştir (Weiss, 1983). Binet zeka testinin çeşitli versiyonlarında çocukların kronolojik yaşına dayanılarak farklı başlama maddeleri seçilmiştir. Artan zorlukta sorular, çocuğun önceden belirlenmiş belli sayıda yanlışa ulaşmasına kadar sürmektedir ve bu noktada test sonlandırılmaktadır (Thissen ve Mislevy, 2000).

Psikometride Binet'in bu uygulamasından sonra, bireye uyarlanmış testler 1950'lerin başına kadar ihmal edilmiştir (Hambleton, Swaminathan ve Rogers, 1991). Daha sonra bireyselleştirilmiş testler Frederic Lord'un öncülüğünde 1970'li yıllarda tekrar önem kazanmıştır. Grup testlerinin teorik yapısını düzenleyerek

bireysel testlere uyarlamış, hem de birçok pratik detaylarla donatmıştır ve bilgisayar yardımı olmadan uygulanabilen testlere örnek teşkil eden esnek düzey (The self-scoring flexilevel) testi tanımlamıştır. Burada bireylerin, testin sorularının yarısına cevap vermesi beklenmektedir. Örneğin 21 maddelik bir testin tüm maddeleri kolaydan-zora doğru sıralandıktan sonra orta zorluktaki test sorusu ile teste başlanır. Bireylerin doğru veya yanlış cevap vermesine göre kolay ya da zor teste yönlendirilmesi yapılır. Bu genellikle çoktan seçmeli soruya doğru yanıt verilmesi ile yanıtın renginin maviye dönüşmesi ve yanlış cevap sonrası ise kırmızıya dönüşmesi şeklinde olur. Mavi rengi alanların mavi kitapçığa, kırmızı cevap alanların ise kırmızı kitapçığa yönlendirilmesi sağlanır. Örneğin mavi testte iki soruyu doğru cevaplamış bir öğrenci daha sonra yanlış cevap verirse kırmızı testin ilk sorusuna tekrar yönlendirilir. Bu test türünün önemli bir dezavantajı bireyleri karmaşık bir madde seçim algoritması ile karşı karşıya bırakmasıdır. Bazı öğrencilerin yönergeyi anlayamamasından dolayı test puanları etkilenebilir (Thissen ve Mislevy, 2000).

Bireyselleştirilmiş testler sadece psikoloji alanında da çalışılmamıştır. Örneğin; Askeri kurumların birçok farklı biriminde (Deniz Araştırmaları, Donanma Personeli Araştırma ve Geliştirme Birimi, Hava Kuvveti İnsan Kaynakları Laboratuvarı, Askeri Araştırma Enstitüsü) uyarlanmış testlerin avantajları fark edilmiş ve birçok teorik araştırma girişimleri desteklenmiştir. Bu süreç boyunca birçok psikometriste ihtiyaç duyulmuş ve bununla beraber gerçek fırsatın yaratılması için ucuz ve yüksek güce sahip bilgisayar olanakları beklenmiştir. 1980'lerde program geliştirilmiş ve bilgisayar ortamında bireye uyarlanmış test uygulamaları resmen başlatılmıştır (Wainer, 2000).

1980'lerin sonlarına doğru bireysel bilgisayarların eğitimde giderek yaygınlaşması ile bu test formatları bilgisayar ile uyumlu formatlara doğru hızlıca geçmeye başlamıştır (Linden ve Glas, 2002). Bir testin bilgisayar aracılığıyla sunulması iki şekilde olabilir. Birincisi kağıt kalem testlerinin bilgisayar aracılığıyla doğrusal yolla (maddelerin tüm bireylere önceden belirlenmiş belirli bir sıra ile) sunulması ve izlenmesinin mümkün olduğu Bilgisayar Temelli Test (Computer Based Test) tir. Bu test kağıt kalem testi ile madde, puanlama yöntemleri bağlamında benzer olup tek farkı testin uygulanma biçimidir. İkinci ise Bireyselleştirilmiş Bilgisayarlı Test



(BBT) tir. BBT esas olarak TOEFL (Test of English as a Foreign Language), GRE (Graduate Record Examination), GMAT (Graduate Management Admission Test) gibi dünya çapında yapılan testlerde ve birçok başka alanda kullanılmaktadır. BBT'lerde kağıt-kalem cinsinden karşılaştığımız şekilde soru formu yoktur. Bu testleri alan bireyler bir bilgisayar karşısına geçerler ve soruları bilgisayar başında cevaplarlar. Burada bilgisayar bir çoklu-ortam aracı, gelişmiş grafik, ses vs. gibi özellikler sunmaktan çok daha fazlasını yapmaktadır. BBT'de testin uygulama şekli test maddelerinin sınavı alan herkesin yeteneklerine göre biçimlendirilmesine olanak sağlamaktadır. Ayrıca bilgisayar ortamında bireyselleştirilmiş testlerin bireylere uygulanma şekli bilgisayar temelli testlerden farklı olarak doğrusaldan bireyselleştirmeye dönüşmektedir (Pearson Assessment, 2009). Yani bilgisayar önceden belirlenmiş bir soru grubunu sıra ile uygulamak yerine, soruları testin gidişine göre seçerek sormaktadır (Kalender, 2004).

Burada bilgisayar teknolojisinin kullanılarak testin uygulanacak olması, kağıt-kalem testlerinde olduğu gibi sınavı alacak tüm bireyler için belli standartların kullanılmasını ortadan kaldırmamaktadır. Kağıt-kalem testlerinin uygulanmasında belirli standartların sağlanmasında göz önüne alınan; testin uygulandığı sınıflarda masa ve sıraların düzeni, ışıklandırma, havalandırma v.b koşullar, BBT uygulamalarında da farklı biçimlerde dikkate alınmaktadır. Bunlar ekranın parlamasının kontrol altına alınması, sistem güvenilirliğinin ve yedeklenmenin sağlanması, ekranın hem grafik hem de metin için yeterli çözüm sunabilmesi, dallanma prosedürlerinin uygun işlemesi, madde sunumu ve test puanlamanın kusursuz işlemesi v.b olabilir. (Wainer, 2000).

Kağıt-kalem testlerinden, BBT uygulamalarına geçişlerin hız kazandığı ilk zamanlarda öncelikle geçerlik ve güvenilirlik çalışmalarına yer verildiği görülmektedir. Literatürde BBT uygulamaları ile kağıt-kalem testlerinin güvenilirlik ve geçerlik açısından karşılaştırıldığı çalışmalar söz konusudur. McBride ve Martin (1983) 15 maddeden oluşan BBT ve kağıt-kalem testlerini geçerlik ve güvenilirliklerini incelemişlerdir. Kağıt- kalem testindeki madde sayısının sadece yarısını kullanarak BBT'de çok daha güvenilir sonuçlar elde edildiğini belirtmişlerdir. Köklü (1990) bireyselleştirilmiş testler ile kağıt-kalem testlerini geçerlik ve güvenilirlik açısından karşılaştırmıştır. Her iki format arasında yetenek

kestirimleri açısından anlamlı bir farklılık bulunmamıştır. Geçerlik çalışması için ölçüt geçerliği belirlenmiş ve öğrencilerin fen kursundan aldıkları puanlar kullanılmıştır. Korelasyon değerleri bireyselleştirilmiş testler ile kağıt-kalem testleri için sırasıyla 0.88 ve 0.81 olarak bulunmuştur. Kaptan (1993) bireyselleştirilmiş testler ile kağıt-kalem testlerinde yetenek kestirimlerini karşılaştırmıştır. Bireylere 50 maddelik matematik kağıt-kalem testi ve 14 maddelik BBT uygulanmıştır. Çalışmada, BBT formatında uygulanan madde sayısının %70 oranda azaldığı belirlenmişken, her iki formdan elde edilen yetenek kestirimleri arasında manidar bir farklılık gözlemlenmemiştir. Mead ve Drasgor (1993) BBT'leri ve kağıt-kalem testlerinin eşdeğerliği araştırmak amacıyla hız ve kalem-kağıt testlerinin bireylerin bilişsel yeterliklerine etkisini incelemiştir ve hız testleri için ortalama korelasyon 0.97 ve güç testleri için ise 0.72 bulmuşlardır. Bu bulgular sonucunda uyarlanmış ve geleneksel test uygulama biçimine göre eşdeğerlik farkı bulunmadığını belirtmişlerdir.

Yapılan çalışmalara kısaca bakıldığında; BBT uygulamasının geleneksel kağıt-kalem testleri ile arasında korelasyon katsayısının genellikle orta ve yüksek değerlerde çıktığı görülmektedir. Ancak BBT uygulamasının, geleneksel uygulama ile karşılaştırıldığında; birçok açıdan avantajlarının olması BBT uygulamasını daha cazip kılmaktadır.

Weiss (1998)'a göre psikometrik bakış açısından, BBT uygulamasının iki önemli avantajı vardır (Hambleton, Swaminathan ve Rogers, 1991). Biri, ölçmenin kesinliğini/doğruluğunu (precision/accuracy of measurement) artırması, diğeri ise, daha kontrollü ve güvenli bir test ortamı sağlamasıdır. Rudner (1998)'e göre BBT'lerin, geleneksel test anlayışına göre üstünlükleri ve uygulamalarının sağladığı faydalar aşağıda maddeler halinde sıralanmıştır:

**i) Test yönetiminin esnekliğini önemli oranda artırır.**

**ii) Testlerin sonuçları bireye hemen sunulabilir.**

**iii) Hem cevapların yazılması için bir kağıda hem de test alma için eğitime ihtiyaç yoktur. Test yönetiminden kaynaklanan hatalar ölçme hatasının bir faktörü olarak elimine edilebilir. Ancak yine de test yönetimin belli düzeyde sağlanmasına ve ortamın standartizasyonuna ihtiyaç duyulmaktadır.**

**iv)** Bireylere kendi hızında ilerleme sağlar. Yani cevaplayıcılar maddelere kendi hızında cevap verirler, diğerlerini beklemek ya da zamana ihtiyacı olanlara stres yüklemek söz konusu değildir.

**v)** Madde ve test güvenliği sağlar. Çünkü hiç kimse aynı soru takımını almadığından cevaplayıcının cevaplarını paylaşması çok anlam ifade etmez.

**vi)** Bilgisayar madde biçimi ve zaman konusunda farklı imkanlar sağlar. Örneğin, bilgisayar ortamında test yapıldığı için maddelerde her türlü görsel ve işitsel öğeler kullanılabilir.

**vii)** BBT uygulamasında kağıt-kalem testlerine kıyasla daha az madde ile kabul edilebilir kesinliğe ulaşıldığından daha az zaman yeterli olabilir. BBT uygulaması aynı güvenilirlik düzeyinde test sonuçları elde etmede zamanı % 50 den fazla azaltabilir.

**viii)** Kısa test uygulaması, bireyin sonucunu önemli derecede etkileyen test yorgunluğunu azaltır.

**ix)** Geleneksel testler genel olarak sadece belirli bir yetenek düzeyi için iken, BBT her yetenek düzeyi için uygundur.

Linacre (2000), yukarıda verilen avantajları ve literatürü göz önüne alarak BBT uygulamasının avantajlarını 5 madde ile özetlemiştir.

**i)** Birey için uygun olmayan soruların bireye yönlendirilmesini engeller: Cevaplayıcılar için çok kolay ya da zor olan maddelerin uygulanması dikkatsiz davranma ve tahmin etme davranışlarına neden olabileceğinden BBT ile bu tür davranışların önüne geçilmiştir.

**ii)** Testlerin kısalması söz konusudur.

**iii)** Testlerin hazırlanması-uygulanması ve raporlanması daha hızlıdır: Düzey belirleme testleri, maddeler yazılır yazılmaz uygulamaya hazırdır. Ayrıca GRE, SAT, TOEFL gibi önemli sınavlarda optik okuyucudan kaynaklanacak tüm sorunlar ortadan kaldırılmış olur.

**iv)** Testten elde edilen sonucu, cevabı yanlış anahtarlanmış madde hemen hemen hiç etkilemez: Geleneksel sınavlarda maddenin yanlış anahtarlandığı durumlarda test sonuçlarının tekrar hesaplanması söz konusu olmaktadır. BBT'lerde maddenin yanlış anahtarlanması durumunda bireyin yetenek kestirimi bu durumdan pek fazla etkilenmeyecektir.

**v)** Testler iyi bir deneyim sunabilir: Uygulanılan ortam oyuna benzer bir mantığa sahip olduğundan BBT'ler ilgi çekici ve iyi bir deneyim olabilir.

BBT uygulamalarının burada bahsedilen üstünlüklerinin yanında bazı sınırlılıklarının olduğunu da söylemek mümkündür (Hambleton ve Stocking, 1984; Sands, Waters ve McBride, 1997; Rudner, 1998; Çıkrıkçı Demirtaşlı, 1999; Kalender, 2004). Bu sınırlılıklar;

**i)** Bilgisayar donanımındaki sınırlılıklar ve maliyeti,

**ii)** Bireylerde bilgisayar kullanmadaki yetersizlik korkusu,

**iii)** Her konu ya da yetenek için uygulanamaması,

**iv)** Ölçülen yeteneğin tek boyutlu olma şartının sağlanamaması,

**v)** Geniş madde havuzuna ihtiyaç duyulması,

**vii)** Bilgisayar donanımının uygulanacak madde türünü sınırlaması (detaylı resim ve grafik çizimi veya sunumu zor olan okuma pasajları),

**vi)** Bireylerin cevaplarını geri dönüp kontrol etme şansının olmaması,

**vii)** Bireylerin farklı madde takımlarını alıyor olmasının bireyde eşitsizlik algısını oluşturması

şeklinde sıralanabilir.

BBT uygulamalarının kuramsal temelinden bahsedilecek olursa; BBT uygulaması beş temel bileşenden oluşmaktadır. Bunlar (a) madde havuzunun oluşturulması (b) ilk maddenin seçimine ilişkin kural (c) bir sonraki maddenin seçimine karar verecek madde seçim algoritması (d) bireyin önceki cevaplarını temel alan yetenek kestirim süreci (e) testin sonlandırılmasına ilişkin kuraldır (Weiss ve Kingsbury, 1984). Bu beş bileşeni temel olarak madde havuzu geliştirme ve test algoritması olarak iki temel başlıkta inceleyebiliriz. Test geliştiriciler buradaki

bileşenlere ait karar verdikten sonra sıra bilgisayarın bileşenlere uygun şekilde programlanmasına gelecektir. Lord'a göre BBT uygulaması için bilgisayar sırasıyla;

1. Bireyin yanıtlarından yola çıkarak henüz uygulanmamış çeşitli test maddelerine vereceği cevabı tahmin edebilecek
2. Bu bilgiyi sonraki uygulanacak madde seçiminde etkili bir şekilde kullanabilecek
3. Testin sonunda yeteneği test edilecek bireye ait sayısal bir değer verebilecek

şekilde programlanmalıdır (Hambleton ve Swaminathan, 1985).

BBT uygulamalarında; test algoritma sürecinin işleminde en uygun matematiksel model Madde Tepki Kuramı (MTK)'dır. MTK bireyin yeteneğinden yararlanarak, bireyin madde üzerinde göstereceği performansın olasılığı üzerinde yoğunlaşır. Bu kuramın en önemli noktası bireylerin yeteneklerini, aldıkları maddelerden bağımsız olarak kestirmektedir. Bireyler farklı maddeler içeren iki farklı test olsa bile kestirilen yetenek düzeyleri farklı olmamaktadır. Kısaca, MTK farklı testi alan, farklı kişilerin yeteneklerini kestirebilmemiz için standart bir çerçeve sunmaktadır (Hambleton, Swaminathan ve Rogers, 1991). MTK' nın üzerine kurulu olduğu iki temel postula Hambleton ve Swaminathan (1985) tarafından aşağıdaki şekilde belirtilmiştir.

i) Bireyin bir test maddesindeki performansı; özellik, örtük özellik veya yetenek gibi faktörlerce yordanabilir.

ii) Bireyin madde performansı ile madde performansının altında yatan özellik kümesi arasındaki ilişki, madde karakteristik eğrisi (MKE) veya fonksiyonu (MKF) adı verilen monoton artan bir fonksiyonla ifade edilebilir.

Açıkça görüldüğü gibi MTK'nin madde bazında matematiksel bir kuram olduğu söylenebilir. Her birey yetenek parametresi ( $\Theta$ ) ve madde parametreleri takımı ile karakterize edilmektedir. MTK; bireyin yeteneği ile madde parametrelerini kıyaslar ve bireyin o maddeyi doğru cevaplayabilme olasılığını kestirir. (Wainer, 2000).

MTK'te kullanılan madde parametre sayısına göre üç farklı model karşımıza çıkmaktadır. 1 parametrelili model; madde güçlük parametresi ( $b_i$ ) ve bireyin yetenek düzeyi  $\theta$  ( $\Theta$ ) arasında ilişki kurmaktadır. 2 parametrelili modelde madde güçlük parametresinin yanına, madde ayırt edicilik ( $a_i$ ) parametresi eklenmektedir. 3 parametrelili modelde ise; iki parametrelili modele şans faktörü parametresi ( $c_i$ ) eklenmektedir. İyi bir madde havuzunda, yüksek  $a_i$  ve düşük  $c_i$  parametrelerine sahip maddelerin bulunması beklenir (Rudner, 1998).

BBT uygulamalarında; çoktan seçmeli maddeler için en uygun MTK modeli 3 parametrelili modeldir (Weiss, 1983; Hambleton, Swaminathan ve Rogers, 1991). Bu modelde  $\theta$  yetenek düzeyindeki bir bireyin bir  $i$  maddesini doğru yanıtlama olasılığı  $P_i(\theta)$ , aşağıdaki eşitlik yardımıyla hesaplanır.

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Burada;

$b_i$  = madde güçlük parametresini,

$a_i$  = madde ayırt edicilik parametresini (düşük ve yetenek grubundaki bireyleri ayırma gücü)

$D = 1.7$  (ölçekleme faktörü) değerini

$c_i$  = şans faktörü parametresini (düşük yetenek düzeyindeki bireylerin maddeyi doğru yanıtlama olasılığı)

temsil etmektedir (Hambleton ve Swaminathan, 1985).

Çalışmanın bundan sonraki kısmında BBT uygulamasında önemli bir yere sahip olan madde havuzu geliştirme ve test algoritmasına ayrıntılı olarak yer verilecektir.

### 1.1.1. Madde Havuzu Geliştirme

MTK'ye dayalı BBT'de verilen bir özellik için, görelî olarak geniş bir madde havuzu geliştirilir ve maddelerin madde bilgi fonksiyonu belirlenir. Binet'in soru bankasına benzer bir şekilde iyi bir BBT bankası yetenek düzeyi  $\theta$ 'nın bütün aralıkları boyunca bilgi sunan maddelere sahiptir (Weiss, 2004).

Genellikle klasik kâğıt-kalem testlerinde, bütün bireyler testin özel bir formunda yer alan tüm maddeleri alırlar. Bu maddeler geniş bir madde havuzundan seçilmiş olabilir ve bir tanesi seçildiği zaman test formundaki maddeler sabit olur. BBT'lerde ise, bireyler geniş bir madde havuzundan her bireyin yeteneğine uygun olarak seçilen madde takımlarından oluşan birçok farklı test formunu alırlar. Madde havuzunun çok geniş farklı yetenek düzeyleri için yüksek nitelikte maddeler içermesi bireyselleştirilmiş testin avantajlarından birisidir. Öğrencilerin başarısını belirlemede kullanılan klasik kâğıt- kalem testlerinde bu durumun aksine genellikle ortalama yetenek düzeyine göre seçilmiş maddeler bulunmaktadır. Geniş bir ranjda yüksek nitelikte madde yazmak için, bireyselleştirilmiş test madde havuzunun madde kalibrasyonu, uygulama ve puanlamanın altında yatan psikometrik modelin varsayımlarını sağlaması gerekmektedir (Flaughner, 2000).

Madde havuzu BBT için kritik bir öneme sahip olmasına rağmen bir BBT madde bankası geliştirmek için özel prosedürler geliştirilmemiştir. Georgiadou (2006)'a göre, bir madde havuzunun kalitesi 2 kriterle değerlendirilebilir.

- a. Toplam madde sayısı bir test oturumu boyunca bilgi verebilmek için yeterli düzeyde olmalıdır.
- b. Havuzdaki maddeler arzu edilen yetenek düzeyi için yeterli bilgiyi sunacak özelliklere sahip olmalıdır (akt: Yoo, 2011).

Flaughner (2000)'a göre madde havuzu geliştirmede izlenmesi gereken plan aşağıdaki gibidir:

1. **Her bir içerik kategorisi için yeterli sayıda madde oluşturulması:** Yeni test maddesi oluşturmada baskın içerik alanı ve madde yazmada gerekli olan kurallar bilgisine ihtiyaç vardır.

**2. Madde niteliğinin gözden geçirilmesi:** Maddelerin mümkün olduğunca yüksek nitelikte yazılması test geliştirme sürecinin önemli bir kısmını oluşturmaktadır. Bu durum BBT'lerde bireylerin puanları, maddelerin göreceli bir kısmından oluştuğu için bilhassa önemlidir. Düşük nitelikli test maddeleri bazı bireylerin puanlarında büyük etkilere yol açarken diğerlerinde bir sorun oluşturmayabilir çünkü farklı bireyler farklı maddeleri alırlar. Madde niteliklerinin gözden geçirilmesi için iki süreç işleyebilir. Birincisi madde niteliklerini test uzmanlarının gözden geçirmesi, ikincisi ise materyalin alt grup ya da cinsiyete göre yanlılık göstermesinin önlenmesi amacıyla test duyarlılığının gözden geçirilmesidir.

**3. İlk ön uygulamanın gerçekleşmesi:** Ön uygulama maddelerin performansının nasıl işlediğine dair kanıtlar sağlamak için gereklidir. Yeni maddelerle ilgili bilgi edinmek için gerçek durumun oluşturulması önemli olduğundan en çok tercih edilen çözüm, az sayıda maddenin uygulanan testlere serpiştirilerek bireylerin puanlarına katılmadan uygulanmasıdır. Bazen uygulanan testin uzunluk ve zaman kısıtlaması ya da yeni maddelerin uygulanan testin içeriğinden çok farklı olması sorun olabileceğinden farklı uygulamalarda gerçekleştirilebilir. Bilgisayar test formatının gerekliliklerinin sağlanamaması ihtimaline karşın ilk aşamada madde havuzu kağıt-kalem formatında oluşturulabilir.

**4. Yeni maddelerden alt testlerin seçilmesi:** Veriler klasik test kuramı (KTK) ve MTK çerçevesinde analiz edilir ve elde edilen bilgilerden yararlanarak madde havuzundan uygun olmayan maddeler düzeltilebilir ya da çıkartılabilir. Wainer, bilgisayar uygulamalarında madde analizinin, geleneksel madde analizi yöntemleri ile MTK tabanlı grafik yöntemlerinin birleşimi ile yapıldığını belirtmiştir. Böyle bir sistem ile maddenin havuza girme durumunu değerlendirme süreci daha basit hale gelmektedir.

KTK çerçevesinde madde gücü ve madde ayırt ediciliği hesaplanırken MTK çerçevesinde a, b ve c parametreleri hesaplanmaktadır. Bireyselleştirilmiş testlerde nitelikli bir madde havuzunun yüksek ayırt



ediciliğe ( $a > 1$ ), madde güçlüğün dikdörtgensel bir dağılıma ve düşük şans parametresine sahip ( $c < 0,2$ ) olması istenir.

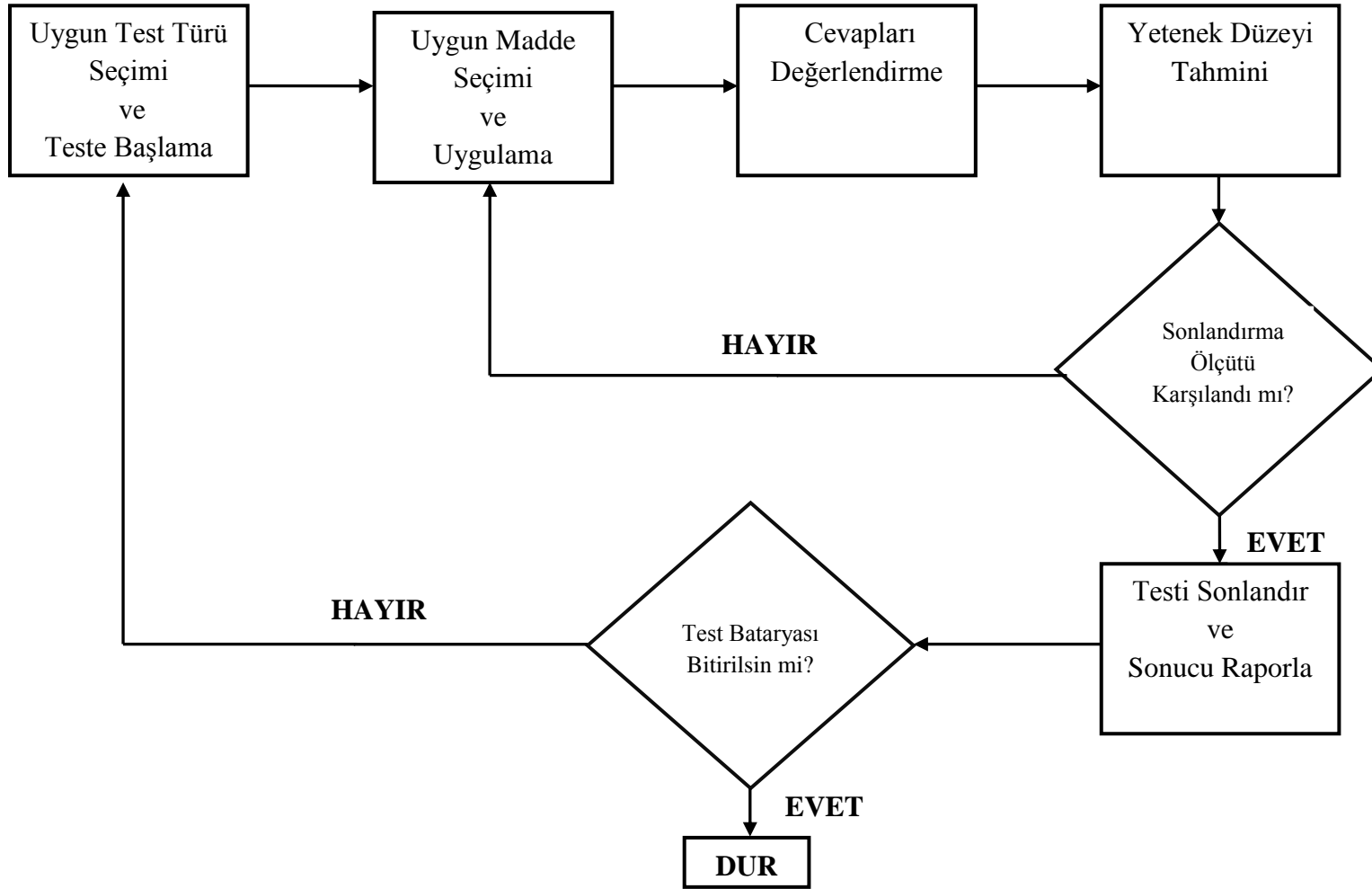
Kingston ve Dorans (1985)'a göre bireysel maddelerin değerlendirilmesi için geliştirilen yöntemlerden birisi de madde-yetenek regresyonun analizidir.

**5. Bilgisayarlı bireyselleştirilmiş ve kağıt kalem testlerinde içerik dengesinin karşılaştırılması:** BBT yaklaşımında her bireyin aldığı test formu farklı olmasına rağmen bu farklı formlardaki maddeler tek bir madde havuzundan seçilmektedir. Bu nedenle BBT ve kağıt-kalem testlerinin madde havuzları birbirleri ile karşılaştırılabilir.

**6. Kağıt-kalem testinin bilgisayarlı bireyselleştirilmiş teste dönüştürülmesi:** Burada testin yönergesinin açık ve anlaşılır yazılması ve testi alan bireylerin bilgisayarı kullanmadaki farklılığının puanlarını etkilememesi gibi bir çok özelliğin göz önünde bulundurulması gerekmektedir.

### 1.1.2. Test Algoritması

BBT uygulamasının daha önce bahsedilen ilk maddenin seçimine ilişkin kural, bir sonraki maddenin seçimine karar verecek madde seçim algoritması, bireyin önceki cevaplarını temel alan yetenek kestirim süreci ve testin sonlandırılmasına ilişkin kural bileşenleri bu başlık altında incelenecektir.



Şekil 1.1. Bireyselleştirilmiş Bilgisayarlı Test Algoritmasının Şeması

Thissen ve Mislevy (2000) ve Eggen (2004)'den uyarlanan BBT için test algoritmasının akış şeması Şekil 1.1.'de yer almaktadır. Şekil 1.1. incelendiğinde; test algoritmasının temel olarak; teste başlama, devam etme ve testi sonlandırma bölümlerinden oluştuğu görülmektedir.

Çalışmanın bundan sonraki kısmında, test algoritmasının bölümleri açıklanmaya çalışılmıştır.

### **1.1.2.1. Teste Başlama**

Klasik testlerde genellikle uygulanacak ilk madde testin en kolay maddesidir. Bu nedenle pek çok birey teste başarı ile başlamaktadır. İyi bir ölçme için ilk başarının önemli olduğuna dair psikometrik kanıt olmamasına rağmen test geliştiricileri genellikle teste zor soru ile başlamanın kaygıyı yükselttiği ve test performansını olumsuz etkilediğini belirtmektedirler. Klasik testlere benzer şekilde ilk sorunun nispeten daha kolay bir soru olarak tercih edilmesi bireyin motivasyonu için BBT'lerde de tercih edilebilmektedir. Ancak bu yaklaşımın dezavantajı uygun güçlük düzeyinde madde bulmanın zor olmasıdır. Bunun üstesinden gelebilmek için daha kolay maddelerin sayısı artırılabilir fakat böyle maddelerin madde havuzundaki oranı da artacağı için madde havuzu kalitesi olumsuz etkilenmektedir. Buna karşın uygulanacak ilk maddenin güçlüğü orta düzeyde ise BBT psikometrik açıdan daha etkin olmaktadır. Çünkü bireyin yeteneği hakkında ön bilginiz yoksa en iyi tahmin orta düzeyde yeteneğe sahip olduğunu kabul etmektir. (Mills ve Stocking, 1996).

BBT'de teste başlamak için kullanılan farklı bir yöntem ise bireyin yeteneğine en uygun soru ile başlanmasıdır. Fakat daha önceki cevaplara dayandırılmış bir yetenek düzeyi olmadığı için testin hangi madde ile başlayacağı belirsizdir. Bu nedenle, benzer testleri almış deneklerin yetenek puanlarının ortalaması ilk sorunun güçlük düzeyini belirlemek için kullanılabilir. Ayrıca deneklerin yaş, aldıkları kurslar gibi demografik bilgileri bilindiği durumlarda, bu özellikleri yansıtan daha küçük farklı bir grubun yetenek ortalaması ilk tahmin olarak kullanılabilir. Ancak testte farklı yetenek düzeyi tahminleri ile başlandığında bireylerin bu demografik bilgilerinin daha sonraki yetenek tahminlerinde kullanılıp

kullanılmayacağı problem olarak karşımıza çıkmaktadır. Çünkü benzer demografik bilgilere sahip iki alt grubun yetenek düzeyleri birbirinden farklı olabilir. Örneğin birisi daha yüksek yetenek düzeyine diğeri ise düşük yetenek düzeyine sahip iki alt grup olduğunu varsayalım. Düşük yetenek düzeyindeki grupta yer alan bir bireye teste başlarken, diğeri grubun ortalama yetenek düzeyindeki bir madde uygulandığında bu birey diğeri gruptaki bir bireyden daha düşük puan alacaktır. Bu durum, bireye uygulanan madde sayısı arttıkça, ilk maddenin hangi güçlük düzeyinde olduğunun önemi azalsa da sorun teşkil edebilir. Bu sorun yardımcı bilgilerin ilk veya ilk birkaç maddeyi belirlemek için uygulanması ve daha sonra ihmal edilmesi ile giderilebilir (Thissen ve Mislevy, 2000).

Bireylerin daha önce aldığı teste benzer başka bir test alması durumunda ise, bu iki test arasındaki ilişkinin çıkarılması da yetenek düzeyi tahmininde kullanılabilir. Örneğin kelime testinden başarılı olmuş bir birey genellikle okuduğunu anlama testinde de başarılı olur. Ancak bu uygulamalarda, ilk yetenek düzeyi tahminlerinin nihai yetenek tahminlerini olumsuz etkileyip etkilemediği sorunu beraberinde getirmektedir. Bu durum, yeteneği kestirme yöntemine göre farklılık gösterir. Örneğin yetenek düzeyi, olabilirlik yaklaşımına göre kestirildiğinde problem olmazken, örnekleme kuramına dayalı olarak kestirildiğinde problem olabilir. Ayrıca aynı maddenin bireye çok sayıda uygulanması, madde kullanım sıklığını (item exposure) artırdığı için testin geçerliği azalır. Çünkü aynı soruya daha önce muhatap olan bir bireyin bu soru ile tekrar karşılaştığında verdiği doğru cevap, gerçek yetenek düzeyinden daha yüksek yetenek düzeyine sahipmiş gibi yorumlanır ve bu yüzden sıradaki maddelerin seçimi gerçeğe uygun olmayan bir şekilde gerçekleşir (Thissen ve Mislevy, 2000).

#### **1.1.2.2. Devam Etme**

Kullanılan yetenek kestirimi yöntemine göre ilk maddenin seçiminin ardından takip edilecek farklı kestirim yöntemleri mevcuttur. Bireyin yeteneğini tahmin etmede kullanılan iki temel yöntem; Maksimum Likelihood Estimation ve Bayes Yöntemi (Expected a Posteriori ve Maksimum a Posteriori) dir. Bireyin yeteneğini tahmin etmede bu kestirim yöntemlerinden Maksimum Likelihood Estimation (MLE)

yöntemi kullanıldığında kestirim için en az bir doğru, bir de yanlış yanıt verilmesi gerekmektedir. Bu nedenle, bu kestirim yöntemi seçildiğinde bilgisayarın sıradaki maddeyi bireyin bir doğru veya bir yanlış cevap verecek şekilde seçmesi gerekmektedir. Yani bireyin ilk maddeye verdiği cevap doğru ise, bilgisayar daha zor bir madde sorarak bireyin yanlış yanıt verme olasılığını artırmaktadır. MLE yetenek düzeyi tahmininde likelihood olasılık fonksiyonunun modu veya maksimum değeri kullanılır (Thissen ve Mislevy, 1982). Log-likelihood fonksiyonunun maksimum değeri bulunduktan sonra fonksiyonun birinci türevi sıfıra eşitlenerek  $\theta_s$  için çözümlenir. MLE yöntemi Newton-Raphson iterasyon sürecini kullanmaktadır. Bu süreçte ilk olarak  $\theta_s$  değeri için bir başlangıç değeri seçilmektedir. (örneğin  $\theta_s=0.0$ ). İkinci adımda ise  $\theta_s$ 'nin bu değerinde log-likelihood fonksiyonunun birinci ve ikinci türevleri hesaplandıktan sonra birinci ve ikinci türevlerin oranı olan  $\epsilon$  hesaplanır. Hesaplanan  $\epsilon$  değeri ilk tahmin edilen  $\theta_s$  değerinden çıkarılarak bu yeni değer  $\theta_s$  için yeni başlangıç değeri olarak kullanılmaktadır. Bu iteratif süreç  $\epsilon$  değeri belirlenen değer altına düşene dek sürdürülmektedir (Embretson ve Reise, 2000).

Eğer birey tüm maddeleri doğru veya yanlış cevaplarsa; MLE yöntemleri yeteneğin doğru tahminde yetersiz kalabilir ve yetenek düzeyini negatif veya pozitif sonsuzlukta tahmin edebilir. Diğer uç noktalarda yer alan bireylere ait cevap örüntüleri de 3 Parametrelili Lojistik modeli kullanımında bu tahmin durumu ile karşılaşabilirler. Bu sınırlılık Bayes kestirim yöntemleri kullanılarak giderilebilir. Bir başka yetenek kestirim yöntemi olan Bayes yöntemlerde bireyin yeteneğinin kestirilmesi için MLE de olduğu gibi bir doğru veya bir yanlış yanıt verme zorunluluğu yoktur. . Bu yöntemlerde bireylerin yetenek tahminleri hakkındaki ön bilgileri likelihood fonksiyonu ile ilişkilendirilir. Bütün bireyler için tek düze bir dağılım kullanılırsa, yetenek tahminleri MLE ile tahmin edilecek değerlere benzer değerler elde edilir (Hambleton, Swaminathan ve Rogers, 1991).

Bayes temelli yetenek kestirim yöntemlerinden BBT uygulamalarında en çok kullanılan Expected a Posteriori (EAP) ve Maksimum a Posteriori (MAP) burada kısaca açıklanacaktır. Yöntemlerin birbirinden en temel farkı bireylerin yeteneğini tahminde; önsel ve sonsal dağılım biçimleri ve sonsal dağılımın özelliğine ilişkin temel varsayımlardır (Wainer ve Mislevy, 2000). Expected a Posteriori (EAP)

sonsal dağılımın ortalamasını ve varyansını bulur ancak her adımda normallik varsayımı yoktur. EAP kestirim yönteminde bireyin yetenek düzeyi hemen ilk maddeye verilen yanıttan sonra kestirilebilir. Bundan dolayı EAP bireyselleştirilmiş test uygulamalarında kullanılabilir (Thissen ve Mislevy, 1982). EAP kestirim yönteminin dezavantajı, EAP'ta standart hata daha yüksek olduğu için, MLE kestirim yöntemi ile elde edilen bir testle aynı güvenilirlik düzeyine sahip bir test elde etmek istendiğinde daha çok soru sorulmasını gerektirmesidir. Maksimum a Posteriori (MAP) Samejima (1996) tarafından,  $\theta$  tahminine ilişkin önsel dağılım kullanılacağı zamanda MLE'ye alternatif olarak önerilmiştir. MAP, EAP yöntemindeki gibi benzer sonsal dağılımı ilişkilendirir ancak yetenek dağılımının ortalamasından ziyade mod değerini kullanmaktadır. Lord (1986) ve Warm (1989) çalışmalarında MAP yetenek tahminin geleneksel uygulamalarda ve BBT uygulamalarında ciddi yanlı sonuç verdiğini göstermişlerdir (akt: Wang ve Vispoel, 1998).

Hangi yetenek kestirim yönteminin daha iyi olduğunu belirlemede, BBT uygulamasının diğer bileşenleri ile birlikte düşünülmelidir. Madde seçiminde hangi yöntemin kullanılacağına karar verme bu aşamada kritik bir rol oynamaktadır. Madde havuzu, bireyin demografik bilgilerinin tahmin sürecinde kullanılıp kullanılmayacağı, madde kullanım sıklığının belirlenmesindeki yöntem seçimi, madde seçiminde içerik dengesinin göz önüne alınması gibi değişkenler ise yetenek kestirim yöntemi seçiminde etkili olan boyutlardır (Linden ve Glas, 2002).

Bireylerin ilk yetenek düzeyi kestirilmesinden sonraki aşama bundan sonra uygulanacak maddelerin seçilmesidir. BBT uygulamasındaki madde seçiminde Maksimum Bilgi Yöntemi (Maximum Information Method) ve Bayes Yöntemi yaygın olarak kullanılmaktadır (Çıkrıkçı Demirtaşlı, 1999). Maksimum bilgi yönteminde, belirli bir yetenek düzeyi için en yüksek bilgiyi sağlayarak yetenek düzeyinin standart hatasına beklenen katkısı minimum olan maddeler seçilir. Maddelerin verdiği bilgi toplanabilir olduğu için bu yöntem seçim kriteri olarak kullanılabilir. Her madde toplam bilgiye katkıda bulunur ve bilgi genel itibarıyla madde parametrelerine dayalıdır (Meijer ve Nering, 1999).

Bayes yöntemlerde; her aşamada bilinmeyen birey parametresinin bir önceki sonsal dağılımı, yeni önsel dağılım olarak hizmet eder. Sonsal varyans, bireyin yetenek düzeyinde kabul edilebilir belirsizlik seviyesine ulaşıncaya kadar devam edilir ve böylece yetenek tahmininin beklenen sonsal varyansı minimize edilebilir (Yoo, 2011). Yani burada yetenek dağılımının varyansını minimize edebilecek maddeler seçilmektedir. Bu yöntemin başarısı önceki dağılımın etkililiğine bağlıdır. Ayrıca bireylere yöneltilen soru sayısının artması da, önceki dağılımlarda yapılan yetenek kestiriminin etkisini azaltacağı için önemlidir.

Linden ve Glas (2002) ise madde seçim yöntemlerini klasik ve modern yöntemler olarak ayırmışlardır. Klasik yöntemlerde madde seçimi maksimum bilgiye dayanmaktadır. Klasik yöntemler kendi içerisinde Maximum-Information Criterion ve Owen's Approximate Bayesian Procedure olarak ikiye ayrılmıştır. Klasik yöntemlerden maksimum bilgi yönteminde son kestirilen yetenek düzeyi için en iyi performansı gösterdiği için seçilen madde, gerçek yetenek düzeyi için daha zayıf performans gösterebilir. Madde seçim kuralları çerçevesinde optimal madde seçimi, madde parametrelerine ait uç değerleri ve geniş tahmin hatalarını da beraberinde getirebilir ancak modern yöntemlerle bu sınırlılığın üstesinden gelinebilir. Modern yöntemler ise kendi içerisinde Maximum Global-Information Criterion, Likelihood-Weighted Information Criterion, Fully Bayesian Criteria, Bayesian Criteria with an Empirical Prior ve Bayesian Criteria with Random Item Parameters olarak beşe ayrılmıştır.

### **1.1.2.3. Testi Sonlandırma**

BT; önceden belirlenmiş madde sayısı uygulandığında, önceden belirlenmiş ölçme kesinliğine (measurement precision) ulaşıldığı zaman veya yine önceden belirlenen süreye ulaşıldığında tamamlanmış kabul edilebilir. Testin sonlandırılmasında genel olarak değişen uzunluk ve sabit uzunluk sonlandırma kuralları kullanılmaktadır. Test geliştiricileri, sonlandırma kuralının seçiminin genel olarak; testin amacına, madde havuzunun özelliğine ve işlevsel kısıtlılıklara dayalı olarak yapıldığını bulmuşlardır (Segall, 2004).

### **1.1.2.3.1.Sabit Uzunluk Sonlandırma Kuralı**

Bu kuralda bireyselleştirilmiş test önceden belirlenmiş sayıda madde uygulanınca sonlanmaktadır. Tüm bireylere testin sonunda ulaşılan ölçme kesinliğinin derecesi göz ardı edilerek aynı sayıda madde uygulanmaktadır. En önemli avantajı kolay olmasıdır. Bu kuralın uygulanmasının sonuçlarından biri de tüm bireyler farklı kesinlik derecelerinde ölçülüyor olması ve genellikle uç yetenek seviyelerinde daha fazla ölçme hatasına neden olmasıdır. Bu duruma ek olarak bireyin yetenek düzeyinde daha az bilgiye sahip maddelerin gereksiz yere uygulanması BBT'nin etkililiğini sınırlayabilir (Choi, Grady ve Dodd, 2010). Segall (2004)'e göre bireyler arasında değişebilen kesinlik indeksleri genellikle MLE veya Bayesian Posterior Variance istatistiklerine dayalıdır.

### **1.1.2.3.2.Değişen Uzunluk Sonlandırma Kuralı**

Bu kuralda bireylerin cevap örüntülerine (item response patterns) bağlı olarak farklı sayıda maddenin uygulanması söz konusudur. Değişen uzunluk sonlandırma kuralını uygulamak iki önemli avantaj sağlamaktadır: ölçmenin etkililiği (efficiency) ve ölçmenin kalitesi (quality). Ölçmenin etkililiği ile bireyin yeteneğinin göreceli olarak daha az madde ile kestirilmesi kastedilmektedir. Bu ölçme için gerekli zamanı azaltmakta ve madde havuzunun kullanımını etkili kılmaktadır (Weiss ve Kingsbury, 1984). Ölçmenin amacı bireylerin yeteneklerini yüksek kesinlikle kestirmektir. Değişen uzunluk sonlandırma kuralında istenilen kesinlik derecesine ulaşmak için bireylere bazen daha fazla madde yöneltilmektedir (Weiss, 1982).

Değişen uzunlukta uygulanan BBT, teorik olarak sabit uzunlukta uygulanan BBT ile eşit veya daha üstün ölçme kalitesi sağlamaktadır. Eğer test puanı belli bir ölçme kesinliğine ulaşmışsa, biraz daha fazla maddenin uygulanması puanın kesinliğini önemli ölçüde etkilememektedir. Test uzunluğu ve kalitesi arasındaki ilişki Spearman-Brown formülüne bakılarak yorumlanabilir çünkü test uzunluğunun değişmesiyle güvenilirliğin nasıl etkilendiği bu formülde açıkça görülmektedir. Eğer BBT, ölçütler çerçevesinde belli bir ölçme kesinliğine ulaştığında sabit uzunlukta uygulanan teste kıyasla daha az madde ile bitirse ölçmenin kalitesinde çok büyük bir farklılık olmayacaktır.



Değişen uzunluk sonlandırma kuralının sabit uzunluğa kıyasla dezavantajı bulunduğu da gözden kaçırılmamalıdır. Bu dezavantaj BBT uygulamasının çok az madde ile sonlandırıldığında ölçme kalitesinin düşük olmasıdır. Madde havuzunda sonlandırma kuralı ile örtüşen bir maddenin kalmaması da başka bir dezavantajdır. Sabit uzunlukta BBT beklendiği gibi sonlandırılırken; değişen uzunlukta BBT de çok fazla maddenin uygulanma ihtimali bulunmaktadır. Ayrıca sabit uzunlukta BBT uygulamalarını sürdürme ve kalite kontrolünü sağlama değişen uzunlukta BBT uygulamasına göre daha kolaydır. Testi alan adaylar sabit uzunlukta testi almanın daha adil olduğunu düşünebilirler. Bu durumda da testte adillik kavramının ölçme etkililiği (measurement precision) ile eş değerde olduğunu testi uygulayanların adaylara belirtmeleri ve tanımlamaları daha doğru olacaktır (Babcock ve Weiss, 2012).

BBT'ler için çok sayıda değişen uzunlukta sonlandırma ölçütleri bulunmaktadır. Bunlar Standart Hata (SH) Sonlandırma Kuralı, Minimum Bilgi (MB) Sonlandırma Kuralı ve Theta Değişimi Sonlandırma Kuralıdır. Sözü edilen sonlandırma kuralları aşağıda açıklanacaktır.

**a) Standart Hata Sonlandırma Kuralı:** Standart Hata Sonlandırma Kuralında bireyin yetenek düzeyi belli bir kesinlik seviyesine ulaşıncaya kadar madde uygulanmaktadır. Bu kural; kabul edilebilir standart hata için yeteri kadar bilgi veren maddelerin bulunmasına ve bireye uymayan maddelerin göreceli olarak az olmasına bağlıdır. Weiss ve Kingsbury (1984)'e göre sonlandırma kuralları içinde en popüler olanı Standart hata sonlandırma kuralıdır.

Madde Tepki Kuramı bağlamında; Maksimum Likelihood puanlamada theta yetenek düzeyinin standart hatası;

$$SEM(\theta) = \frac{1}{\sqrt{-\partial^2 \log L / \partial \theta^2}}$$

ile hesaplanır. Burada

$\bar{\theta}_i$  : tahmin edilen  $\theta$  değerini

log L: log likelihood fonksiyonunu

göstermektedir (Samejima, 1977). Log likelihood fonksiyonuna ait hesaplama aşağıda gösterilmiştir.

$$\log L(x_{1p}, x_{2p}, \dots, x_{np}) = \sum_{i=1}^n x_{ip} \log[P_i(\theta)] + (1 - x_{ip}) \log[Q_i(\theta)]$$

Burada;

i: madde indeksi,

n bireyler tarafından cevaplanan madde sayısı,

$P_i(\theta)$  maddenin doğru cevaplanma olasılığı

$Q_i(\theta)$  maddenin doğru cevaplanmama olasılığını

göstermektedir (Embretson ve Reise, 2000).

Yapılan araştırmalar, BBT uygulamalarında standart hata sonlandırma kuralı ile her bir yetenek düzeyi kestirildiğinde ölçme kesinliğinin yüksek olduğunu göstermektedir. Böyle bir performansın görülebilmesi için, madde havuzunun her bir yetenek düzeyinde standart hata değerinin elde edilmesine imkan vermesi gerekmektedir (Dodd, Koch ve De Ayala, 1993; Wang ve Wang, 2001; Babcock ve Weiss, 2012).

**b) Minimum Bilgi Sonlandırma Kuralı:** Bu kuralda havuzda bireylerin en son yetenek düzeyinde bilgi sağlayabilen madde kalmayıncaya kadar BBT uygulamasının sürdürülmesi söz konusudur (Gialluca ve Weiss, 1979; Maurelli ve Weiss, 1981; akt: Babcock ve Weiss, 2012).

Minimum bilgi sonlandırma kuralında, tahmin edilen Fisher'in madde bilgi fonksiyonu  $I_i(\theta)$ ,

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

ile gösterilir (Samejima, 1977). Burada;

i madde indeksini,

$P'_i(\theta)$  madde karakteristik fonksiyonun ikinci türevini

göstermektedir. Bazı araştırmalar minimum bilgi sonlandırma kuralının etkili bir sonlandırma kuralı olduğunu ortaya koymaktadır (Brown ve Weiss, 1979).

**c) Theta Değişimi Sonlandırma Kuralı:** Bu kural, Theta ( $\Theta$ ) veya Theta'ya yakınsayan yetenek düzeyindeki değişimi ölçüt almaktadır. BBT'de daha fazla madde almak psikometrik bilginin artması ile sonuçlanmaktadır, böylece BBT'de bireye uygulanan her madde sonrasında bireyin yetenek düzeyi değişmektedir. BBT uygulamasının başlangıcında yetenek düzeyindeki değişim fazla iken; genellikle giderek daha azalmaktadır (Weiss ve Kingsbury, 1984). Yetenek düzeyine yakınsama da (convergence of  $\Theta$  estimate) BBT için iyi bir sonlandırma kuralı olabilir (Babcock ve Weiss, 2012).

Bunların dışında madde bankası tükenmesi; madde bankalarının küçük olması ve birey tarafından her maddenin alındığı durumlarda test sonlandırılabilir. Ayrıca BBT yazılımları verilen cevap örüntüsünü sürekli takip eder. Bireyin çok hızlı ya da çok yavaş olması durumunda test yöneticisi bireyle görüşüp testi durdurmak veya ertelemek için karar alabilir. Bireye minimum sayıda madde yönlendirilmediyse veya teste dahil olan tüm konulardan soru gelmediyse testin durdurulması kesinlikle önerilmez (Linacre, 2000).

BBT uygulamalarında kullanılabilecek çok sayıda testi sonlandırma kuralı bulunmaktadır. Bu kuralların her biri testin standart hatası ve yetenek kestirimlerinin yanlılığı üzerinde farklı etkilerde bulunmaktadır (Blais ve Raiche, 2000; Yi, Wang ve Ban, 2001; Hart, Cook, Mioduski, Teal ve Crane, 2006; Choi, Grady ve Dodd, 2011; Babcock ve Weiss, 2012). Testi sonlandırma kuralı ile birlikte madde havuzu, başlama kuralları ve testi sürdürmede kullanılan yöntemler de testin ölçme kesinliğini etkileyebilmektedir. Bu araştırmada farklı sonlandırma kuralları; testten elde edilen RMSE, yanlılık ve uyum değerleri ile test uzunluğu bakımından karşılaştırılacaktır.

## 1.2. Problem Cümlesi

Bireyselleştirilmiş Bilgisayarlı Test uygulamasında ölçme kesinliği ve/veya test uzunluğu

- a. Testin sabit uzunluğuna,
- b. Yetenek düzeyinin standart hatasına,
- c. Yetenek düzeyinin standart hatası ve en az madde sayısına,
- d. Yetenek düzeyindeki mutlak değişime,
- e. Yetenek düzeyindeki mutlak değişime ve en az madde sayısına

dayalı sonlandırma kurallarına göre nasıl bir değişim göstermektedir?

### 1.2.1. Alt Problemler

1. Bireyselleştirilmiş Bilgisayarlı Test uygulamasında, madde havuzu büyüklüğü 250; başlama kuralı  $b=0$  ve yetenek kestirim yöntemi Maksimum Likelihood Estimation seçildiğinde; ölçme kesinliği ve/veya test uzunluğu farklı sonlandırma kurallarına göre nasıl değişmektedir?
2. Bireyselleştirilmiş Bilgisayarlı Test uygulamasında, madde havuzu büyüklüğü 250; başlama kuralı  $-1 < b < +1$  ve yetenek kestirim yöntemi Maksimum Likelihood Estimation seçildiğinde; ölçme kesinliği ve/veya test uzunluğu farklı sonlandırma kurallarına göre nasıl değişmektedir?
3. Bireyselleştirilmiş Bilgisayarlı Test uygulamasında, madde havuzu büyüklüğü 500; başlama kuralı  $b=0$  ve yetenek kestirim yöntemi Maksimum Likelihood Estimation seçildiğinde; ölçme kesinliği ve/veya test uzunluğu farklı sonlandırma kurallarına göre nasıl değişmektedir?
4. Bireyselleştirilmiş Bilgisayarlı Test uygulamasında, madde havuzu büyüklüğü 500; başlama kuralı  $-1 < b < +1$  ve yetenek kestirim yöntemi Maksimum Likelihood Estimation seçildiğinde; ölçme kesinliği ve/veya test uzunluğu farklı sonlandırma kurallarına göre nasıl değişmektedir?
5. Bireyselleştirilmiş Bilgisayarlı Test uygulamasında, madde havuzu büyüklüğü 250; başlama kuralı  $b=0$  ve yetenek kestirim yöntemi Expected

a Posteriori seçildiğinde; ölçme kesinliği ve/veya test uzunluğu farklı sonlandırma kurallarına göre nasıl değişmektedir?

6. Bireyselleştirilmiş Bilgisayarlı Test uygulamasında, madde havuzu büyüklüğü 250; başlama kuralı  $-1 < b < +1$  ve yetenek kestirim yöntemi Expected a Posteriori seçildiğinde; ölçme kesinliği ve/veya test uzunluğu farklı sonlandırma kurallarına göre nasıl değişmektedir?
7. Bireyselleştirilmiş Bilgisayarlı Test uygulamasında, madde havuzu büyüklüğü 500; başlama kuralı  $b=0$  ve yetenek kestirim yöntemi Expected a Posteriori seçildiğinde; ölçme kesinliği ve/veya test uzunluğu farklı sonlandırma kurallarına göre nasıl değişmektedir?
8. Bireyselleştirilmiş Bilgisayarlı Test uygulamasında, madde havuzu büyüklüğü 500; başlama kuralı  $-1 < b < +1$  ve yetenek kestirim yöntemi Expected a Posteriori seçildiğinde; ölçme kesinliği ve/veya test uzunluğu farklı sonlandırma kurallarına göre nasıl değişmektedir?

### **1.3. Sınırlılıklar**

1. Bu çalışma, Madde Tepki Kuramı modellerinden 3 PLM ve madde seçim yöntemlerinden Maksimum Fisher Information (MFI) ile sınırlıdır.
2. Bu çalışmada madde kullanım sıklığı ve içerik dengelemesi (content balancing) göz önüne alınmamıştır.

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

Gelişen teknoloji ile birlikte BBT uygulamaları günümüzde eğitim alanında sıkça kullanılmaktadır. Yurtdışında ve ülkemizde uygulamaları bulunan TOEFL, GRE, GMAT gibi önemli sınavlar BBT uygulaması ile gerçekleşmektedir. BBT'nin sağladığı avantajlar göz önüne alındığında; eğitim alanında ve diğer alanlarda uygulamaların giderek artacak olması bir gerçektir. Ayrıca, ülkemizde de mevcut çeşitli okul, program ve işe almada kullanılan geniş ölçekli test uygulamalarının ve öğrenci değerlendirmede kullanılan kağıt-kalem sınavlarının yerini BBT uygulamalarının alabileceği aşikardır.

Yurtdışında uzun süredir yapılan BBT uygulamalarına ilişkin çalışmalar, ülkemizde sınırlı sayıda olup halen yetersiz kalmaktadır. Araştırmada BBT'nin önemli bir bileşeni olan testi sonlandırma kurallarının farklı değerler almasının ölçme kesinliği ve test uzunluğunu nasıl etkileyebileceği incelenecektir. Ayrıca sonlandırma kurallarının, BBT uygulamasının diğer bileşenleri ile birlikte değerlendirilmesi de araştırma sonuçlarına zenginlik katacaktır. Araştırma kapsamında kullanılan değişkenler çerçevesinde, ölçme kesinliği ve test uzunluğu bağlamında madde havuzu büyüklüğü, başlama kuralı seçimi ve yetenek kestirim yöntemi seçimine dair öneriler sunulmuştur. Özetle; bu çalışma ile test geliştiricilere; BBT'nin güvenilirliği ve geçerliğini artırmasına ilişkin kullanabilecekleri bilgiler verilmesi amaçlanmıştır.

Araştırma kapsamında ele alınan sonlandırma kurallarına ait koşulların; farklı madde havuzu büyüklüğü, başlangıç kuralları ve yetenek kestirim yöntemleri bağlamında karşılaştırılması hem ulusal hem de uluslararası literatüre katkı sağlayacaktır. Araştırmanın ele alınan değişkenler ve koşulları bakımından; ülkemizde ilk olması, araştırmanın önemli bir boyutunu oluşturmaktadır.

## 2. İLGİLİ ARAŞTIRMALAR

Wang, Hanson ve Lau (1999) BBT uygulamalarında farklı sonlandırma kuralları ile yetenek kestirim yöntemlerinin ölçme kesinliğini nasıl etkilediğini belirlemeye çalışmışlardır. Çalışmada sabit uzunluk ve değişen uzunluk sonlandırma kuralı kullanılmıştır. Değişen uzunluk sonlandırma kuralı olarak; sabit sonsal varyans sonlandırma kuralı kullanılmıştır. Sabit uzunluk sonlandırma kuralı olarak 30 madde kullanılmıştır. Sabit sonsal varyans, sonlandırma kuralı olarak ise 0.85 güvenilirlik katsayısına karşılık gelen 0,15 sonsal varyans değeri seçilmiştir. Yetenek kestirim yöntemleri olarak; Maksimum a Posteriori (MAP), beta dağılımını kullanan MAP'nin genişletilmiş (EU-MAP), beta dağılımını kullanan Expected a Posteriori (EU-EAP), Weighted Likelihood Estimation (WLE), Maksimum Likelihood Estimation (MLE) kullanılmıştır. Karşılaştırmalar önsel dağılımın şekline, farklı madde havuzu özelliklerine, yanlılık, RMSE ve standart hataya dayalı gerçekleştirilmiştir. Sabit uzunluk sonlandırma kuralının kullanıldığı durumda; WLE yöntemi testin yanlılığını azaltmada MLE'ye kıyasla daha etkilidir. Ancak değişen uzunluk sonlandırma kuralında, WLE yöntemi ile daha yüksek yanlılık değeri elde edilmiştir. Değişen uzunluk sonlandırma kuralında WLE, EU-EAP veya EU-MAP yöntemlerinin, yanlılığın önemli olduğu durumlarda kullanılabilmesi ifade edilmiştir. WLE yönteminin sabit sonsal varyans (değişen uzunluk) sonlandırma kuralında tercih edilmemesi önerilmiştir. Amacına göre EU-EAP ve EU-MAP yöntemlerinden birinin seçilebileceği belirtilmiştir. Araştırmacılar; yanlılık değeri önemli ise EU-MAP yönteminin, standart hata değerinin önemli olduğu durumda ise EU-EAP yönteminin kullanılabilmesini çalışmada öneri olarak sunmuşlardır.

McBride, Wetzel ve Hetter (2001) en iyi madde seçme stratejisini belirlemede amacıyla madde bankası büyüklüğü ve karakteristiği ile yeteneğin rolünü incelemişlerdir. Çalışma 3 aşamada simülasyon olarak gerçekleştirilmiştir. Birinci aşamada dağılımın ve yetenek kestiriminin etkisi incelenmiştir. İkinci aşamada ise iki farklı madde seçme yönteminin test güvenliğine etkisi incelenmiştir. Üçüncü aşamada değişken ve sabit uzunluklu testlerin farklılığını incelemiştir. Bayes

sonsal varyansı 0,064 ve 0,053 olan 2 deęişken uzunluklu test ve 15 maddelik sabit test -2,25 ve +2,25 yetenek aralığında karşılaştırılmıştır. Ortalama 10 ve 30 madde uzunluęundaki testlerde 0,053 sonlandırma kuralı ölçme kesinlięi açısından daha iyi sonuçlar vermiştir. -2,25 ile +1,75 yetenek ranjında bu sonlandırma kuralı ortalama 15 maddeden daha az madde ile sonlandırılmıştır. Ölçme etkililięi daha düşük olan 0,064 sonlandırma kuralında test uzunluęu ise ortalama 10 ve 30 madde arasındadır. Yetenek düzeyi +0,25'in altında ortalama test uzunluęu 18 iken; +0,25 ile +1,75 arasında ise ortalama test uzunluęu 15 maddelik sabit uzunluktan oluşan testten daha kısa olmaktadır. Yetenek düzeyi +1,75 ve üzerinde ise test uzunluęu yetenek düzeyi ile birlikte istenmeyen şekilde artmıştır.

Yi, Wang ve Ban (2001) BBT uygulamasında farklı sonlandırma kurallarının ve ölçek dönüşümlerinin farklı yetenek kestirim yöntemleri ile ölçme kesinlięine etkisini iki aşamada incelemişlerdir. Sonlandırma kuralı olarak; sabit uzunluk kuralı ile deęişen uzunluk sonlandırma kurallarından olan standart hata ve minimum bilgi kurallarını karşılaştırmışlardır. Yetenek kestirim yöntemlerinden Maksimum Likelihood Estimation (MLE), Weighted Likelihood Estimation (WLE), Expected a Posteriori (EAP) ve Maksimum a Posteriori (MAP) kullanılmıştır. Parametreleri farklı (gerçek data ve gerçek datadan türetilen) iki madde havuzu kullanılmıştır. Başlama kuralı olarak  $b=0$  seçilmiştir. Madde seçim yöntemi olarak Maksimum Bilgi kullanılmıştır. Sabit uzunluk sonlandırma kuralında 30 madde; standart hata sonlandırma kuralında MLE ve WLE için 0,32(SE) ve Bayes yöntemler için ise önsel standart sapma 0,32 olarak belirlenmiştir. Tüm test için maksimum test uzunluęu 60 madde olarak belirlenmiştir. Simülasyon çalışması 0,32 (SE)'ye veya daha düşük deęere ulaştığında ya da maksimum 60 maddeye ulaşıldığında sonlandırılmıştır. Sonuçlar hata göstergelerine (RMSE, yanlılık ve yetenek düzeyinin standart hatası) göre yorumlanmıştır. Çalışmanın sonucunda sonlandırma kurallarının ve ölçek dönüşümlerinin yetenek tahmini üzerinde önemli etkisi bulunduęu belirtilmiştir. Genel olarak farklı sonlandırma kuralları kullanımının hata göstergeleri üzerinde etkisi bulunmuştur. Standart hata sonlandırma kuralının, düşük ve yüksek yetenek düzeylerinde (özellikle madde havuzunun tüm bireylerin yetenek düzeylerini kapsayacak kadar geniş ranjda madde içermediğinde) testin etkililięini azaltabildięi ifade edilmiştir. Bilgi sonlandırma



kuralı kullanıldığında; hata göstergelerinin, çalışmanın koşullarında farklı değerler aldığı bulunmuştur. MLE yöntemi kullanıldığında yanlılığın yönü ve WLE'nin etkililiğinin farklı sonlandırma kurallarında değişmekte olduğu belirtilmiştir WLE yöntemi; sabit sonlandırma ve bilgi sonlandırma kuralında (araştırmada kullanılan tüm ölçeklerde) MLE'ye göre daha iyi sonuçlar üretmiştir. Sabit sonlandırma kuralı kullanıldığı zaman WLE yöntemi ile elde edilen yanlılık değeri MLE'ye kıyasla biraz yüksek iken, standart hata değeri daha düşüktür. Tüm Bayes kestirim yöntemleri ile MLE ve WLE'ye kıyasla daha düşük standart hata ancak daha yüksek yanlılık değeri elde edilmiştir. Ayrıca iki farklı madde havuzu karakteristiğinin madde sayısı kadar hata göstergelerini etkilediği elde edilmiştir.

İşeri (2002) Ortaöğretim Kurumları Seçme Yerleştirme Sınavı ve Özel Okullar Sınavının matematik kısımlarının oluşturduğu soru bankasını kullanarak matematik başarısının BBT ile ölçülmesini gerçekleştirmiştir. Araştırmanın sonucunda BBT'lerin yetenek düzeyini daha az soru ile güvenilir biçimde ölçmekte olduğu bulunmuştur. Öğrencilerin cevaplarını kontrol edebildiği takdirde, yüksek yetenek düzeyindeki öğrencilerin düşük yetenek düzeyinde olarak kestirildi. Bayes yönteminin daha iyi kestirim sağladığı, sabit sayılı ve sabit hatalı sonlandırmanın her ikisinin de iyi sonuçlar sağladığı bulunmuştur.

Wang ve Wang (2001) değişik yetenek kestirim yöntemlerini BBT uygulamalarında çoklu puanlanan (genelleştirilmiş kısmi puanlanan) model için karşılaştırmışlardır. Monte Carlo ile simülasyon çalışması gerçekleştirmişlerdir. Yetenek kestirim yöntemlerinden Warm's Weighted likelihood Estimation (WLE), Expected a Posteriori (EAP) ve Maksimum a Posteriori (MAP) kullanılmıştır. Burada sonlandırma kurallarının etkisi de belirlenmeye çalışılmıştır. Sabit uzunluk ve sabit güvenilirlik sonlandırma kuralları kullanılmıştır. Sabit uzunluk sonlandırma kuralında 5, 10, 15 ve 20 maddelik dört farklı uzunlukta test belirlenmiştir. Sabit test güvenilirliği için ise; 0,7; 0,8 ve 0,9 güvenilirlik katsayılarına karşılık gelen 0,55(SE); 0,45(SE) ve 0,32(SE) değerleri seçilmiştir. Maksimum test uzunluğu 33 madde ile sınırlandırılmıştır. Çalışmada üç farklı madde havuzu kullanılmıştır. İlki; 3 ile 5 arasında değişen kategoriye sahip 263 maddeden oluşan gerçek data, diğer ikisi ise buradan elde edilen değerlerden türetilen 66 ve 33 maddeden oluşan havuzlardır. Araştırmanın bağımlı değişkenlerini RMSE, yanlılık, yetenek

kestiriminin standart hatası ve uygulama etkililiği (belirlenen standart hata değerine ulaşınca kadar uygulanan madde sayısı) oluşturmaktadır. Sonuçlar sabit uzunluk ve sabit güvenilirlik sonlandırma kuralına göre yapılan; ANOVA analizine göre yorumlanmıştır. Buna göre yetenek kestirim yöntemleri; yanlılık değerinin belirlenmesinde sabit uzunluk sonlandırma kuralında %54, sabit güvenilirlik sonlandırma kuralında ise %74 etkilemektedir. Yetenek kestirim yönteminin standart hatasının belirlenmesinde sabit uzunluk sonlandırma kuralında %16, sabit güvenilirlik sonlandırma kuralında ise %2 etkilemektedir. RMSE değerinin belirlenmesinde ise sabit uzunluk sonlandırma kuralında %5, sabit güvenilirlik sonlandırma kuralında ise %6 etkilemektedir. Test uzunluğu ise; sabit uzunluk sonlandırma kuralında, yanlılığın belirlenmesinde %31, yetenek kestiriminin standart hatasının belirlenmesinde %66, RMSE değerinin belirlenmesinde ise %31 etkilemektedir. Sabit güvenilirlik sonlandırma kuralında ise güvenilirlik değeri yanlılık katsayısının belirlenmesinde %15, yetenek kestiriminin standart hatasının belirlenmesinde %59 ve RMSE değerinin belirlenmesinde ise %88 etkilemektedir. WLE yöntemi sabit uzunluk sonlandırma kuralında MLE'ye kıyasla tüm hata göstergelerinde, EAP ve MAP'a kıyasla da yanlılık değerinde daha iyi sonuçlar vermiştir. MLE yöntemi ile Bayes yöntemlere kıyasla daha düşük yanlılık değeri elde edilmiştir. Sonlandırma kurallarının özellikle WLE ve MLE yöntemlerinde, bağımlı değişkenler üzerinde önemli etkiye sahip olduğu belirtilmiştir. Araştırmacılar ayrıca, sonlandırma kurallarının bağımlı değişkenler üzerindeki etkisinin, madde havuzunun etkisinden daha fazla olduğunu belirtmişlerdir.

Blais ve Raiche (2002) BBT'ler de iki farklı sonlandırma kuralına göre yetenek dağılımlarının değişimini incelemişlerdir. Rasch modeli kullanılarak normal dağılıma sahip 2000 kişilik veri üretilmiştir. Başlangıç kuralı olarak  $b=0$ , madde seçme kuralı olarak Urry yöntemi, yetenek tahmininde Bayesi kestirim yöntemi belirlenmiştir. Maksimum standart hata (0,20-0,80 aralığında değişen) ve madde sayısına dayalı (1-60 aralığında değişen) sonlandırma kuralına göre dağılımların yetenek düzey ranjları ve standart hataları ile basıklık ve çarpıklık katsayıları incelenmiştir. Standart hata sonlandırma kuralına göre SH'nin 0,40 ve altında; basıklık ve çarpıklık değerlerinin normal dağılıma yakın olduğu ve yetenek düzey aralığının ise daha geniş ranjda olduğu belirlenmiştir. Madde sayısına dayalı

sonlandırma kuralında ise 13 ve daha fazla maddenin uygulanması durumunda dağılımın normale yaklaştığı belirlenmiştir.

Simms ve Clark (2005) birden çok alt ölçekten oluşan SNAP (Schedule for Nonadaptive and Adaptive Personality) testinin BBT formunun geçerliğini sağlamaya çalışmışlardır. SNAP testinin BBT uygulamasında 2 aşamalı sonlandırma kuralı kullanmışlardır. İlk aşamada bireye uygulanacak minimum madde sayısı seçilmiş olup ikinci aşamada test ise değişen uzunluk sonlandırma kuralı seçilmiştir. Değişen uzunluk kuralı olarak minimum bilgi veya standart hata kuralları belirlenmiştir. Bireyler için değişen uzunluk sonlandırma kurallarından hangisi ilk olarak sağlanmış ise test sonlandırılmıştır. Araştırmacılar testi alan bireylerin %82'sinin maksimum madde sayısını aldıklarını hatta birkaçının madde havuzunun büyüklüğünü aşan sayıda madde alabileceği belirlenmiştir. Farklı sonlandırma kurallarının bazı alt ölçekler için zayıf psikometrik özelliğe sahip olabileceği belirtilmiştir.

Ivie (2007) çalışmasını iki aşamada yürütmüştür. Çalışmanın amacı test inceleme firmalarının ilk 5 veya 10 maddeye daha fazla dikkat ve süre ayrılması halinde bireyin yetenek tahmininin daha iyi olacağı yönündeki iddialarının geçerliğinin araştırılmasıdır. Birinci çalışma testin nasıl çalıştığına ve puanını nasıl iyileştirdiğine dair farklı miktarlardaki bilginin etkisini incelemektedir. Sonuç olarak test alma stratejisine sahip bireylerin performansının sahip olmayan bireylerden daha yüksek olduğu belirtilmiştir. İkinci çalışmada BBT'nin kararlılığını ve asıl yetenek düzeyi tahmininin belirli test parametreleri madde bankası parametreleri (madde havuzu büyüklüğü, ayırt edicilik parametresi ve tahmin parametresinin); birey parametreleri (bireyin yeteneğinin gerçek değerinden fazla olup olmadığını) ve test algoritması parametreleri değiştiğinde kararlılığını araştıran bir dizi simülasyon çalışması yapılmıştır. Madde havuzu genişliği arttıkça daha yüksek ayırt edicilik parametrelerinin ve daha düşük şans parametrelerinin daha kararlı yetenek düzeyi tahmini sağladığı görülmüştür.

Riley, Conrad, Bezruczko ve Dennis (2007) çalışmalarında GAIN-SPS (Global Appraisal of Individual Needs Substance Problem Scale) ölçeğinin BBT formunun farklı yetenek düzeyindeki bireyler için başlama ve sonlandırma kurallarının etkisini

incelemiştir. Bu ölçek; madde bağımlılık düzeyini belirlemede kullanılmakta olup, tamamlanması ise 1-2 saatte değişmektedir. Araştırmacılar farklı yetenek düzeyindeki bireyler için farklı başlama ve sonlandırma kurallarının kullanıldığı test oluşturmuşlardır. BBT sonuçları ortalama standart hata, ortalama madde sayısı, yetenek parametrelerindeki yanlışlık ve RMSE değerleri ve bireylerin sınıflama tutarlığı ve tüm ölçek maddelerinin uygulandığı durum ile korelasyonu bağlamında değerlendirilmiştir. Orta düzeyde yetenek düzeyleri için 0,35 logit belirlemişler ve ayrıca üst ve alt yetenek düzeyleri için 0,50 ve 0,60 ve 0,75 logitlik ölçme hatasına dayalı sonlandırma kurallarını kullanmışlardır. Araştırmanın sonucunda farklı yetenek düzeyleri için değişen sonlandırma kurallarının verilen maddelerin oranını anlamlı olarak azalttığı belirlenmiştir. %13 ve %66 indirgeme oranını rapor etmişlerdir.

Wang (2009) BBT'lerde madde bankasının psikometrik özelliklerini, farklı test yöntemlerinin yetenek kestirimine etkisini, klasik kısa yanıt testler ile yetenek tahminlerini karşılaştırmıştır. Farklı test yöntemleri ile BBT'lerden elde edilen yetenek parametrelerinde tekrarlı ölçümler için MANOVA yapılmıştır. Yapısal eşitlik modellemesi ile tüm test, BBT ve kısa testlerden elde edilen yetenek tahminleri arasındaki korelasyon incelenmiştir. BBT ve kısa formdan elde edilen yetenek kestirimlerinin standart hatası ilişkili örneklem t testi yapılmıştır. Sonuçta madde bankası tek boyutlu bulunmuş, BBT'ler için EAP kestirim yöntemi, tüm test için MLE kestirim yönteminden daha karşılaştırılabilir ve daha güvenilir bulunmuştur. Ayrıca BBT'ler, kısa formlarına göre daha farklı ve istatistiksel olarak daha anlamlı tahmin yöntemleri önermektedir.

Kalender (2011) farklı BBT yöntemi ile elde edilen yetenek kestirimlerini farklı yetenek kestirim ve sonlandırma kurallarını dikkate alarak Öğrenci Seçme Sınavı (ÖSS) fen alt testinin kağıt kalem formatı sonuçları ile karşılaştırmıştır. Çalışma iki aşamada yürütülmüştür. ÖSS'nin BBT ve kağıt-kalem formatlarından elde edilen yetenek kestirimlerinde simülasyon programı uygulamıştır. Yetenek kestirim yöntemi MLE ve EAP; sonlandırma kuralı olarak ise standart hata ile sabit uzunluk kullanılmıştır. Burada BBT uygulamasının ÖSS için EAP yetenek kestirim yöntemi ile 0,30 ve daha yüksek standart hata kuralı ile kullanılabileceği bulunmuştur. İki formattan elde edilen yetenek kestirimleri arasındaki ilişki 0.95 bulunmuştur.

Çalışmanın ikinci aşamasında BBT uygulaması bir grup öğrenciye uygulanmıştır. Bu uygulamada iki formattan elde edilen yetenek kestirimleri arasındaki ilişki 0.74 bulunmuştur. Her aşamada da BBT uygulaması ile soru sayısında düşüş kaydedilmiştir. Sonuç olarak BBT'lerin kağıt-kalem testlerine oranla daha yüksek güvenilirliğe ve yetenek kestirimlerini ise daha az soru ile sağladığını ifade etmişlerdir.

Choi, Grady ve Dodd (2011) BBT'de kullanılmak üzere yeni durdurma kuralı önerisinde bulunmuşlardır. Özellikle sonlandırma kurallarında sıklıkla rastlanılan bir problem olan ölçme etkililiği ve kesinliğinin test sürecinde dengelenmesi durumu göz önünde bulundurulmuştur. Yeni durdurma kuralı olan tahmin edilen standart hata varyansı (TSHV) (PSER- predictive standart error variance) durdurma kuralı; teste yeni maddelerin uygulanması sonucunda tahmin edilen varyansı kullanarak standart hatadaki azalmayı belirlemektedir. Bu yeni durdurma kuralı yeni maddelerin uygulanmasından kaynaklanan ölçüm kesinliğindeki tahmin edilen değişimi dikkate alarak ölçüm kesinliği ve test etkililiği arasındaki dengeyi sağlamaya çalışmaktadır. Ayrıca; minimum bilgi sonlandırma kuralı testte, o ana kadar ölçülen maddelerdeki psikometrik bilgiyi dikkate almaz iken TSHV dikkate almaktadır. Çalışmada uygulanan madde sayısına dayalı olarak TSHV'nin özellikleri araştırılmıştır. Uygulanan madde sayısının artırılması durumunda minimum standart hata sonlandırma kuralı ve minimum bilgi sonlandırma kuralı ile yetenek düzeyi tahmininin duyarlılığı çerçevesinde karşılaştırılmıştır. Çalışmada simülasyon ile çoklu puanlanan madde havuzu oluşturulmuştur. Her üç sonlandırma kuralı ile elde edilen RMSE ve yanlılık değerleri farklı durumlarda karşılaştırılmıştır. Bunlar madde havuzları (30 ve 90), yetenek düzeylerinin dağılımları ( $\theta = 0$  ve  $+1,5$ 'a göre simetrik dağılım), Standart hata ve TSHV'ye göre farklı zamanlar da testin sonlandırılması (yüksek durum-SH' ya göre test sonlandırılmış; düşük durum-TSHV' ye göre test sonlandırılmış) dir. Sonuçta TSHV'nin çoğu koşulda diğer sonlandırma kurallarına göre daha düşük RMSE değeri verdiği; aynı RMSE değerine diğerlerine kıyasla daha az ortalama madde ile ulaştığı bulunmuştur. Test bilgi fonksiyonunun maksimum olduğu noktada en düşük RMSE değerine ulaştığının bulunması da çalışmanın önemli bulguları arasındadır.

Babcock ve Weiss (2012) farklı koşullar altında sabit uzunlukta testi sonlandırma kuralları ile değişen uzunlukta sonlandırma kurallarından elde edilen sonuçları karşılaştırmışlardır. Farklı sayıda maddeden ve bilgi şekillerinden (information shapes) oluşan test bankalarında çeşitli sonlandırma kurallarının işleyişini belirlemek için Monte Carlo simülasyon çalışması yürütmüşlerdir. Madde sayısına göre 100 ve 500 maddelik; b parametresinin tek biçimli dağılması (uniform) veya pik yapmasına (peaked) göre eşleştirilmiş olmak üzere toplam dört tip madde havuzu ile 13 farklı yetenek düzeyinde ( $-3 < \Theta < +3$ ) 1000 tane olmak üzere toplam 13000 birey parametresi üretilmiştir. Burada standart hata, maksimum bilgi, theta değişim (theta yakınsama) ve sabit uzunluk sonlandırma kuralları kullanılmıştır. Ayrıca standart hata ile maksimum bilgi sonlandırma kurallarının iki kombinasyonu kullanılmıştır. Başlangıç kuralı olarak  $\Theta=0$  düzeyi belirlenmiştir. Madde seçmede Maksimum Fisher Information ve yetenek tahmininde ise Maksimum Likelihood yöntemi kullanılmıştır. Araştırmada her bir madde havuzu için sonlandırma kuralları, RMSE ve yanlılık değerine göre karşılaştırılmıştır. Sonuçta literatürde yaygın olarak elde edilen sonuca zıt olarak; değişen uzunluk sonlandırma kuralından elde edilen sonuçların sabit uzunluk sonlandırma kuralından elde edilenlere eşit veya biraz daha iyi olduğu bulunmuştur. Ayrıca düşük sayıda madde ile sonlandırılan testlerde özellikle düşük yetenek düzeyine sahip bireylerin yetenek seviyelerinin tahminlerinde kesinliğin düşük olduğu ve yeterince uzun testlerde daha fazla madde uygulanmasının ise kesinliği çok büyük miktarlarda artırmadığı belirtilmiştir. Çalışma sonucunda elde edilen önemli bir bulgu ise BBT uygulamasının nasıl sonlandırıldığı önemli değil ancak çok az sayıda madde uygulanmasının olumsuz etkiye neden olduğu söylenebilir. Özellikle geniş madde havuzlarında  $SH=0,38$  sonlandırma şartından, 10 maddeden daha az madde ile testin sonlandırılmasında daha yüksek RMSE ve yanlılık değerleri elde edilmiştir. Standart Hata sonlandırma kuralında değerin 0,315 ve aşağısında kullanılması önerilmiştir. Bu kuralda özellikle düşük yetenek düzeyindeki bireylerde sabit uzunluk sonlandırma kuralına göre daha yüksek kesinlikte ölçümler yapılabildiği de belirtilmiştir. Yetenek düzeyi değişimde ise Standart Hata sonlandırma kuralına yakın sonuçlar elde edilmiş ve özellikle geniş madde havuzunda Maksimum Bilgi sonlandırma kuralında çok fazla madde uygulanmıştır.

## 2.1. İlgili Arařtırmalar Özet

İlgili arařtırmalar incelendiğinde; farklı sonlandırma kurallarının kullanılmasının genellikle yetenek kestirim yöntemleri, madde havuzunun dağılım özelliđi ve büyüklüğü ile birlikte çalışıldıđı görölmektedir. Çalışmalarda gerçek veri ve/veya simülasyon verinin kullanılmıřtır. Sonlandırma kuralları genellikle sabit uzunluk ve deđişen uzunluk olarak seçilmektedir. Deđişen uzunluk sonlandırma kurallarında en çok tercih edilen standart hata sonlandırma kuralı olarak karřımıza çıkmaktadır. Ayrıca minimum-bilgi ve theta yakınsama sonlandırma kuralları da arařtırmalarda tercih edilmektedir. Deđişen uzunluk sonlandırma kuralının maksimum madde sayısı ile kombinasyonunun kullanıldıđı arařtırmalar da mevcuttur. Sonuçlar hata göstergeleri olan RMSE, yanlılık deđerine ve bazılarında uygulanan madde sayısına göre yorumlanmıřtır.

Arařtırmalarda her bir sonlandırma kuralından yola çıkarak; ölçme kesinliđi ve etkililiđi ile ilgili kullanılan deđişkenler bağlamında sonuç ve öneriler getirilmiřtir. Yine sabit uzunluk ve deđişen uzunluk sonlandırma kurallarının birbiri ile karřılařtırılması da söz konusudur. Ayrıca bir çalışmada mevcut sonlandırma kuralları ile arařtırmacılar tarafından yeni önerilen sonlandırma kuralının karřılařtırılması da mevcuttur.

## 3. YÖNTEM

Bu bölümde araştırmanın türü, verilerin üretilmesi, araştırma koşulları ve verilerin analizine yer verilmektedir.

### 3.1. Araştırmanın Türü

Bu çalışmada BBT'lerde ölçme kesinliği ve test uzunluğunun farklı testi sonlandırma kurallarına göre kontrollü koşullarda incelenmektedir. Kontrollü koşulların oluşturularak uygun verilerin türetilmesi bakımından araştırma, bir simülasyon çalışmasıdır. Araştırmada simülasyon verileri ile yöntemleri etkileyecek değişkenler kontrol edilmekte ve yöntemlerin üstünlükleri/zayıflıkları değerlendirilmektedir. Araştırma bu yönüyle de yöntemlerin geliştirilmesine katkı sağlayacağından temel araştırma olarak kabul edilebilir (Karasar, 2004).

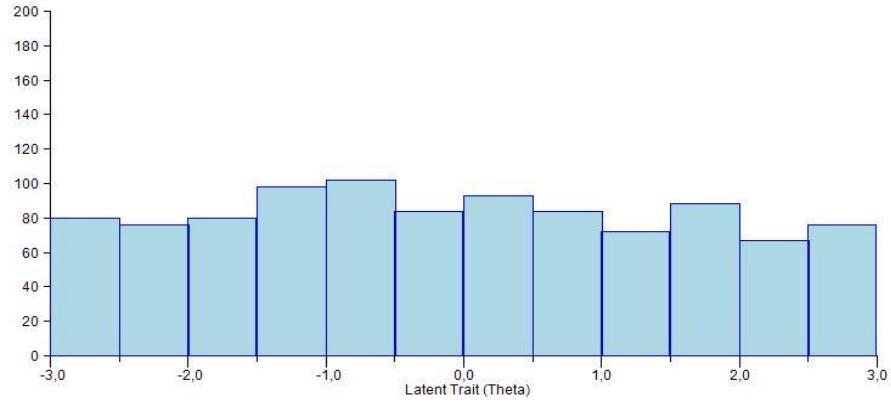
### 3.2. Verilerin Türetilmesi

Literatürde BBT uygulamaları için kullanılan çok çeşitli simülasyon programları yer almaktadır. Bu çalışmada SimulCAT simülasyon veri üretme programından (Han, 2011) yararlanılmıştır. Veri üretimi için öncelikle bireylere ait yetenek parametre, ikinci aşamada ise madde parametre değerleri elde edilmiştir. Aşağıda yetenek ve madde parametre değerlerinin elde edilmesi açıklanmıştır.

#### 3.2.1. Yetenek Parametrelerinin Elde Edilmesi

Testi alan bireylerin yetenek parametresi ( $\theta$  değeri) -3 ve +3 arasında tek biçimli dağılacak şekilde oluşturulmuştur. Cevaplayıcı sayısı 1000 olarak belirlenmiştir. Bireylerin yetenek parametrelerine göre dağılımı Şekil 3.1'de sunulmuştur.





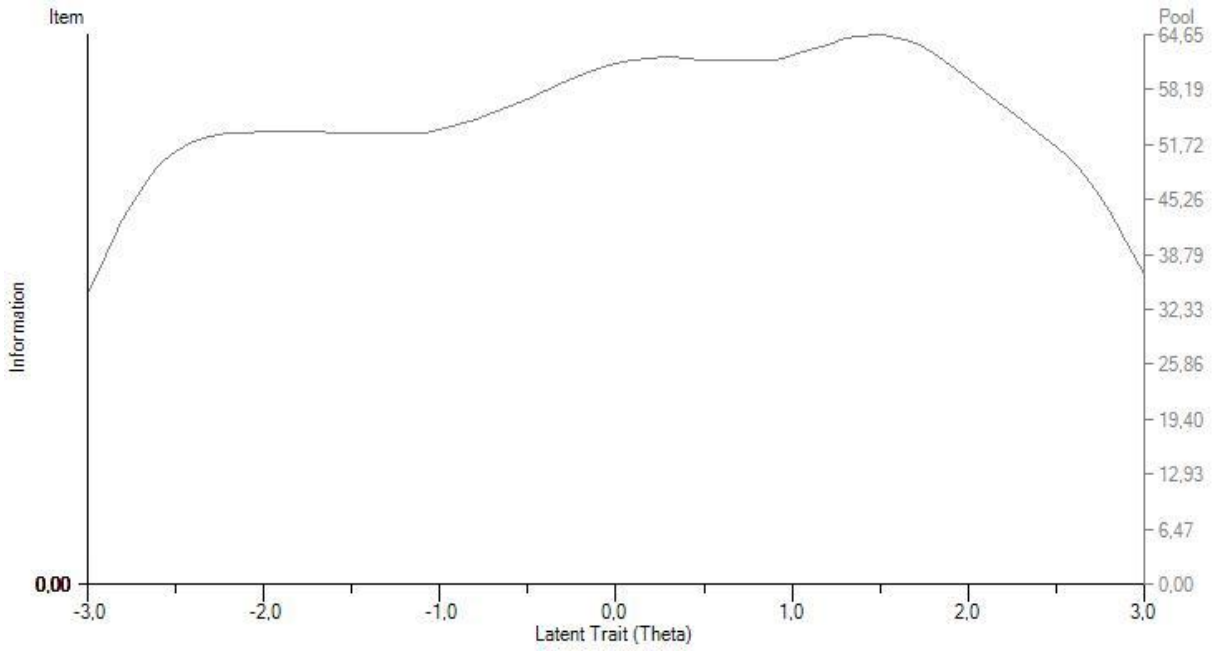
### Şekil 3. 1. Bireylerin Yetenek Parametrelerine Göre Dağılımı

Bireylerin yetenek parametrelerine göre kümülatif dağılımına ilişkin histogram eğrisi Ek-1'de sunulmuştur.

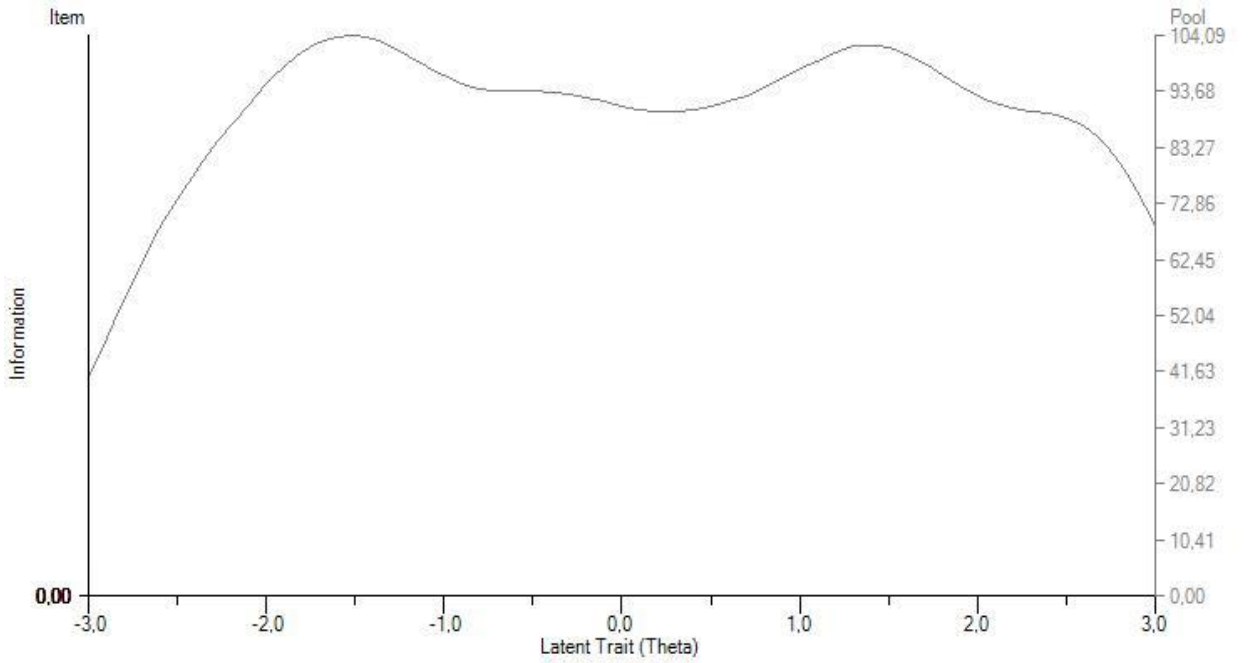
#### 3.2.2. Madde Parametrelerinin Elde Edilmesi

Araştırmada kullanılmak üzere çoktan seçmeli maddelerden oluşan 2 farklı büyüklükte madde havuzu oluşturulmuştur. Araştırma kapsamında kullanılan veri üretme programı, 3PLM ile uyumlu veri üretmektedir. Programda 250 ve 500 den oluşan 2 farklı havuzun parametreleri ayrı ayrı türetilmiştir.

Araştırmada sadece madde havuzu büyüklüğün etkisi incelenmek istendiğinden her iki büyüklükteki havuz için madde parametre aralıkları aynı belirlenmiştir. Her iki madde havuzundaki, maddelerin parametreleri tek biçimli olacak şekilde; a parametresi  $[0,50;2,00]$ ; b parametresi  $[-3,00;+3,00]$  ve c parametresi ise  $[0,05;0,20]$  aralığında belirlenmiştir. Şekil 3.2. ve Şekil 3.3.de 250 ve 500 maddelik havuza ait havuz bilgi fonksiyonları yer almaktadır.



**Şekil 3. 2. 250 Maddelik Havuza Ait Havuz Bilgi Fonksiyonu**



**Şekil 3. 3. 500 Maddelik Havuza Ait Havuz Bilgi Fonksiyonu**

Araştırma kapsamında oluşturulan; 250 ve 500 maddelik havuza ait, cevaplama fonksiyonları Ek-2 ve Ek-3'te sunulmuştur

### 3.3. BBT Koşulları

Araştırmada verilerin türetilmesinden sonra her bir alt problem kapsamında yer alan koşullar için program çalıştırılmıştır. Tüm alt problemlerde madde seçme yöntemi olarak, MFI sabit olarak ele alınmıştır. Yetenek kestirim yöntemi MLE ve EAP yöntemleri, teste başlama koşulları ise  $b=0$  ve  $-1 < b < +1$  olarak belirlenmiştir.

Testi sonlandırmada sabit uzunluk, yetenek düzeyinin standart hatası ve theta değerinin yakınsaması kuralları kullanılmıştır.

1. *Sabit uzunluk*: Bu kural için 15 ve 20 maddelik testler olmak üzere iki koşul belirlenmiştir. Blaise ve Raiche (2002) çalışmalarında madde sayısının sonlandırma kuralı olarak temel alındığı çalışmalarda en az 13 maddenin uygulanmasını önermişlerdir.

2. *Yetenek düzeyinin standart hatası  $SE(\theta)$* : Standart hata sonlandırma kuralı araştırmada iki farklı kurala göre ele alınmıştır.

a)  *$SE(\theta)$  ve  $r^2$  değerine göre sonlandırma*: Bu kural için, BBT uygulamasında tahmin edilen güvenilirlik değeri belirli bir düzeye ulaştığında sonlandırılması ilişkin koşullar belirlenmiştir. BBT uygulamasında kestirilen yetenek düzeyi ve güvenilirlik değeri sonlandırma kuralı üzerinde aynı etkiye sahiptir (Wang, Hanson ve Lau, 1999).

Ayrıca güvenilirlik katsayısı ( $r$ ) ve  $SE(\theta)$  arasındaki ilişki;

$$r^2 = 1 - SE(\theta)^2$$

şeklinde tanımlanmaktadır.

Blaise ve Raiche (2002) çalışmalarında  $[-3,00; +3,00]$  aralığında değişen yetenek düzeylerinde standart hatanın 0,40' a eşit veya daha düşük olmasının ölçme kesinliği için uygun olduğunu önermişlerdir. Ayrıca; Babcock ve Weiss (2012)  $SE(\theta)$  değerlerinin belirlenmesinde güvenilirliğin karesini ( $r^2$ ) göz önüne almışlardır.

Bu çalışmada  $SE(\theta)$  değerinin en az 0.40 olması ölçütü ve  $r^2$  ölçütü birlikte ele alınmıştır. Buna göre sırasıyla;  $r^2 = 0,85$ ,  $r^2 = 0,90$  ve  $r^2 = 0,95$

değerlerine karşılık gelen  $SE(\theta)$  değerleri 0,385, 0,315 ve 0,220 sonlandırma kuralı olarak alınmıştır.

b) *SE( $\theta$ ) ve minimum madde sayısına göre sonlandırma*: Minimum madde sayısı kısıtlaması diğer sonlandırma kuralları ile birlikte kullanılması değişen uzunlukta sonlandırma için uygulanan kullanışlı bir çözümdür (Babcock ve Weiss, 2012). Bu kuralda  $SE(\theta)$  değerinin belirlenen aralığa düşmesi ve belirlenen minimum sayıda maddenin uygulanması gerekliliği söz konusudur. Eğer  $SE(\theta)$  değeri belirlenen değerin altına düşse dahi belirlenen maddeden daha az uygulama söz konusu ise test sonlandırılmayacaktır. Bock ve Mislevy (1982) [-3,00;+3,00] aralığında değişen yetenek düzeyinde en az 10 maddenin uygulanması durumunda yanlılığın sifıra yaklaşacağını belirtmişlerdir. Bu çalışmada  $SE(\theta)$  değerinin 0,385; 0,315 ve 0,220 değerlerinin altına düşmesi ve en az madde sayısı 10 olarak belirlenmiştir.

3. *Yetenek kestirimi ( $\theta$ ) değerinde mutlak değişim yakınsaması ( $\theta$  değerinde yakınsama)*:  $\theta$  değerinde yakınsama sonlandırma kuralı araştırmada iki farklı kurala göre ele alınmıştır.

a) *Yetenek kestirimleri arasındaki fark*: Yetenek kestirimleri arasındaki fark 0.05 ve 0.02 olarak belirlenmiştir (Babcock ve Weiss, 2012).

b) *Yetenek kestirimleri arasındaki farkının belirlenen değerin altına düşmesi ve minimum madde sayısı*: Yetenek kestirimi ( $\theta$ ) değerleri arasındaki fark 0,05 ve 0,02 ve en az madde sayısı 10 seçilmiştir. Bu kuralda; ( $\theta$ ) değerleri arasındaki farkın belirlenen aralığa düşmesi ve belirlenen minimum sayıda maddenin uygulanması gerekliliği söz konusudur. Eğer ( $\theta$ ) değerleri arasındaki fark; belirlenen değerin altına düşse dahi 10 maddeden daha az uygulama söz konusu ise test sonlandırılmayacaktır.

Araştırmada ele alınacak 5 ayrı sonlandırma kuralına göre oluşturulan 12 sonlandırma koşulu Tablo 3.1.'de yer almaktadır.

**Tablo 3. 1. Araştırma Kapsamında Kullanılan Sonlandırma Koşulları**

<b>Koşullar</b>	<b>Sonlandırma Koşulu</b>
Koşul 1	Sabit uzunluk-15 madde
Koşul 2	Sabit uzunluk-20 madde
Koşul 3	SE( $\theta$ ): 0,385
Koşul 4	SE( $\theta$ ): 0,315
Koşul 5	SE( $\theta$ ): 0,220
Koşul 6	SE( $\theta$ ): 0,385 ve en az 10 madde
Koşul 7	SE( $\theta$ ): 0,315 ve en az 10 madde
Koşul 8	SE( $\theta$ ): 0,220 ve en az 10 madde
Koşul 9	( $\theta$ ) değeri farkı: 0,05
Koşul 10	( $\theta$ ) değeri farkı: 0,02
Koşul 11	( $\theta$ ) değeri farkı: 0,05 ve en az 10 madde

Araştırmada ele alınan yetenek kestirim yöntemi, madde havuzu büyüklüğü, başlama kuralı ve sonlandırma kuralı değişkenlerine ait koşulların yer aldığı; Tablo 3.2. aşağıda verilmiştir.

**Tablo 3. 2. Araştırmada Kapsamında Ele Alınan Değişkenler ve Koşulları**

<b>Değişkenler</b>	<b>Koşullar</b>
Yetenek kestirimi	ML, EAP
Madde havuzu	250, 500
Başlama kuralı	$b=0, -1 < b < +1$
Sonlandırma kuralı	
• Sabit uzunluk	15, 20
• Standart hata ve $r^2$	0.385, 0.315 ve 0.222
• Standart hata ve minimum madde sayısı	Se: 0.385, 0.315, 0.222 Madde sayısı: 10
• Yetenek kestirimlerinin yakınsaması	0.05 ve 0.02
• Yetenek kestirimlerinin yakınsaması ve minimum madde sayısı	0.05 ve 0.02 Madde sayısı: 10

Tablo 3.2.'de yer alan değişkenler ve koşulları çaprazlandığında 8 alt problem ve her bir alt problem için 12 sonlandırma koşulu oluşmaktadır. Araştırmanın genelinde 96 koşul için program çalıştırılmıştır. Harwell (1996) örneklem yanlılığının ortadan kaldırılması için en az 25 tekrarın kullanılmasını önermektedir (akt: Evans, 2010). Araştırmada her koşul için 25 tekrar kullanılmıştır.

### 3.4. Verilerin Analizi

Her bir alt problem için ölçme kesinliği ve test uzunlukları hesaplanmış ve birbirleri ile karşılaştırılmıştır.

**1. Ölçme Kesinliği:** Ölçme kesinliği, ölçmenin tutarlığı ile ilgilidir. Araştırma kapsamında; ölçme kesinliğinin belirlenmesi için, hata göstergeleri olan RMSE, yanlılık, uyum değerleri incelenmiştir. Hata göstergelerinin değeri ile ölçme kesinliğinin ters orantılı olarak değişmesinden yola çıkılarak sonuçlar yorumlanmıştır.

- i. **RMSE (Root Mean Squared Error):** BBT uygulaması ile elde edilen bireye ait tahmin edilen yetenek düzeyi ile gerçek yetenek düzeyi arasındaki mutlak farklılığa ilişkin istatistiktir. RMSE değeri aşağıdaki eşitlik yardımıyla hesaplanır.

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

Burada;

n: toplam birey sayısı,

$\theta_i$  i.bireyin gerçek yetenek düzeyi değeri;

$\hat{\theta}_i$  .i.bireyin kestirilen yetenek düzeyi değeri;

göstermektedir..

- ii. **Yanlılık (Bias):** BBT uygulaması ile elde edilen bireye ait tahmin edilen yetenek düzeyi ile gerçek yetenek düzeyi arasındaki ortalama anlamlı farklılık istatistiğidir. Yanlılık değeri

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

ile hesaplanır. RMSE değeri ile Bias değeri arasındaki ilişkiyi;

$$RMSE^2 = Bias^2 + SE^2$$

göstermek mümkündür. Burada RMSE değerinin, yanlışlık ve tahminin standart hatasını içerdiği görülmektedir.

- iii. **Uyum (Fidelity):** Uyum katsayısı bireylerin gerçek yetenek düzeyi ile tahmin edilen yetenek düzeyi arasındaki Pearson korelasyon olarak tanımlanır (Weiss,1982). Bireye ait tahmin edilen yetenek düzeyi ile gerçek yetenek düzeyi arasındaki korelasyon katsayısı;

$$r = \frac{cov(\hat{\theta}, \theta)}{SS(\hat{\theta})SS(\theta)}$$

ile hesaplanır. Bu değer yüksek elde edilmesi; tahmin edilen yetenek düzeyi ( $\hat{\theta}_i$ ) ile gerçek yetenek düzeyi ( $\theta_i$ ) arasındaki yüksek uyumu belirtmektedir.

**2. Test Uzunluğu:** Çalışma kapsamında, her bir koşul için bireylerin aldıkları maddelerin değeridir. Bu değişken sabit uzunluk sonlandırma kuralı için hesaplanmamıştır.

## 4. BULGULAR VE YORUM

Bu bölümde araştırmamızın elde edilen bulgularına ve bu bulgularda yapılan yorumlara yer verilmiştir.

### 4.1. Bulgular

#### 4.1.1. Birinci Alt Probleme İlişkin Bulgular

Bu alt problem için, BBT uygulamasında madde havuzu büyüklüğü 250; başlama kuralı  $b=0$  ve yetenek kestirim yöntemi MLE seçildiğinde; ölçme kesinliği ve test uzunluğu incelenmiştir. Tablo 3.1.'de yer alan koşullara göre; elde edilen yetenek kestirimlerine ait RMSE, yanlılık, uyum katsayısı ve test uzunluğuna ait ortalama değerler Tablo 4.1.'de yer almaktadır.

**Tablo 4. 1. Madde Havuzu Büyüklüğü 250, Başlama kuralı  $b=0$  ve Yetenek Kestirimi MLE Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri\***

Sonlandırma Koşulları		RMSE	Yanlılık	Uyum Katsayısı	Test Uzunluğu	
K1	Sabit Uzunluk	15	0,302	0,026	0,984	15,0
K2		20	0,237	0,015	0,990	20,0
K3	SE( $\theta$ )	0.385	0,546	0,097	0,949	7,0
K4		0.315	0,386	0,044	0,974	9,6
K5		0.220	0,222	0,010	0,991	19,4
K6	SE( $\theta$ ) ve en az 10 madde	0.385	0,397	0,052	0,972	10,2
K7		0.315	0,356	0,040	0,978	10,6
K8		0.220	0,222	0,010	0,991	19,4
K9	(θ) değerinde yakınsama	0.05	0,305	0,009	0,984	15,9
K10		0.02	0,266	0,006	0,988	28,2
K11	(θ) değerinde yakınsama ve en az 10 madde	0.05	0,246	0,012	0,989	16,6
K12		0.02	0,190	0,002	0,994	29,4

\* Değerlerin tümü 25 tekrardan elde edilen verilerin ortalaması hesaplanarak verilmiştir.



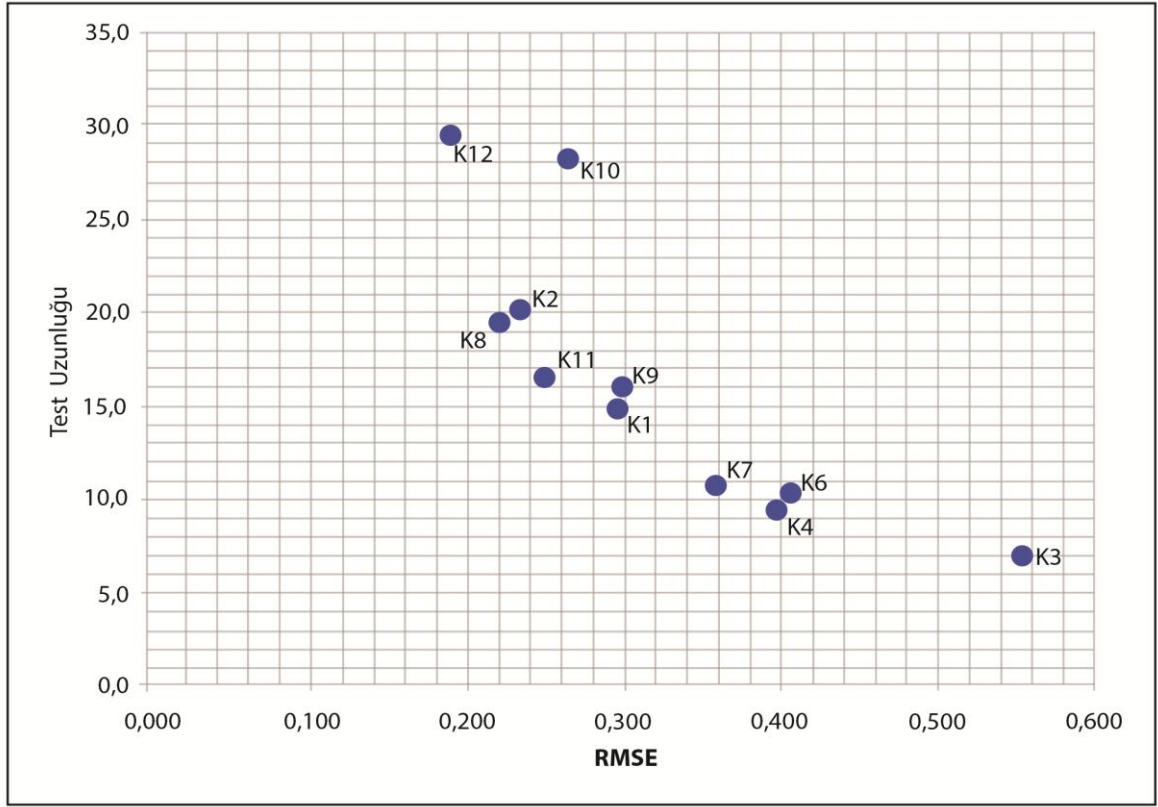
Tablo 4.1. incelendiğinde; ele alınan koşullar için en düşük RMSE değerinin **0,02 theta yakınsama-en az 10 madde** sonlandırma koşulunda elde edildiği görülmektedir. Bu sonlandırma koşulunu sırasıyla **0,220 SE(theta)**, **0,220 SE(theta)** ve **en az 10 madde** ile **sabit uzunluk 20 madde** sonlandırma koşulları izlemektedir. En yüksek RMSE değerinin ise sırasıyla **0,385 SE(theta)**, **0,385 SE(theta)**-en az 10 madde ve **0,315 SE(theta)** sonlandırma koşullarının kullanıldığı durumda elde edildiği görülmektedir.

Yanlılık değerleri incelendiğinde RMSE değerleri ile çoğunlukla paralellik gösterdiği gözlenmiştir. En düşük yanlılık değeri **0,02 theta yakınsama-en az 10 madde** koşulunda elde edilmiştir. Bu koşulu sırasıyla **0,02** ve **0,05 theta yakınsama** izlemektedir. En yüksek yanlılık değerleri ise sırasıyla **0,385 SE(theta)**, **0,385 SE(theta)**-en az 10 madde ve **0,315 SE(theta)** sonlandırma koşullarında elde edilmiştir.

Uyum katsayıları incelendiğinde, paralel olarak benzer sonuçlar çıkmıştır. Hesaplanan yetenek düzeyi ile gerçek yetenek düzeyi arasındaki uyum en yüksek **0,02 theta yakınsama-en az 10 madde**, en düşük uyum ise **0,385 SE(theta)** sonlandırma koşulunun kullanıldığı durumda elde edilmiştir.

Her bir koşul için elde edilen ortalama test uzunluklarının incelendiğinde testin en az madde ile **0,385 SE(theta)** (ortalama 7 madde), en çok madde ile **0,02 theta yakınsama-en az 10 madde** sonlandırma koşullarında (ortalama 29 madde) sonlandırıldığı gözlemlenmiştir.

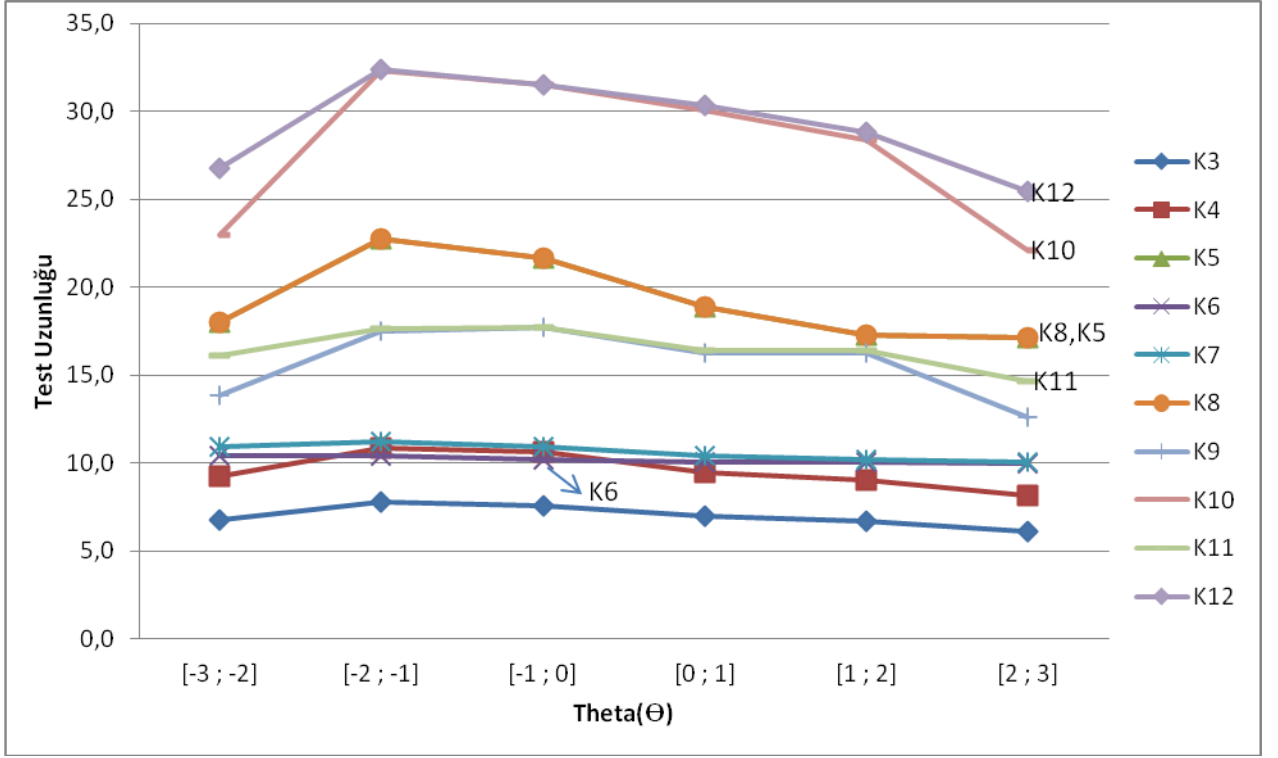
Aşağıda yer alan Şekil 4.1.'de, bu alt problem için, Tablo 3.1.'de yer alan her bir koşula ait ortalama test uzunluğu ve RMSE değerlerinin ilişkisi yer almaktadır.



**Şekil 4. 1. Birinci Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki**

Şekil 4.1. incelendiğinde; test uzunluğunun arttıkça RMSE değerlerinin azaldığı görülmektedir. Burada **0,02 ve 0,05 theta yakınsama** koşulunda testin ortalama uzunluğu, **theta yakınsama-en az 10 madde** göz önüne alındığı koşula kıyasla daha kısa olup, daha yüksek RMSE değeri elde edilmiştir. **0,385 ve 0,315 SE(theta)** sonlandırma koşullarında ise testin ortalama uzunluğunun; **standart hata-en az 10 madde** nin göz önüne alındığı sonlandırma kuralına kıyasla daha kısa olduğu ve daha yüksek RMSE değeri elde edildiği ayrıca belirlenmiştir.

Test uzunluklarının, bireylerin yetenek düzeyine göre, sürekli dağılımı Ek-4'de yer almaktadır. Ayrıca yetenek düzeyi (-3 ve +3 theta arasında değişen) 6 eşit birime bölünmüş ve 8 koşula (K3-K12) ait ortalama test uzunlukları, Şekil 4.2.'de verilmiştir.



**Şekil 4. 2. Birinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu**

Şekil 4.2. incelendiğinde değişen uzunluk sonlandırma koşullarında; test uzunluğu genellikle [-2;-1] yetenek düzeyinde en yüksek; [2;3] yetenek düzeyinde de en düşük değerine ulaştığı görülmektedir. Ortalama test uzunluğunun 10 maddeden düşük olduğu **0,385 ve 0,315 SE(theta)** sonlandırma koşullarının hem tek başına hem de en az 10 madde koşulu ile birlikte kullanımında yetenek düzeylerinde daha az değişim söz konusudur. **0,220 SE(theta)** koşulunda ise [1;3] theta yetenek düzeyinde ortalama test uzunluğunun değişmediği gözlenmiştir.

#### 4.1.2. İkinci Alt Probleme İlişkin Bulgular

Bu alt problem için, BBT uygulamasında madde havuzu büyüklüğü 250; başlama kuralı  $-1 < b < +1$  ve yetenek kestirim yöntemi MLE seçildiğinde; ölçme kesinliği ve test uzunluğu incelenmiştir.

Tablo 3.1.'de yer alan koşullara göre; elde edilen yetenek kestirimlerine ait RMSE, yanlılık, uyum katsayısı ve test uzunluğuna ait ortalama değerler Tablo 4.2.'de yer almaktadır.

**Tablo 4. 2. Madde Havuzu Büyüklüğü 250, Başlama kuralı  $-1 < b < 1$  ve Yetenek Kestirimi MLE Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri\***

Sonlandırma Koşulları			RMSE	Yanlılık	Uyum Katsayısı	Test Uzunluğu
K1	Sabit Uzunluk	15	0,297	0,026	0,984	15,0
K2		20	0,232	0,014	0,990	20,0
K3	SE( $\theta$ )	0.385	0,554	0,100	0,947	7,1
K4		0.315	0,403	0,050	0,972	9,7
K5		0.220	0,223	0,011	0,991	19,5
K6	SE( $\theta$ ) ve en az 10 madde	0.385	0,406	0,054	0,971	10,2
K7		0.315	0,359	0,038	0,977	10,7
K8		0.220	0,223	0,011	0,991	19,5
K9	(θ) değerinde yakınsama	0.05	0,298	0,010	0,985	15,9
K10		0.02	0,265	0,005	0,988	28,2
K11	(θ) değerinde yakınsama ve en az 10 madde	0.05	0,250	0,013	0,989	15,0
K12		0.02	0,189	0,005	0,994	29,4

\* Değerlerin tümü 25 tekrardan elde edilen verilerin ortalaması hesaplanarak verilmiştir.

Tablo 4.2. incelendiğinde; en düşük RMSE değerinin **0,02 theta yakınsama-en az 10 madde** sonlandırma koşulunda elde edildiği görülmektedir. Bu değeri sırasıyla **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) ve **sabit uzunluk 20 madde** izlemektedir. En yüksek RMSE değeri ise **0,385 SE(theta)** sonlandırma

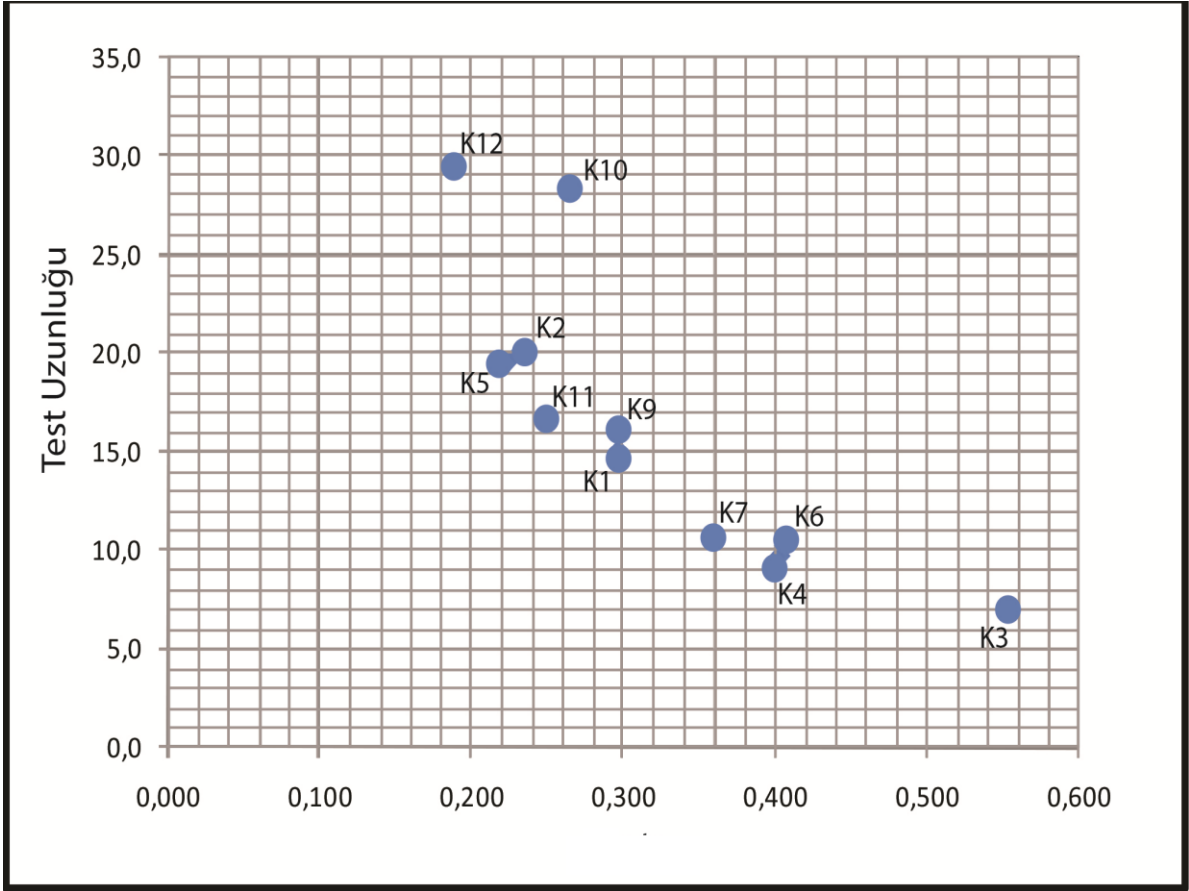
koşulunun kullanıldığı durumda elde edilmiştir. Bu değeri sırasıyla **0,385 SE(theta)- en az 10 madde** ve **0,315 SE(theta)** koşulları izlemektedir.

Yanlılık değerleri genellikle RMSE değeri ile paralellik göstermektedir. En düşük yanlılık değeri sırasıyla **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) ve **0,05 theta yakınsama** koşullarında elde edilmiştir. En yüksek yanlılık değeri ise **0,385 SE(theta)** sonlandırma koşulunda elde edilmiştir. Bu değeri sırasıyla **0,385 SE(theta)-en az 10 madde** ve **0,315 SE(theta)** izlemektedir.

Uyum katsayıları bu alt problemde tüm koşullar için birbirine yakın ve yüksek elde edilmiştir. Hesaplanan yetenek düzeyi ile gerçek yetenek düzeyi arasındaki uyum en yüksek sırasıyla **0,02 theta yakınsama-en az 10 madde** ve **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı); en düşük ise **0,385 SE(theta)** ve **0,315 SE(theta)** sonlandırma koşulunun kullanıldığı durumda elde edilmiştir.

Her bir koşul için elde edilen ortalama test uzunluklarının incelendiğinde, en düşük değer **0,385 SE(theta)** (ortalama 7 madde) elde edildiği görülmektedir. Bu değeri sırasıyla **0,315 SE(theta)** ve **0,385 SE(theta) - en az 10 madde** sonlandırma koşulları izlemektedir. Ortalama test uzunluğunun en yüksek olduğu sonlandırma koşulları sırasıyla **0,02 theta yakınsama-en az 10 madde** (ortalama 29 madde); **0,02 theta yakınsama** ve **sabit uzunluk 20 madde** dir.

Aşağıda yer alan Şekil 4.3.'te bu alt problem için her bir sonlandırma koşuluna ait ortalama test uzunluğu ve RMSE değerlerinin ilişkisi yer almaktadır.

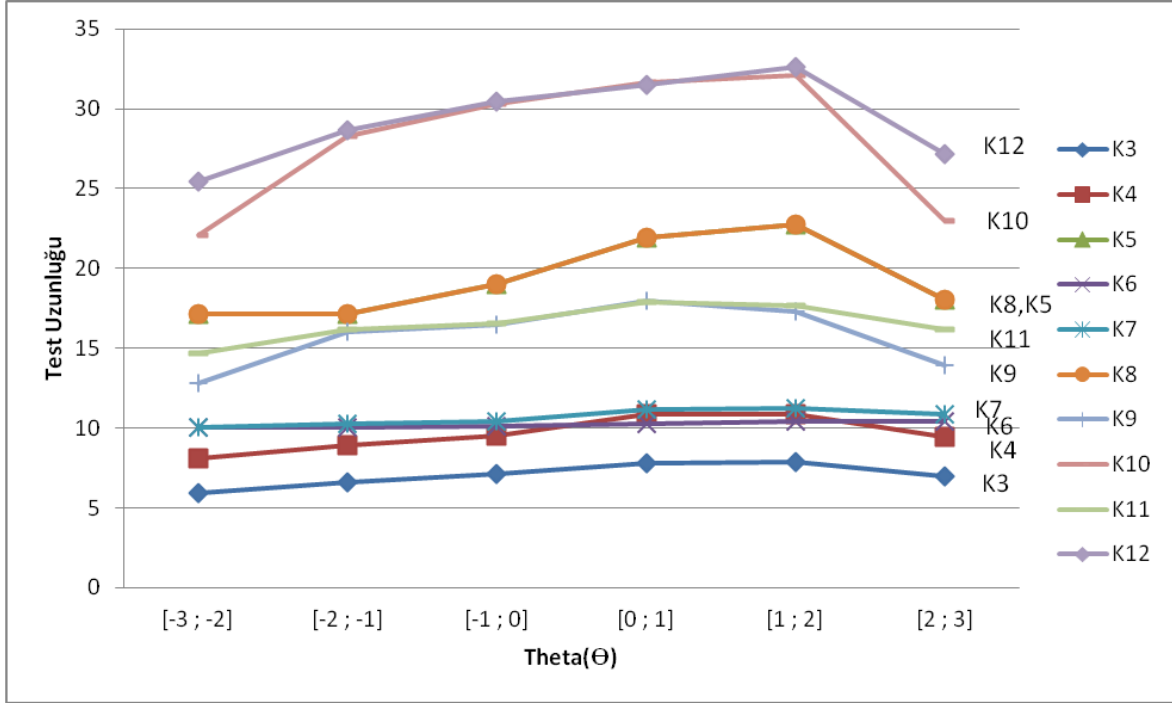


**Şekil 4. 3. İkinci Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki**

Şekil 4.3. incelendiğinde **0,05 ve 0,02 theta yakınsama** ile **0,385 ve 0,315 SE(theta)** sonlandırma koşullarında en az 10 madde sınırlamasının kullanılması ile ortalama test uzunluğu artış gösterse de RMSE değerindeki düşüşler dikkat çekicidir. **0,385 SE(theta)** sonlandırma koşulunu en az 10 madde şartı ile birlikte kullandığımızda ortalama 3 madde fazla uygulayarak RMSE değerinde 0,149'luk bir düşüşün olduğu da ayrıca göze çarpmaktadır. **Sabit uzunluk 15 madde ve 0,05 theta yakınsama** koşullarında ortalama RMSE ve test uzunlukları arasında fark çok düşük olmasına karşın yanlılık değerine bakıldığında 0,05 theta yakınsamanın kullanıldığı koşulun lehine olduğu görülmektedir.

Test uzunluklarının, bireylerin yetenek düzeyine göre, sürekli dağılımı Ek-5'de yer almaktadır. Ayrıca yetenek düzeyi (-3 ve +3 theta arasında değişen) 6 eşit birime

bölünmüş ve 8 koşula (K3-K12) ait ortalama test uzunlukları, Şekil 4.4.'te verilmiştir.



**Şekil 4. 4. İkinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu**

Şekil 4.4. incelendiğinde değişen uzunluk sonlandırma koşullarında; test uzunluğu genellikle [1;2] theta yetenek düzeyinde en yüksek değerine ulaştığı; [2;3] theta yetenek düzeyinde de hızlı bir düşüş gösterdiği görülmektedir. Ortalama test uzunluğunun 10 maddeden düşük olduğu **0,385 ve 0,315 SE(theta)** sonlandırma koşullarının en az 10 madde koşulu ile birlikte kullanımında ise yetenek düzeylerinde daha az değişim söz konusudur. **0,220 SE(theta)** koşulunda ise [-3;-1] theta yetenek düzeyinde test uzunluğunun değişmediği gözlenmiştir.

### 4.1.3. Üçüncü Alt Probleme İlişkin Bulgular

Bu alt problem için, BBT uygulamasında madde havuzu büyüklüğü 500; başlama kuralı  $b=0$  ve yetenek kestirim yöntemi MLE seçildiğinde; ölçme kesinliği ve test uzunluğu incelenmiştir.

Tablo 3.1.'de yer alan koşullara göre; elde edilen yetenek kestirimlerine ait RMSE, yanlılık, uyum katsayısı ve test uzunluğuna ait ortalama değerler Tablo 4.3.'te yer almaktadır.

**Tablo 4. 3. Madde Havuzu Büyüklüğü 500, Başlama kuralı  $b=0$  ve Yetenek Kestirimi MLE Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri\***

Sonlandırma Koşulları		RMSE	Yanlılık	Uyum Katsayısı	Test Uzunluğu	
K1	Sabit Uzunluk	15	0,273	0,021	0,987	15,0
K2		20	0,217	0,013	0,992	20,0
K3	SE( $\theta$ )	0.385	0,516	0,079	0,954	6,8
K4		0.315	0,382	0,042	0,974	8,9
K5		0.220	0,227	0,012	0,991	16,3
K6	SE( $\theta$ ) ve en az 10 madde	0.385	0,359	0,041	0,977	10,1
K7		0.315	0,341	0,032	0,979	10,4
K8		0.220	0,227	0,012	0,991	16,3
K9	(θ) değerinde yakınsama	0.05	0,293	0,010	0,985	15,8
K10		0.02	0,262	0,007	0,988	28,6
K11	(θ) değerinde yakınsama ve en az 10 madde	0.05	0,228	0,009	0,991	16,6
K12		0.02	0,170	0,005	0,995	29,9

\* Değerlerin tümü 25 tekrardan elde edilen verilerin ortalaması hesaplanarak verilmiştir.

Tablo 4.3. incelendiğinde; en düşük RMSE değerinin **0,02 theta yakınsama- en az 10 madde** sonlandırma koşulunda elde edildiği görülmektedir. Bu değeri sırasıyla **sabit uzunluk 20 madde** ile **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) koşulları izlemektedir. En yüksek RMSE değeri ise **0,385**



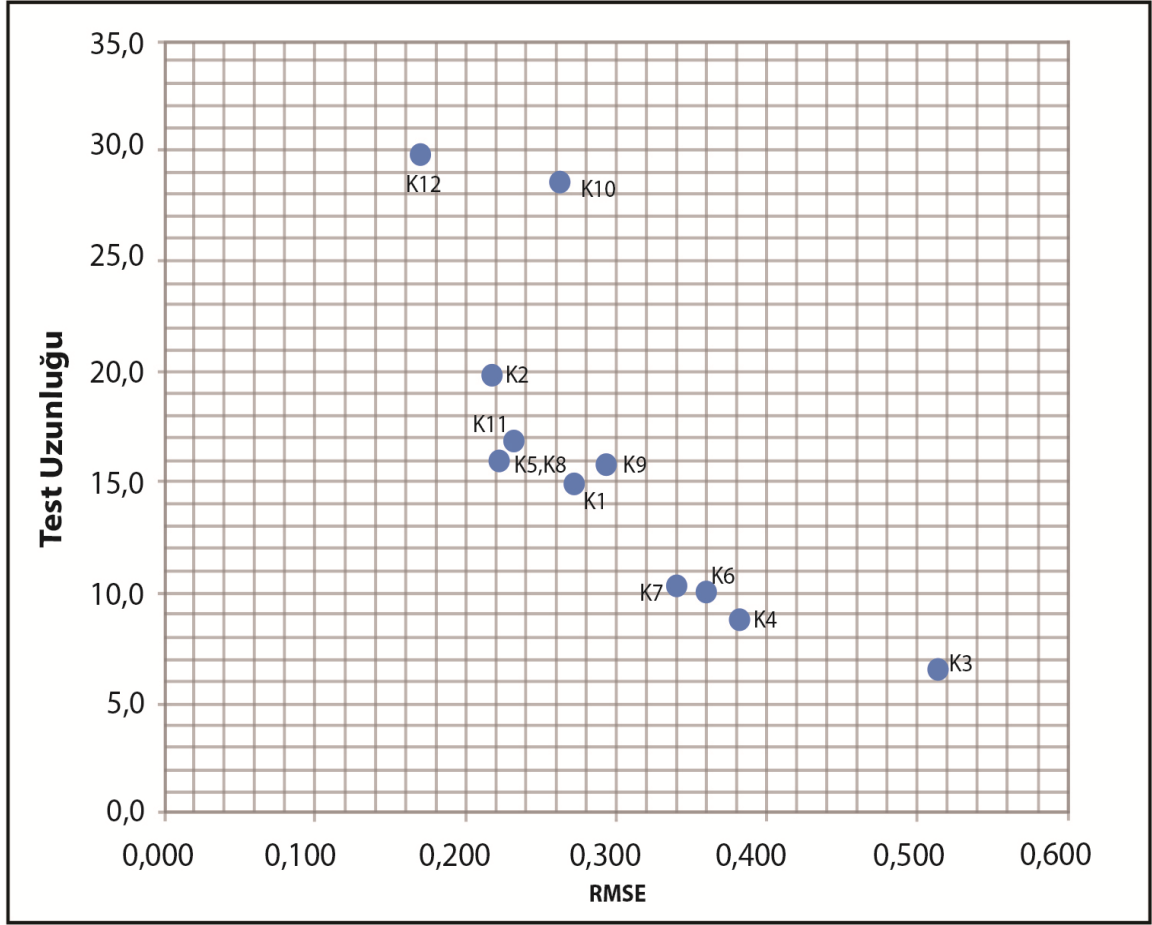
**SE(theta)** sonlandırma koşulunun kullanıldığı durumda elde edilmiştir. Bu koşulu sırasıyla **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** koşulları izlemektedir.

Yanlılık değerlerine bakıldığında; en düşük yanlılık değerleri sırasıyla **0,02 theta yakınsama- en az 10 madde, 0,02 ve 0,05 theta yakınsama-en az 10 madde** koşullarında elde edilmiştir. En yüksek yanlılık değeri ise **0,385 SE(theta); 0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşullarının kullanıldığı durumda gözlenmiştir.

Uyum katsayıları incelendiğinde katsayılar birbirine oldukça yakın ve yüksek olduğu göze çarpmaktadır. En yüksek uyum katsayısı sırasıyla **0,02 theta yakınsama- en az 10 madde; sabit uzunluk 20 madde ve 0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) ile **0,05 theta yakınsama- en az 10 madde** koşullarından elde edilmiştir. En düşük uyum katsayısının ise **0,385 SE(theta); 0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşullarında elde edildiği görülmektedir.

Her bir koşul için elde edilen ortalama test uzunluklarının incelendiğinde; testin ortalama en az madde ile **0,385 SE(theta)** koşulunda (ortalama 7 madde) sonlandırıldığı görülmektedir. Bu kuralı sırasıyla **0,315 SE(theta)** ve **0,05 theta yakınsama** izlemektedir. Testin ortalama en çok madde **0,02 theta yakınsama-en az 10 madde, 0,02 theta yakınsama ve sabit uzunluk 20 madde** koşullarında sonlandırıldığı görülmektedir.

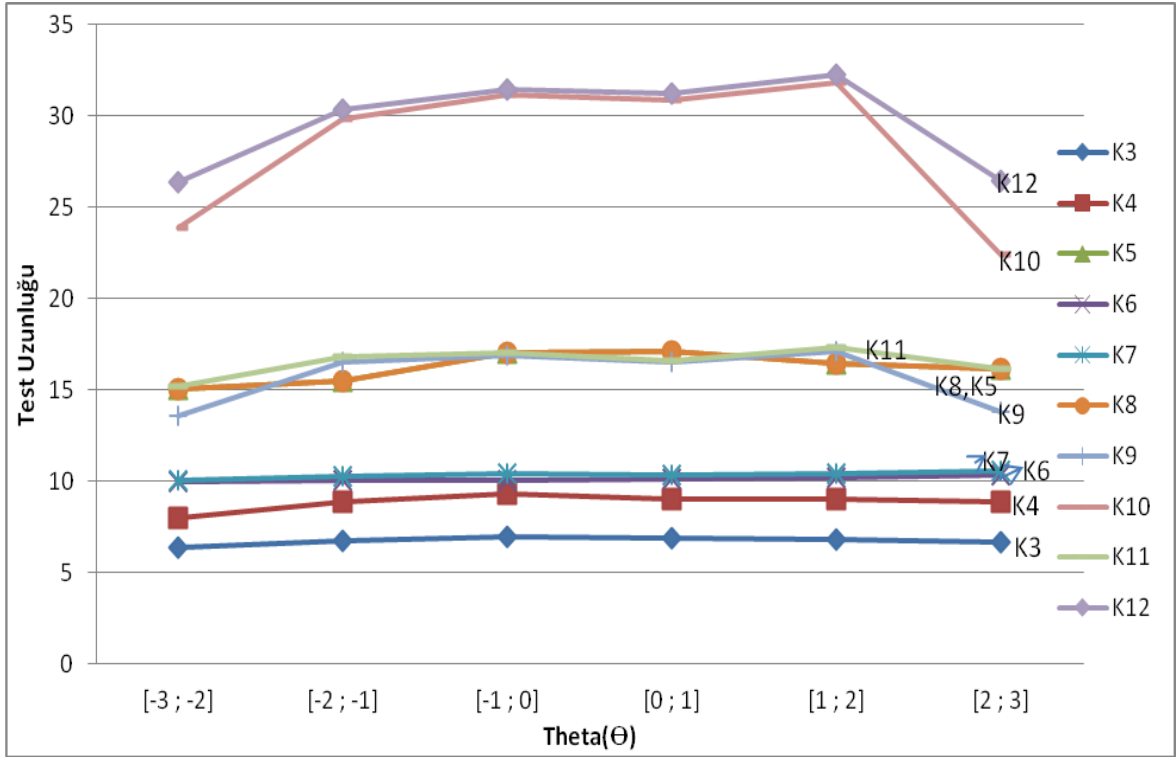
Aşağıda yer alan Şekil 4.5.'te bu alt problem için her bir sonlandırma koşuluna ait ortalama test uzunluğu ve RMSE değerlerinin ilişkisi yer almaktadır.



**Şekil 4. 5. Üçüncü Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki**

Şekil 4.5.'te **sabit uzunluk 15 madde** ve **0,05 theta yakınsama** sonlandırma koşullarında karşılaştırıldığında; **sabit uzunluk 15 madde** koşulunda daha düşük ortalama test uzunluğu ile daha düşük RMSE değeri elde edilmiştir. **0,02 theta yakınsama** ile **0,220 SE(theta)** koşulları incelendiğinde, **0,220 SE(theta)** koşulunda daha düşük madde sayısı ile daha düşük RMSE değeri elde edildiği belirlenmiştir. **0,02 ve 0,05 theta yakınsama** koşullarının minimum madde sayısı ile birlikte kullanımında RMSE ve yanlılık değerinde düşüş gözlenmiştir. Benzer durum **0,385 ve 0,315 SE(theta)** kuralları için de geçerlidir.

Test uzunluklarının, bireylerin yetenek düzeyine göre, sürekli dağılımı Ek-6'da yer almaktadır Ayrıca yetenek düzeyi (-3 ve +3 theta arasında değişen) 6 eşit birime bölünmüş ve 8 koşula (K3-K10) ait ortalama test uzunlukları, Şekil 4.6.'da verilmiştir.



**Şekil 4. 6. Üçüncü Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu**

Şekil 4.6. incelendiğinde theta yakınsama sonlandırma koşullarında; test uzunluğu genellikle [1;2] theta yetenek düzeyinde en yüksek değerine ulaştığı; [2;3] theta yetenek düzeyinde de hızlı bir düşüş gösterdiği görülmektedir. Sonlandırma koşullarından **0,385** ve **0,315 SE(theta)** ile **0,05 theta yakınsamanın** tek başına ve en az 10 madde koşulu ile birlikte kullanımında ise yetenek düzeylerinde daha az değişim söz konusudur. **0,220 SE(theta)** koşulunda ise [-3;-1] theta yetenek düzeyinde test uzunluğunun değişmediği gözlenmiştir.

#### 4.1.4. Dördüncü Alt Probleme İlişkin Bulgular

Bu alt problem için, BBT uygulamasında madde havuzu büyüklüğü 500; başlama kuralı  $1 < b < +1$  ve yetenek kestirim yöntemi MLE seçildiğinde; ölçme kesinliği ve test uzunluğu incelenmiştir.

Tablo 3.1.'de yer alan koşullara göre; elde edilen yetenek kestirimlerine ait RMSE, yanlılık, uyum katsayısı ve test uzunluğuna ait ortalama değerler Tablo 4.4.'te yer almaktadır.

**Tablo 4. 4. Madde Havuzu Büyüklüğü 500, Başlama kuralı  $-1 < b < 1$  ve Yetenek Kestirimi MLE Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri\***

Sonlandırma Koşulları			RMSE	Yanlılık	Uyum Katsayısı	Test Uzunluğu
K1	Sabit Uzunluk	15	0,272	0,021	0,987	15,0
K2		20	0,217	0,013	0,992	20,0
K3	SE( $\theta$ )	0.385	0,567	0,098	0,945	6,9
K4		0.315	0,403	0,050	0,971	9,1
K5		0.220	0,229	0,013	0,991	16,5
K6	SE( $\theta$ ) ve en az 10 madde	0.385	0,391	0,048	0,973	10,2
K7		0.315	0,362	0,041	0,977	10,5
K8		0.220	0,229	0,013	0,991	16,5
K9	(θ) değerinde yakınsama	0.05	0,301	0,014	0,984	16,0
K10		0.02	0,268	0,009	0,988	28,8
K11	(θ) değerinde yakınsama ve en az 10 madde	0.05	0,231	0,009	0,991	16,7
K12		0.02	0,169	0,002	0,995	30,0

\* Değerlerin tümü 25 tekrardan elde edilen verilerin ortalaması hesaplanarak verilmiştir.

Tablo 4.4. incelendiğinde; en düşük RMSE değerinin **0,02 theta yakınsama-en az 10 madde** sonlandırma koşulunda elde edildiği görülmektedir. Bu koşulu sırasıyla **sabit uzunluk 20 madde** ve **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) izlemektedir. RMSE değerinin en yüksek elde edildiği sonlandırma

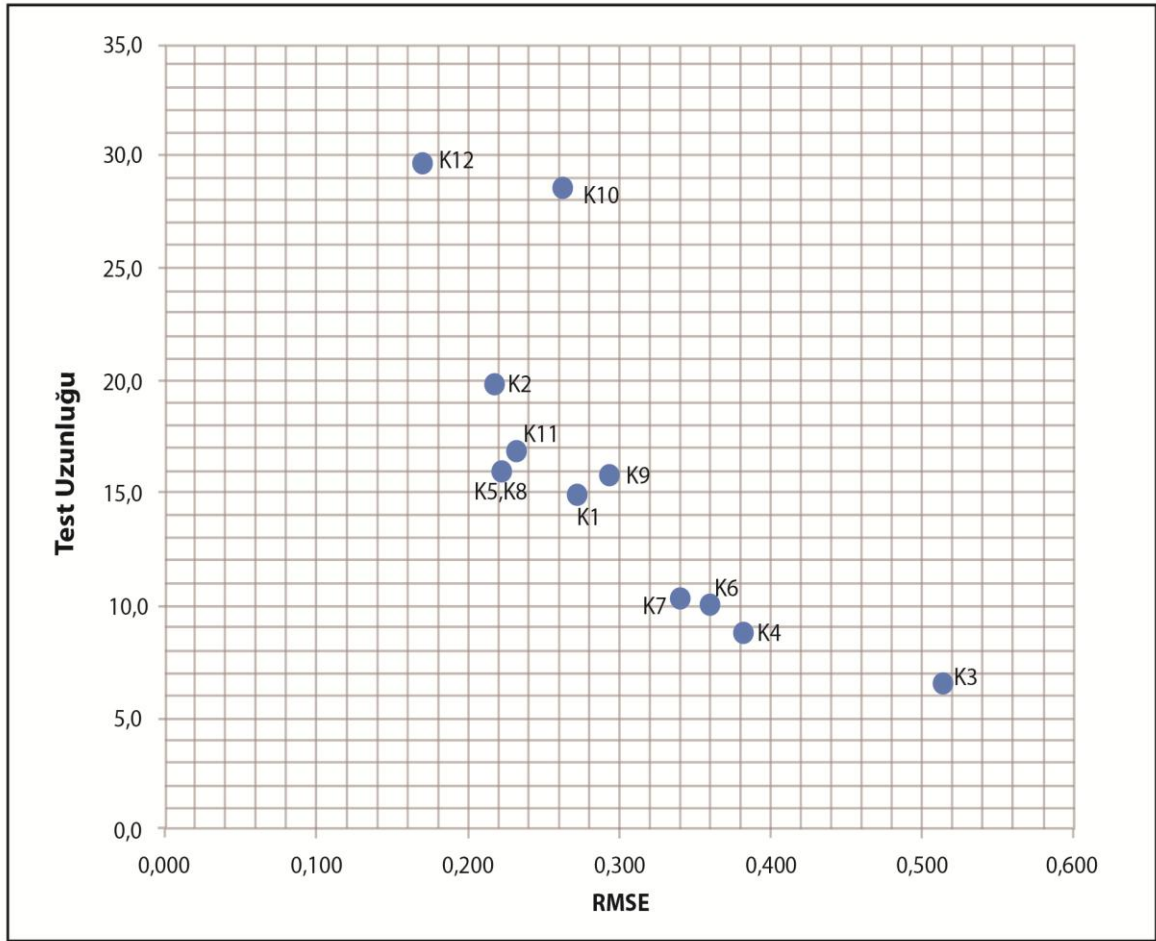
koşulları sırasıyla **0,385 SE(theta); 0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 maddedir.**

Tablo 4.4.'de yanlılık değerleri RMSE değerleri ile büyük ölçüde paralellik göstermektedir. En düşük yanlılık değerleri sırasıyla **0,02 theta yakınsama-en az 10 madde; 0,05 theta yakınsama-en az 10 madde** ile **0,02 theta yakınsama** sonlandırma koşullarında elde edilmiştir. En yüksek yanlılık değeri ise **0,385 SE(theta)** sonlandırma koşulunda elde edilmiş olup bunu sırasıyla **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşulları izlemektedir.

Bu alt problem için uyum katsayıları tüm koşullarda birbirine yakın ve yüksek elde edilmiştir. Hesaplanan ile gerçek theta arasındaki uyum sırasıyla en yüksek **0,02 theta yakınsama-en az 10 madde; sabit uzunluk 20 madde ve 0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) ile **0,05 theta yakınsama- en az 10 madde** koşullarında elde edilmiştir. En düşük uyum katsayısının ise sırasıyla **0,385 SE(theta); 0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşullarının kullanıldığı durumda elde edildiği belirlenmiştir.

Ortalama test uzunlukları incelendiğinde; testin ortalama en az madde ile **0,385 SE(theta)** (ortalama 7 madde) sonlandırma koşulunda sonlandırıldığı belirlenmiştir. Bu koşulu sırasıyla **0,315 SE(theta)** ve **0,385 SE(theta)- en az 10 madde** sonlandırma koşulları izlemektedir. Testin ortalama en çok madde ile sırasıyla **0,02 theta yakınsama-en az 10 madde** (ortalama 30 madde), **0,02 theta yakınsama** ve **sabit uzunluk 20 madde** koşullarında sonlandırıldığı gözlemlenmiştir.

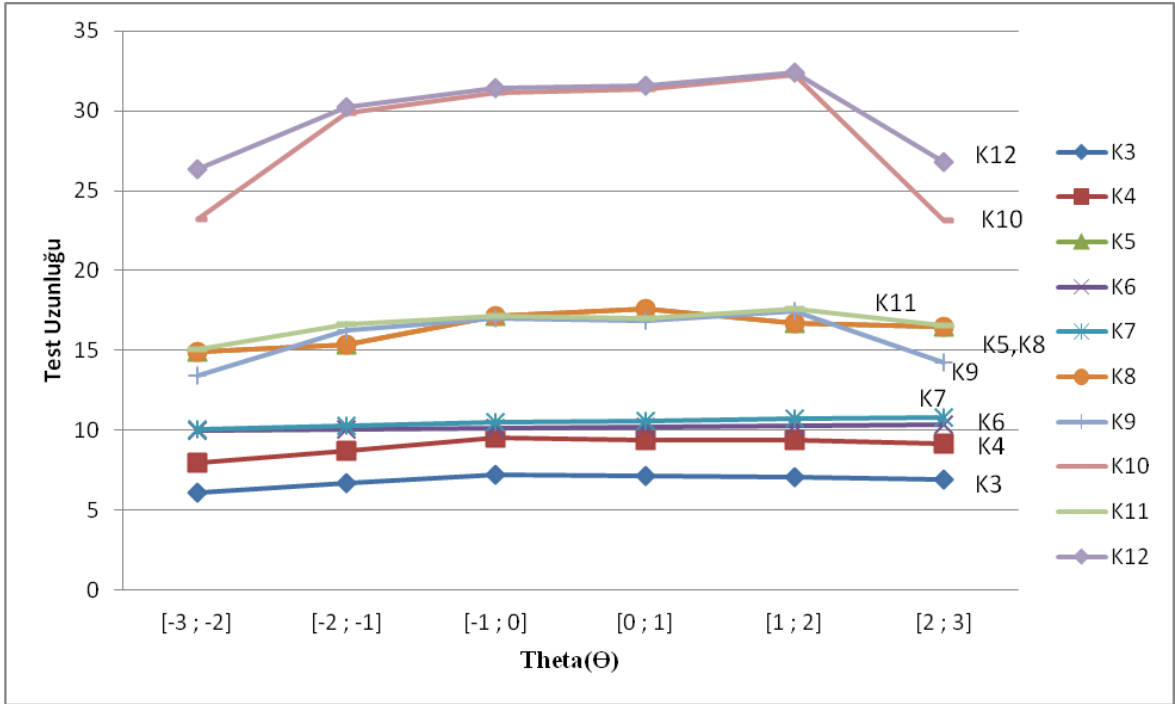
Aşağıda yer alan Şekil 4.7.'de bu alt problem için her bir sonlandırma koşuluna ait ortalama test uzunluğu ve RMSE değerlerinin ilişkisi yer almaktadır.



**Şekil 4. 7. Dördüncü Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki**

Şekil 4.7. incelendiğinde; dördüncü alt problem için **sabit uzunluk 15 madde** sonlandırma koşulunda daha düşük ortalama test uzunluğu ile **0,05 theta yakınsama** koşuluna kıyasla daha düşük RMSE değeri elde edilmiştir. Ayrıca en az 10 madde koşulunun, theta değerinde yakınsama ve standart hata sonlandırma koşulları ile birlikte kullanımının testin ortalama uzunluğunda pek bir farklılığa neden olmamakla beraber RMSE değerlerinde iyileşmeye neden olması söz konusudur.

Test uzunluklarının, bireylerin yetenek düzeyine göre, sürekli dağılımı Ek-7'de yer almaktadır Ayrıca yetenek düzeyi (-3 ve +3 theta arasında değişen) 6 eşit birime bölünmüş ve 8 koşula (K3-K10) ait ortalama test uzunlukları, Şekil 4.8.'de verilmiştir.



**Şekil 4. 8. Dördüncü Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu**

Şekil 4.8. incelendiğinde theta yakınsama sonlandırma koşullarında; test uzunluğu genellikle [1;2] theta yetenek düzeyinde en yüksek değerine ulaştığı; [2;3] theta yetenek düzeyinde de hızlı bir düşüş gösterdiği görülmektedir. **0,220 SE(theta)** koşulunda ise [-3;-1] theta yetenek düzeyinde test uzunluğunun çok önemli değişmediği ve [0;1] theta yetenek düzeyinde en yüksek değerine ulaştığı gözlenmiştir. Sonlandırma koşullarından **0,385** ve **0,315 SE(theta)** tek başına ve en az 10 madde koşulu ile birlikte kullanımında ise yetenek düzeylerinde daha az değişim söz konusudur. Ayrıca **0,385** ve **0,315 SE(theta)** koşullarının en yüksek test uzunluğu değerinin [-1;0] yetenek düzeyinde aldığı belirlenmiştir.

#### 4.1.5. Beşinci Alt Probleme İlişkin Bulgular

Bu alt problem için, BBT uygulamasında madde havuzu büyüklüğü 250; başlama kuralı  $b=0$  ve yetenek kestirim yöntemi EAP seçildiğinde; ölçme kesinliği ve test uzunluğu incelenmiştir.

Tablo 3.1.'de yer alan koşullara göre; elde edilen yetenek kestirimlerine ait RMSE, yanlılık, uyum katsayısı ve test uzunluğuna ait ortalama değerler Tablo 4.5.'te yer almaktadır.

**Tablo 4. 5. Madde Havuzu Büyüklüğü 250, Başlama kuralı  $b=0$  ve Yetenek Kestirimi EAP Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri\***

Sonlandırma Koşulları			RMSE	Yanlılık	Uyum Katsayısı	Test Uzunluğu
K1	Sabit Uzunluk	15	0,269	0,010	0,988	15,0
K2		20	0,227	0,007	0,991	20,0
K3	SE( $\theta$ )	0.385	0,463	0,016	0,965	6,9
K4		0.315	0,363	0,009	0,978	9,3
K5		0.220	0,229	0,002	0,991	19,1
K6	SE( $\theta$ ) ve en az 10 madde	0.385	0,348	0,013	0,980	10,1
K7		0.315	0,337	0,011	0,981	10,5
K8		0.220	0,229	0,002	0,991	19,1
K9	(θ) değerinde yakınsama	0.05	0,250	0,006	0,989	16,9
K10		0.02	0,189	0,003	0,994	30,0
K11	(θ) değerinde yakınsama ve en az 10 madde	0.05	0,245	0,007	0,990	16,9
K12		0.02	0,189	0,003	0,994	30,0

\* Değerlerin tümü 25 tekrardan elde edilen verilerin ortalaması hesaplanarak verilmiştir.

Tablo 4.5. incelendiğinde; en düşük RMSE değerinin **0,02 theta yakınsama (tek başına ve en az 10 madde ile kullanımı)** sonlandırma koşulunda elde edildiği görülmektedir. Bu koşulu sırasıyla **sabit uzunluk 20 madde, 0,220 SE(theta)** (tek



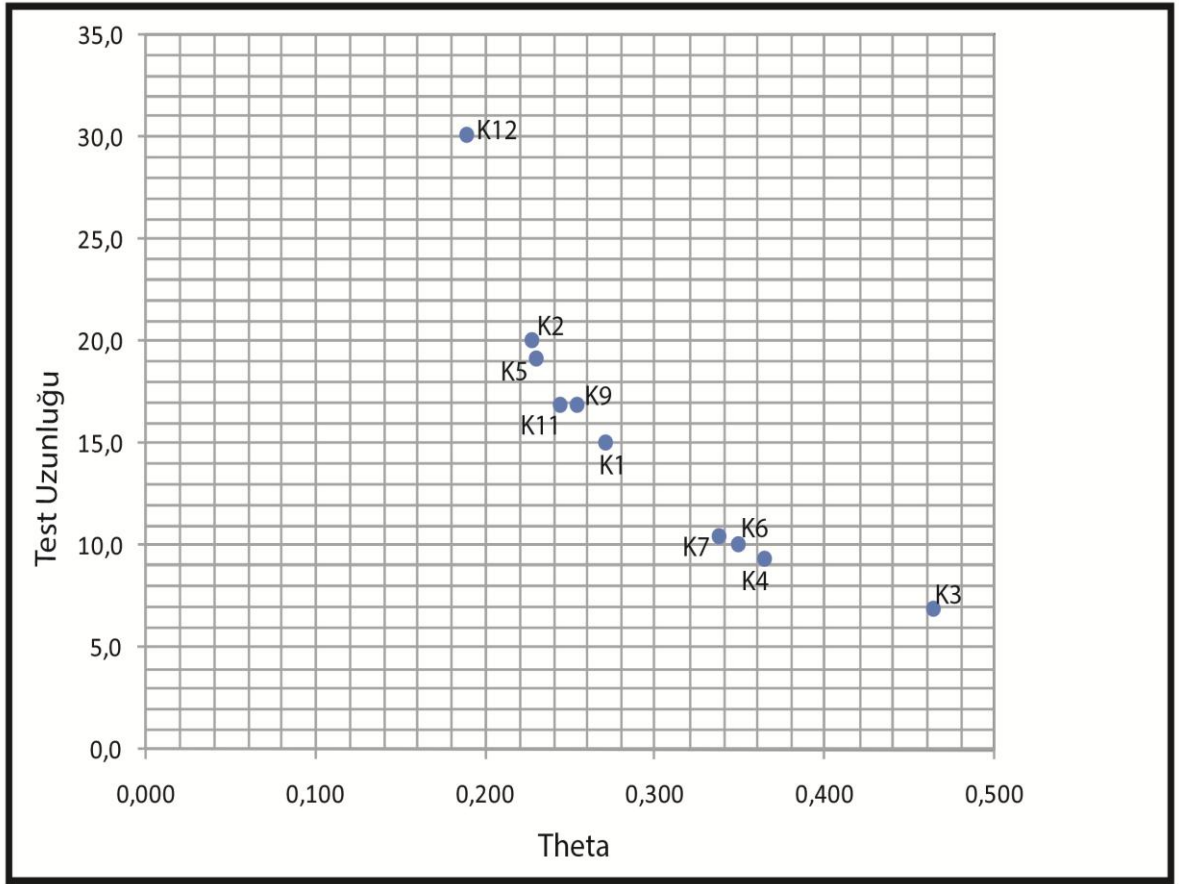
başına ve en az 10 madde ile kullanımı) izlemektedir. RMSE değerinin en yüksek elde edildiği sonlandırma koşulları sırasıyla **0,385 SE(theta); 0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 maddedir.**

Yanlılık değerleri RMSE değerleri ile büyük ölçüde paralellik göstermektedir. En düşük yanlılık değeri **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) koşulunda elde edilmiştir. Bu koşulu sırasıyla **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) ve **0,05 theta yakınsama** koşulları izlemektedir. En yüksek yanlılık değeri ise **0,385 SE(theta)** sonlandırma koşulunda elde edilmiş olup bunu sırasıyla **0,385 SE(theta)-en az 10 madde** ve **0,315 SE(theta)-en az 10 madde** koşulları izlemektedir.

Bu alt problem için uyum katsayıları tüm koşullarda birbirine çok yakın ve yüksek elde edilmiştir. Hesaplanan ile gerçek theta arasındaki uyum sırasıyla en yüksek **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı); **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) ve **sabit uzunluk 20 madde** koşullarında elde edilmiştir. Uyum katsayısı en düşük olarak ise sırasıyla **0,385 SE(theta); 0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşullarının kullanıldığı durumda elde edilmiştir.

Ortalama test uzunlukları incelendiğinde; testin ortalama en az madde ile **0,385 SE(theta)** (ortalama 7 madde) sonlandırma koşulunda sonlandırıldığı belirlenmiştir. Bu koşulu sırasıyla **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşulları izlemektedir. Testin ortalama en çok madde ile sırasıyla **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) (ortalama 30 madde), **sabit uzunluk 20 madde** ve **0,220 SE(theta)** koşullarında sonlandırıldığı gözlemlenmiştir.

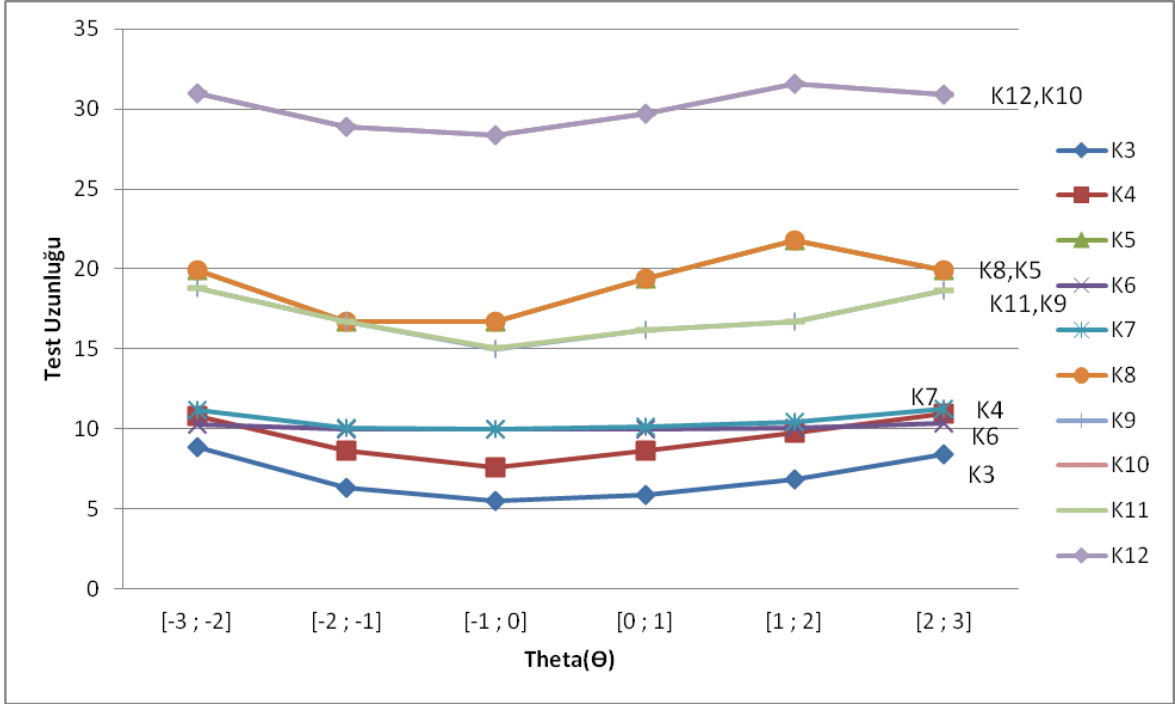
Aşağıda yer alan Şekil 4.9.'da bu alt problem için her bir sonlandırma koşuluna ait ortalama test uzunluğu ve RMSE değerlerinin ilişkisi yer almaktadır.



**Şekil 4. 9. Beşinci Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki**

Şekil 4.9. incelendiğinde; beşinci alt problem için en az 10 madde koşulunun, theta değerinde yakınsama ile birlikte kullanımının RMSE ve testin ortalama uzunluğunda kullanılmadığı koşula kıyasla önemli bir değişime neden olmadığı görülmektedir. Theta değerinde yakınsama kuralından farklı olarak, en az 10 madde koşulu **0,385** ve **0,315 SE(theta)** sonlandırma koşulları ile birlikte kullanıldığında RMSE değerlerinde iyileşme meydana gelmiştir.

Test uzunluklarının, bireylerin yetenek düzeyine göre, sürekli dağılımı Ek-8'de yer almaktadır Ayrıca yetenek düzeyi (-3 ve +3 theta arasında değişen) 6 eşit birime bölünmüş ve 8 koşula (K3-K10) ait ortalama test uzunlukları, Şekil 4.10.'da verilmiştir.



**Şekil 4. 10. Beşinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu**

Şekil 4.10. incelendiğinde değişen uzunluk sonlandırma koşullarında; test uzunluğunun genellikle [1;2] theta yetenek düzeyinde en yüksek değerine ulaştığı ve bu düzeyden sonra ise hızlı bir düşüş gösterdiği görülmektedir. Ayrıca [-3;-2] ile [2;3] theta yetenek düzeylerinde aynı ortalama test uzunluğu elde edilmiştir. Theta yakınsama kurallarında ve **0,220 SE(theta)** koşullarının tek başına veya en az madde ile kullanımı durumunda yetenek düzeylerinde test uzunluğunun değişmediği görülmektedir. Sonlandırma koşullarından **0,385** ve **0,315 SE(theta)** koşullarının en az 10 madde koşulu ile birlikte kullanımında ise yetenek düzeylerinde daha az değişim söz konusudur. Ayrıca **0,385** ve **0,315 SE(theta)** koşullarının en düşük test uzunluğu değerinin [-1;0] yetenek düzeyinde aldığı belirlenmiştir.

#### 4.1.6. Altıncı Alt Probleme İlişkin Bulgular

Bu alt problem için, BBT uygulamasında madde havuzu büyüklüğü 250; başlama kuralı  $-1 < b < +1$  ve yetenek kestirim yöntemi EAP seçildiğinde; ölçme kesinliği ve test uzunluğu incelenmiştir.

Tablo 3.1.'de yer alan koşullara göre; elde edilen yetenek kestirimlerine ait RMSE, yanlılık, uyum katsayısı ve test uzunluğuna ait ortalama değerler Tablo 4.6.'da yer almaktadır.

**Tablo 4. 6. Madde Havuzu Büyüklüğü 250, Başlama kuralı  $-1 < b < 1$  ve Yetenek Kestirimi EAP Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri\***

Sonlandırma Koşulları			RMSE	Yanlılık	Uyum Katsayısı	Test Uzunluğu
K1	Sabit Uzunluk	15	0,270	0,009	0,988	15,0
K2		20	0,231	0,007	0,991	20,0
K3	SE( $\theta$ )	0.385	0,478	0,022	0,962	6,9
K4		0.315	0,362	0,005	0,978	9,4
K5		0.220	0,230	0,002	0,991	19,2
K6	SE( $\theta$ ) ve en az 10 madde	0.385	0,350	0,013	0,980	10,1
K7		0.315	0,336	0,011	0,981	10,6
K8		0.220	0,230	0,002	0,991	19,2
K9	(θ) değerinde yakınsama	0.05	0,246	0,004	0,990	17,0
K10		0.02	0,190	0,001	0,994	30,3
K11	(θ) değerinde yakınsama ve en az 10 madde	0.05	0,245	0,007	0,990	17,0
K12		0.02	0,190	0,003	0,994	30,2

\* Değerlerin tümü 25 tekrardan elde edilen verilerin ortalaması hesaplanarak verilmiştir.

Tablo 4.6. incelendiğinde; en düşük RMSE değerinin **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) sonlandırma koşulunda elde edildiği görülmektedir. Bu koşulu sırasıyla **0,220 SE(theta)** ve **sabit uzunluk 20 madde** koşulları izlemektedir. RMSE değerinin en yüksek elde edildiği sonlandırma

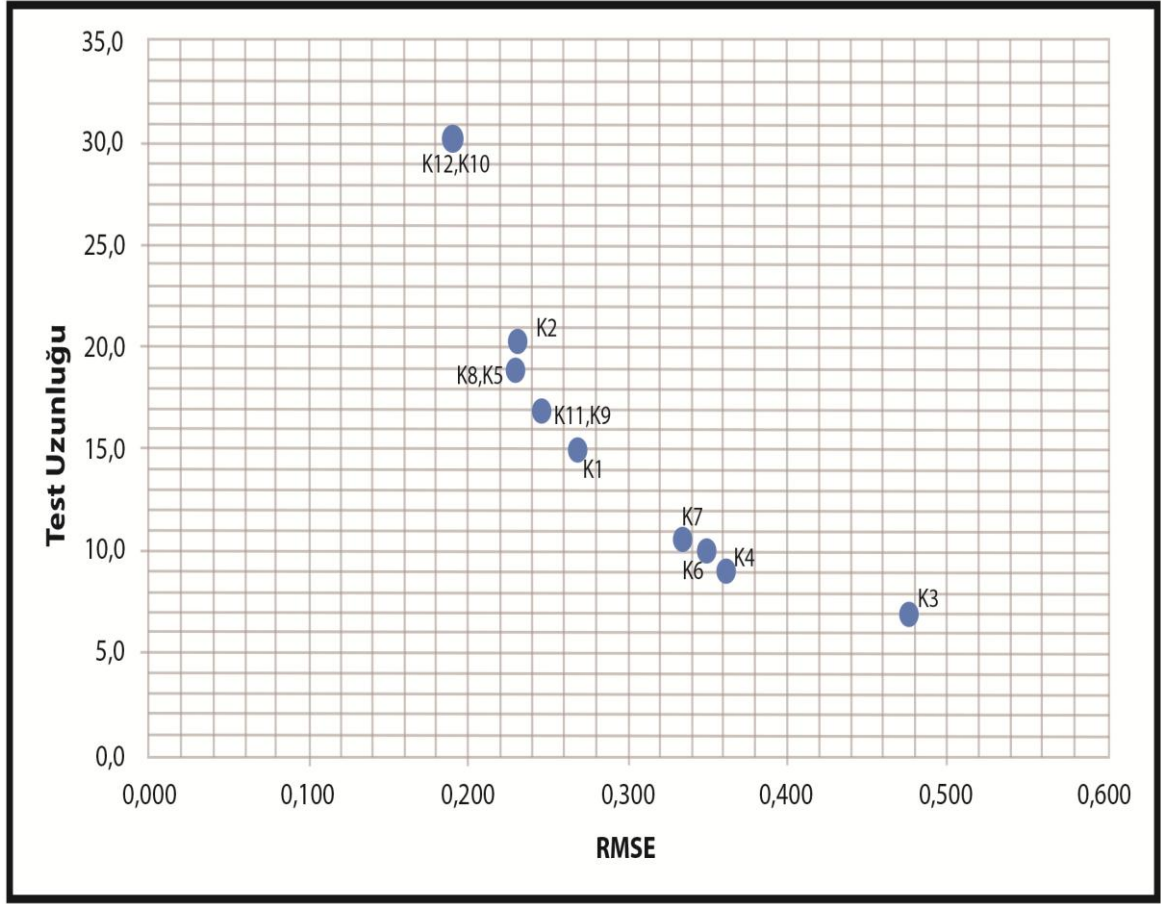
koşulları sırasıyla **0,385 SE(theta)**; **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** koşullarıdır.

Tablo 4.6.'da yer alan yanlılık değerleri incelendiğinde; **0,02 theta yakınsama** koşulunda en düşük yanlılık değeri elde edilmiş olup daha sonra sırasıyla **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) ve **0,02 theta yakınsama-en az 10 madde** koşulları gelmektedir. En yüksek yanlılık değeri ise **0,385 SE(theta)** sonlandırma koşulunda elde edilmiştir. Bu koşulu ise sırasıyla **0,385 SE(theta)-en az 10 madde** ve **0,315 SE(theta)-en az 10 madde** koşulları izlemektedir.

Gerçek theta ile hesaplanan theta değerleri arasındaki uyum katsayıları tüm koşullarda birbirine çok yakın ve yüksek elde edilmiştir. Uyum katsayısı sırasıyla en yüksek **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı); **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) ve **sabit uzunluk 20 madde** koşullarında elde edilmiştir. Uyum katsayısı en düşük değerini ise **0,385 SE(theta)**, **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşullarının kullanıldığı durumda almaktadır.

Ortalama test uzunlukları incelendiğinde; testin ortalama en az madde ile **0,385 SE(theta)** (ortalama 7 madde) sonlandırma koşulunda sonlandırıldığı belirlenmiştir. Bu koşulu sırasıyla **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşulları izlemektedir. Testin ortalama en çok madde ile sırasıyla **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) (ortalama 30 madde), **sabit uzunluk 20 madde** ve **0,220 SE(theta)** koşullarında sonlandırıldığı gözlemlenmiştir.

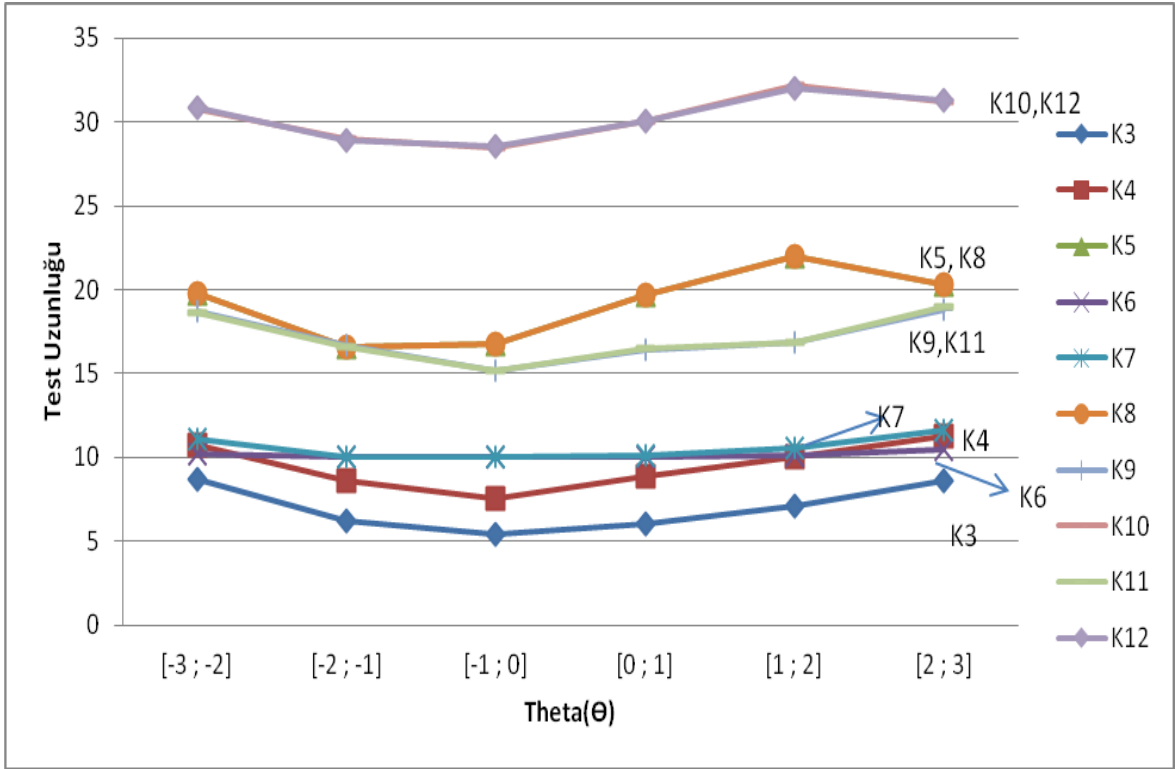
Aşağıda yer alan Şekil 4.11.'de bu alt problem için her bir sonlandırma koşuluna ait ortalama test uzunluğu ve RMSE değerlerinin ilişkisi yer almaktadır.



**Şekil 4. 11. Altıncı Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki**

Altıncı alt problem için en az 10 madde koşulunun, theta değerinde yakınsama kuralı ile birlikte kullanımının RMSE değeri ve testin ortalama uzunluğunda; kullanılmadığı koşula kıyasla önemli bir değişime neden olmadığı belirtilebilir. Farklı olarak en az 10 madde koşulu diğer alt problemlerdekine benzer şekilde **0,385 ve 0,315 SE(theta)** sonlandırma koşulları ile birlikte kullanıldığında RMSE değerinde iyileşmeye neden olmaktadır.

Test uzunluklarının, bireylerin yetenek düzeyine göre, sürekli dağılımı Ek-9'da yer almaktadır. Ayrıca yetenek düzeyi (-3 ve +3 theta arasında değişen) 6 eşit birime bölünmüş ve 8 koşula (K3-K10) ait ortalama test uzunlukları, Şekil 4.12.'de verilmiştir.



**Şekil 4. 12. Altıncı Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu**

Şekil 4.12. incelendiğinde değişen uzunluk sonlandırma koşullarından **0,02 theta yakınsama** ve **0,05 theta yakınsama** koşullarının tek başına veya en az 10 madde ile kullanımının tüm yetenek düzeylerinde ortalama test uzunluğu dağılımı benzer çıkmıştır ve en yüksek değerini [1;2] theta yetenek düzeyinde almaktadır. Değişen uzunluk sonlandırma koşullarının hepsinde [-1;0] theta yetenek düzeyinde en düşük uzunlukta test uygulanmıştır. **0,220 SE(theta)** koşulunun tek başına ve en az 10 madde ile kullanımı durumunda ise en yüksek test uzunluğu değerini uç yetenek düzeylerinde aldığı görülmektedir. Ayrıca **0,385 ve 0,315 SE(theta)** koşullarının en düşük test uzunluğu değerinin [-1;0] yetenek düzeyinde aldığı belirlenmiştir.

#### 4.1.7. Yedinci Alt Probleme İlişkin Bulgular

Bu alt problem için, BBT uygulamasında madde havuzu büyüklüğü 500; başlama kuralı  $b=0$  ve yetenek kestirim yöntemi EAP seçildiğinde; ölçme kesinliği ve test uzunluğu incelenmiştir.

Tablo 3.1.'de yer alan koşullara göre; elde edilen yetenek kestirimlerine ait RMSE, yanlılık, uyum katsayısı ve test uzunluğuna ait ortalama değerler Tablo 4.7.'de yer almaktadır.

**Tablo 4. 7. Madde Havuzu Büyüklüğü 500, Başlama kuralı  $b=0$  ve Yetenek Kestirimi EAP Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri\***

Sonlandırma Koşulları		RMSE	Yanlılık	Uyum Katsayısı	Test Uzunluğu	
K1	Sabit Uzunluk	15	0,247	0,004	0,990	15,0
K2		20	0,208	0,002	0,993	20,0
K3	SE( $\theta$ )	0.385	0,466	0,014	0,964	6,2
K4		0.315	0,363	0,009	0,978	8,4
K5		0.220	0,235	0,002	0,991	15,8
K6	SE( $\theta$ ) ve en az 10 madde	0.385	0,324	0,004	0,983	10,1
K7		0.315	0,313	0,005	0,984	10,3
K8		0.220	0,235	0,002	0,991	15,8
K9	(θ) değerinde yakınsama	0.05	0,228	0,003	0,991	16,7
K10		0.02	0,168	0,001	0,995	30,5
K11	(θ) değerinde yakınsama ve en az 10 madde	0.05	0,227	0,003	0,991	16,8
K12		0.02	0,168	0,000	0,995	30,5

\* Değerlerin tümü 25 tekrardan elde edilen verilerin ortalaması hesaplanarak verilmiştir.

Tablo 4.7. incelendiğinde; en düşük RMSE değerinin **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) sonlandırma koşulunda elde edildiği



görülmektedir. Bu koşulu sırasıyla **sabit uzunluk 20 madde** ve **0,05 theta yakınsama-en az 10 madde** koşulları izlemektedir.

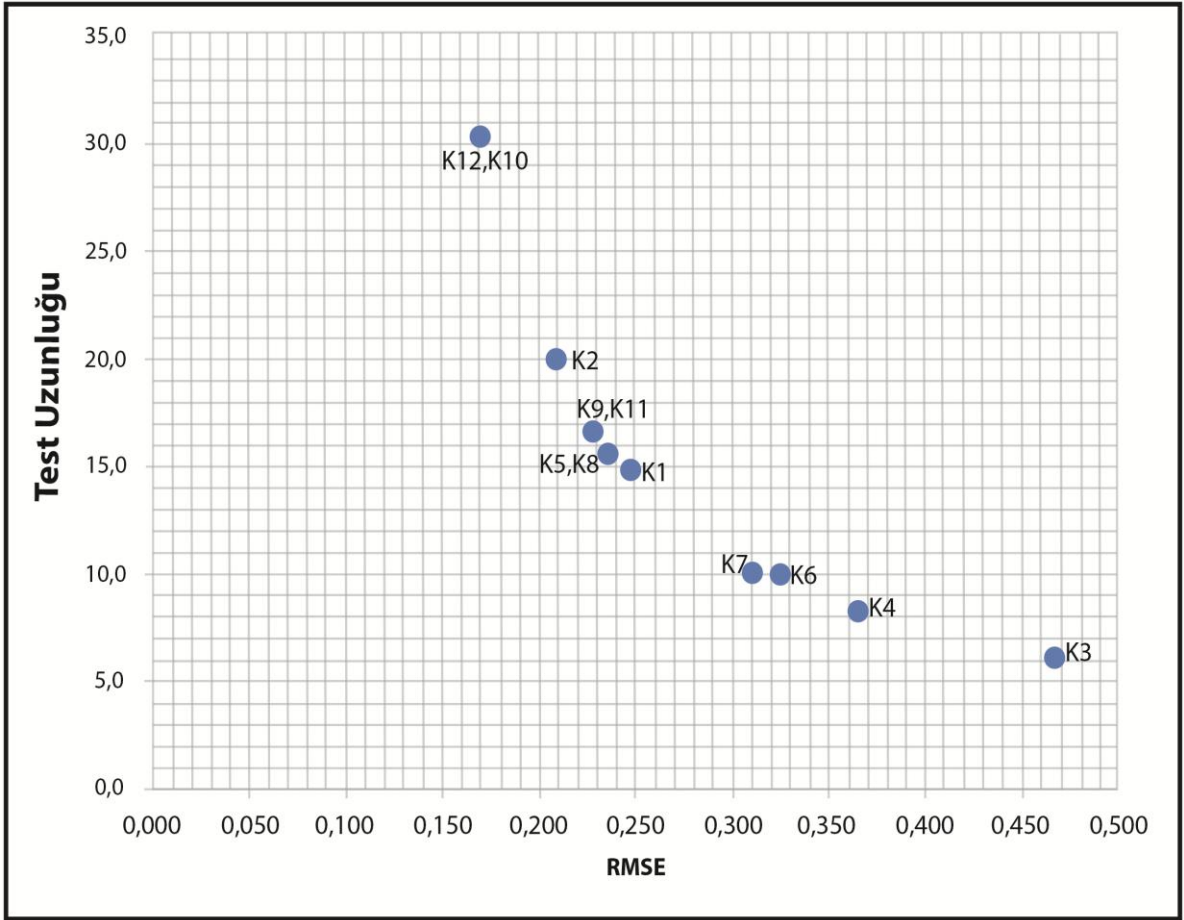
RMSE değerinin en yüksek elde edildiği sonlandırma koşulları sırasıyla **0,385 SE(theta); 0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** koşullarıdır.

Tablo 4.7.'de yer alan yanlılık değerleri incelendiğinde; **0,02 theta yakınsama-en az 10 madde** koşulunda en düşük yanlılık değeri elde edilmiş olup daha sonra sırasıyla **0,02 theta yakınsama** ve **0,220 SE(theta)**(tek başına ve en az 10 madde) ile **sabit uzunluk 20 madde** koşulları gelmektedir. En yüksek yanlılık değeri ise **0,385 SE(theta)** sonlandırma koşulunda elde edilmiştir. Bu koşulu ise sırasıyla **0,315 SE(theta)** ve **0,315 SE(theta)-en az 10 madde** koşulları izlemektedir.

Gerçek theta ile hesaplanan theta değerleri arasındaki uyum katsayıları tüm koşullarda birbirine çok yakın ve yüksek elde edilmiştir. Uyum katsayısı sırasıyla en yüksek **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı); **sabit uzunluk 20 madde** ve **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) koşullarında elde edilmiştir. Uyum değerinin en düşük **0,385 SE(theta), 0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşullarının kullanıldığı durumda elde edilmiştir.

Ortalama test uzunlukları incelendiğinde; testin ortalama en az madde ile **0,385 SE(theta)** (ortalama 6 madde) sonlandırma koşulunda sonlandırıldığı belirlenmiştir. Bu koşulu sırasıyla **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşulları izlemektedir. Testin ortalama en çok madde ile sırasıyla **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) (ortalama 31 madde); **sabit uzunluk 20 madde** ve **0,05 theta yakınsama-en az 10 madde** koşullarında sonlandırıldığı gözlemlenmiştir.

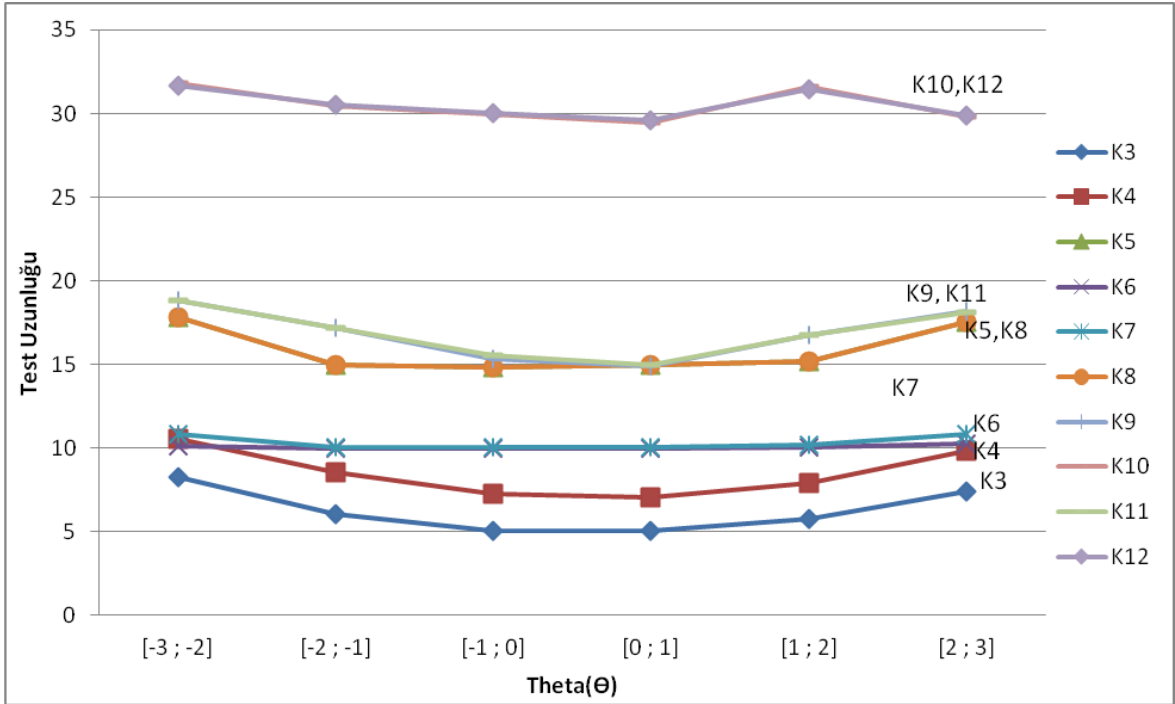
Aşağıda yer alan Şekil 4.13.'de bu alt problem için her bir sonlandırma koşuluna ait ortalama test uzunluğu ve RMSE değerlerinin ilişkisi yer almaktadır.



**Şekil 4. 13. Yedinci Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki**

Bu alt problemde de en az 10 madde koşulu, theta yakınsama kuralları ile birlikte kullanıldığında RMSE ve test uzunluğunda bir değişime neden olmadığı görülmektedir. Diğer alt problemdekine benzer şekilde, **0,385** ve **0,315 SE(theta)** sonlandırma koşulu ile birlikte kullanıldığında RMSE ve yanlılık değerlerinde iyileşmeye neden olmaktadır.

Test uzunluklarının, bireylerin yetenek düzeyine göre, sürekli dağılımı Ek-10'da yer almaktadır Ayrıca yetenek düzeyi (-3 ve +3 theta arasında değişen) 6 eşit birime bölünmüş ve 8 koşula (K3-K10) ait ortalama test uzunlukları, Şekil 4.14.'te verilmiştir.



**Şekil 4. 14. Yedinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu**

Şekil 4.14. incelendiğinde değişen uzunluk sonlandırma koşullarının çoğunluğunda [0;1] theta düzeyinde ortalama test uzunluğu en düşük değerini almaktadır. Standart hata sonlandırma kuralında ise uç yetenek düzeylerinde yüksek diğer yetenek düzeylerinde daha düşük test uzunluğu görülmektedir. Theta yakınsama kurallarında ve **0,220 SE(theta)** koşullarının tek başına veya en az madde ile kullanımı durumunda yetenek düzeylerinde test uzunluğunun değişmediği görülmektedir. **0,315 ve 0,385 SE(theta)** koşullarının **en az 10 madde** koşulu ile birlikte kullanılmasında ise yetenek düzeylerinde çok az değişim gözlenmiştir.

#### 4.1.8. Sekizinci Alt Probleme İlişkin Bulgular

Bu alt problem için, BBT uygulamasında madde havuzu büyüklüğü 500; başlama kuralı  $-1 < b < +1$  ve yetenek kestirim yöntemi EAP seçildiğinde; ölçme kesinliği ve test uzunluğu incelenmiştir.

Tablo 3.1.'de yer alan koşullara göre; elde edilen yetenek kestirimlerine ait RMSE, yanlılık, uyum katsayısı ve test uzunluğuna ait ortalama değerler Tablo 4.8.'de yer almaktadır.

**Tablo 4. 8. Madde Havuzu Büyüklüğü 500, Başlama kuralı  $-1 < b < 1$  ve Yetenek Kestirimi EAP Seçildiğinde; Ortalama RMSE, Yanlılık, Uyum ve Test Uzunluğu Değerleri\***

Sonlandırma Koşulları			RMSE	Yanlılık	Uyum Katsayısı	Test Uzunluğu
K1	Sabit Uzunluk	15	0,248	0,005	0,990	15,0
K2		20	0,209	0,004	0,992	20,0
K3	SE( $\theta$ )	0.385	0,477	0,017	0,963	6,3
K4		0.315	0,370	0,012	0,977	8,6
K5		0.220	0,234	0,002	0,991	15,9
K6	SE( $\theta$ ) ve en az 10 madde	0.385	0,330	0,006	0,982	10,1
K7		0.315	0,319	0,006	0,983	10,3
K8		0.220	0,234	0,002	0,991	15,9
K9	$(\theta)$ değerinde yakınsama	0.05	0,226	0,002	0,991	16,9
K10		0.02	0,168	0,001	0,995	30,6
K11	$(\theta)$ değerinde yakınsama ve en az 10 madde	0.05	0,227	0,003	0,991	16,9
K12		0.02	0,168	0,001	0,995	30,6

\* Değerlerin tümü 25 tekrardan elde edilen verilerin ortalaması hesaplanarak verilmiştir.

Tablo 4.8. incelendiğinde; en düşük RMSE değerinin **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) sonlandırma koşulunda elde edildiği görülmektedir. Bu koşulu sırasıyla **sabit uzunluk 20 madde** ve **0,05 theta yakınsama** koşulları izlemektedir. RMSE değerinin en yüksek elde edildiği

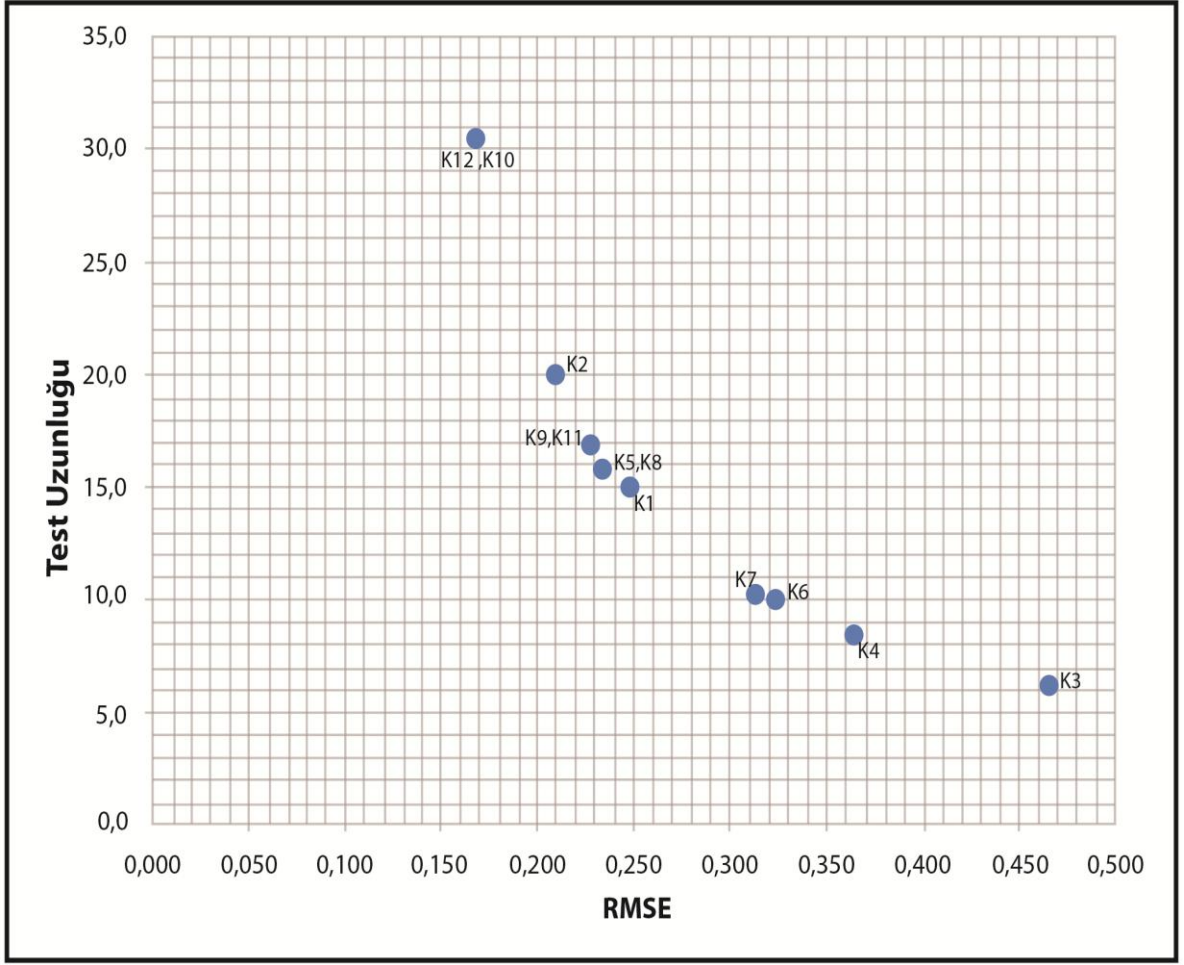
sonlandırma koşulları sırasıyla **0,385 SE(theta)**; **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** koşullarıdır.

Tablo 4.8.'de yer alan yanlılık değerleri incelendiğinde ise; **0,02 theta yakınsama**(tek başına ve en az 10 madde ile kullanımı) koşulunda en düşük yanlılık değeri elde edilmiştir. Bu koşulu ise sırasıyla **0,220 SE(theta)** (tek başına ve 10 madde ile kullanımı) ve **0,05 theta yakınsama** (tek başına ve 10 madde ile kullanımı) koşulları izlemektedir. En yüksek yanlılık değeri ise **0,385 SE(theta)** sonlandırma koşulunda elde edilmiş olup sırasıyla **0,315 SE(theta)** ve **0,385-0,315 SE(theta)-en az 10 madde** koşulları izlemektedir.

Gerçek theta ile hesaplanan theta değerleri arasındaki uyum katsayıları tüm koşullarda birbirine çok yakın ve yüksek elde edilmiştir. Uyum katsayısı sırasıyla en yüksek **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı); **sabit uzunluk 20 madde** ve **0,220 SE(theta)** (tek başına ve en az 10 madde ile kullanımı) koşullarında elde edilmiştir. Uyum katsayısı en düşük değerini sırasıyla **0,385 SE(theta)**, **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** ile **0,05 theta yakınsama** sonlandırma koşullarının kullanıldığı durumda elde edilmiştir.

Ortalama test uzunlukları incelendiğinde; testin ortalama en az madde ile **0,385 SE(theta)** (ortalama 6 madde) sonlandırma koşulunda sonlandırıldığı belirlenmiş olup sırasıyla **0,315 SE(theta)** ve **0,385 SE(theta)-en az 10 madde** sonlandırma koşulları gelmektedir. Testin ortalama en çok madde ile sırasıyla **0,02 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) (ortalama 31 madde); **sabit uzunluk 20 madde** ve **0,05 theta yakınsama** (tek başına ve en az 10 madde ile kullanımı) koşullarında sonlandırıldığı gözlemlenmiştir.

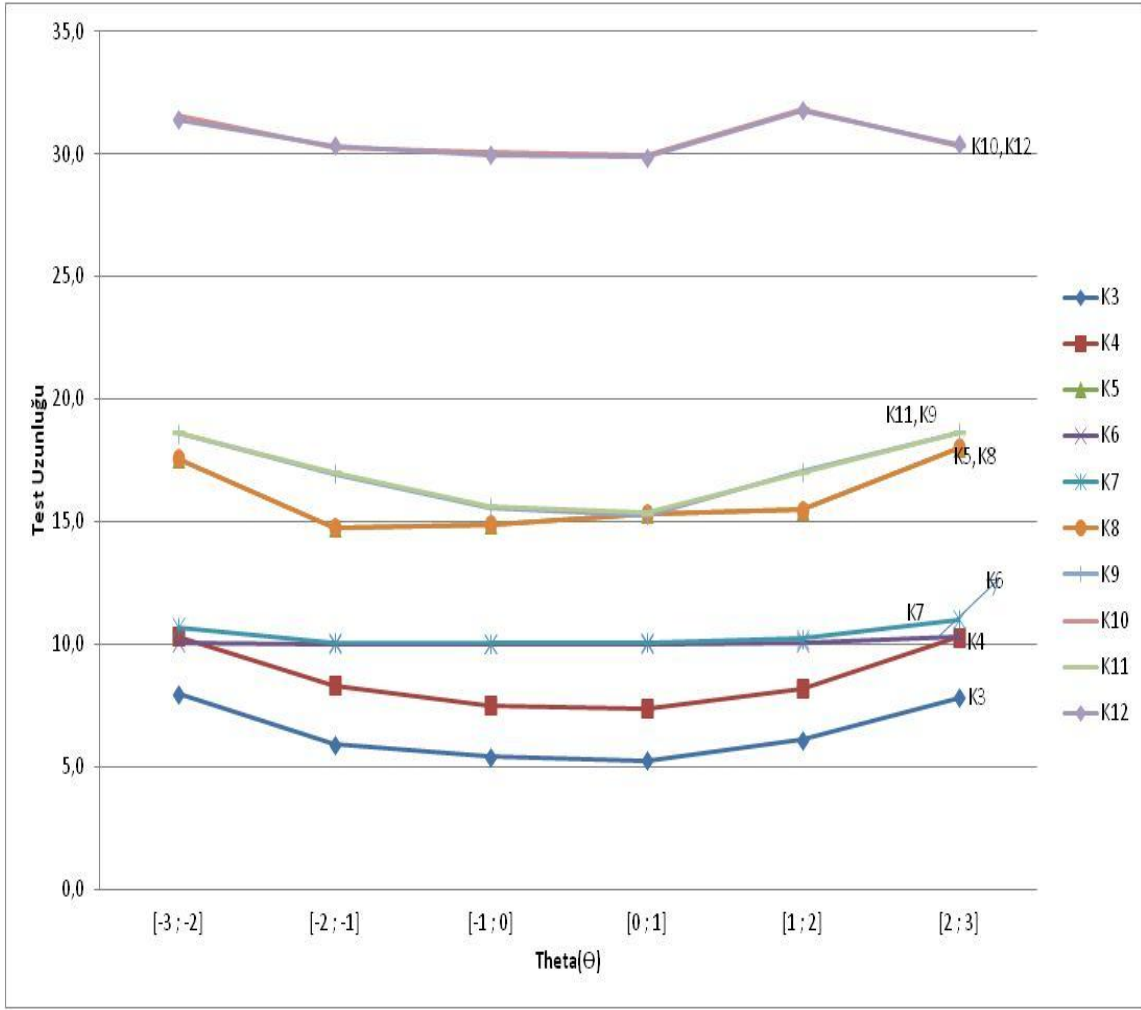
Aşağıda yer alan Şekil 4.15.'te bu alt problem için her bir sonlandırma koşuluna ait ortalama test uzunluğu ve RMSE değerlerinin ilişkisi yer almaktadır.



**Şekil 4. 15. Sekizinci Alt Probleme Ait Test Uzunluğu ve RMSE Değerleri Arasındaki İlişki**

Bu alt problemde de en az 10 madde koşulu diğer alt problemlerdekine benzer şekilde theta yakınsama kuralı ile birlikte kullanıldığında RMSE ve ortalama test uzunluğu değerlerinde önemli bir değişime neden olmamaktadır. Ancak **0,385 ve 0,315 SE(theta)** sonlandırma koşulu ile birlikte kullanıldığında RMSE ve yanlılık değerlerinde iyileşmeye neden olmaktadır.

Test uzunluklarının, bireylerin yetenek düzeyine göre, sürekli dağılımı Ek-11'de yer almaktadır Ayrıca yetenek düzeyi (-3 ve +3 theta arasında değişen) 6 eşit birime bölünmüş ve 8 koşula (K3-K10) ait ortalama test uzunlukları, Şekil 4.16.'da verilmiştir.



**Şekil 4. 16. Sekizinci Alt Probleme Ait Yetenek Düzeylerinde Ortalama Test Uzunluğu**

Şekil 4.16. incelendiğinde değişen uzunluk sonlandırma koşullarının çoğunluğunda [0;1] theta düzeyinde ortalama test uzunluğu en düşük değerini almaktadır. Standart hata sonlandırma kuralında ise uç yetenek düzeylerinde yüksek diğer yetenek düzeylerinde daha düşük test uzunluğu görülmektedir. Theta yakınsama kurallarında ve **0,220 SE(theta)** koşulunun tek başına veya en az 10 madde ile kullanımı durumunda yetenek düzeylerinde test uzunluğunun değişmediği görülmektedir. **0,385 ve 0,315 SE(theta)** koşullarının en az 10 madde ile birlikte kullanılması yetenek düzeylerinde test uzunluğu açısından çok önemli bir farklılık gözlenmemiştir.

## 4.2. Tartışma ve Yorum

Çalışmada ölçme kesinliğini belirlemek amacıyla RMSE, yanlılık ve uyum değerleri hesaplanmıştır. Ayrıca her bir koşul için test uzunluğu değeri de verilmiştir. Her bir alt problemde toplam 5 farklı sonlandırma kuralına göre oluşturulan 12 koşulda ortalama RMSE, yanlılık ve uyum değerlerinin farklılık gösterdiği belirlenmiştir.

Sabit uzunluk sonlandırma kuralına ait 15 ve 20 maddelik koşul ölçme kesinliği bakımından kıyaslandığında tüm alt problemlerde 20 maddelik test daha etkili çıkmıştır. Sabit uzunluk kuralında tüm bireylere testin belirli bir ölçme kesinliğine ulaşması beklenmeksizin aynı sayıda maddenin uygulanması söz konusudur. Sabit uzunluk 20 madde sonlandırma koşulunun daha az hata göstergesi değerlerini içermesi, daha fazla madde uygulanmasından kaynaklanmaktadır. Bu bulgular Babcock ve Weiss (2012)' in araştırma sonuçlarıyla desteklenir niteliktedir. Babcock ve Weiss (2012) dört farklı özelliğe sahip madde havuzlarında toplam 14 sonlandırma koşuluna ait ortalama RMSE, yanlılık ve ortalama test uzunluklarını vermişlerdir. Ortalama test uzunluğu arttıkça RMSE ve yanlılık değerlerinde düşme yani ölçme kesinliğinde artış belirlenmiştir. Choi, Graddy ve Dodd (2011) çalışmalarında 3 farklı sonlandırma kuralında uygulanan ortalama madde sayılarını ve RMSE değerlerini vermişlerdir. Ortalama madde sayısının düşük olduğu minimum bilgi sonlandırma kuralında RMSE değeri de diğerlerine kıyasla daha düşük çıkmıştır.

Standart hata sonlandırma kuralına ait 0,385 SE(theta), 0,315 SE(theta) ve 0,220 SE(theta) sonlandırma koşulları tüm alt problemlerde benzer şekilde sonuç üretmiştir. Ölçme kesinliği en yüksek 0,220 SE(theta); en düşük ise 0,385 SE(theta) koşulunda elde edilmiştir. Testlerin güvenilirliği arttıkça ölçme sonuçlarına ait hata gösterge değerlerinin azalması yani ölçme kesinliğinin artması beklenen bir durumdur. Bu bulgular Babcock ve Weiss (2012)'in araştırma sonuçlarıyla desteklenmektedir. Babcock ve Weiss (2012), 3 farklı güvenilirlik katsayısına bağlı olarak oluşturduğu bu değerleri kullanmışlardır ve ölçme kesinliği bu araştırma sonuçlarıyla paralel bulunmuştur. Ortalama test uzunluklarına da bakıldığında en düşük değer 0,385 SE(theta) sonlandırma koşulunda en yüksek değerin ise 0,220 SE(theta) koşulunda elde edildiği görülmektedir. Özellikle Bayes



yöntemlerde test uzunluğu kısalдықça sonuçların daha yanlı olabileceği ile ilgili çalışmalara literatürde rastlanmaktadır.

Theta düzeyinde yakınsama sonlandırma kuralında 0,02 ve 0,05 yakınsama koşulları kullanılmıştır. Araştırmada alt problemlerde 0,02 theta yakınsama koşulunda ölçme kesinliği daha yüksek çıkmıştır. Ortalama test uzunluklarına bakıldığında da ortalama test uzunluğu değeri de 0,02 theta yakınsama koşulunda daha yüksek çıkmıştır.

Sabit sonlandırma, standart hata ve theta yakınsama sonlandırma kuralı içerisinde; 0,02 theta yakınsama koşulunun en düşük RMSE değerini aldığı görülmüştür. Literatürde araştırmaların sonuçlarını destekleyecek çalışmalara rastlanmaktadır. Hart, Mioduski ve Stratford (2005) ve Hart, Cook, Mioduski, Teal ve Crane (2006) ve araştırmalarında standart hata ve theta yakınsama sonlandırma kurallarının  $SE(\theta) < 0,4$  ve theta değişimi  $< 0,1$ 'in altına düşmesi koşullarını uygulamışlardır. Çalışmalarında her iki koşulu birlikte kombine etmişlerdir. Theta yakınsama kuralının BBT uygulamalarında theta tahmininde iyi sonuçlar ürettiği belirlenmiştir (akt: Babcock ve Weiss, 2012). Ayrıca sonlandırma koşulları testte uygulanan madde sayıları bağlamında karşılaştırıldığında yine en yüksek test uzunluğunun 0,02 theta yakınsama koşulunda elde edildiği görülmektedir.

Alt problemlerin genelinde değişen uzunluklu sonlandırma kurallarından olan theta yakınsama kuralından 0,02 theta yakınsama koşulunun kullanıldığı durumda sabit uzunluk kuralının her iki koşuluna kıyasla ölçme kesinliğinin daha yüksek olduğu ortaya çıkmaktadır. Ayrıca standart hata sonlandırma kuralının 0,220  $SE(\theta)$  koşulu ile sabit uzunluk 15 madde koşuluna kıyasla daha düşük RMSE ve yanlılık değerleri elde edilmiştir. Literatürde de değişen uzunluklu sonlandırma kurallarının sabit uzunluk sonlandırma kuralına kıyasla daha güçlü ölçme sonuçları sunduğuna dair araştırmalar söz konusudur (Weiss, 1982; Babcock ve Weiss, 2012)

Araştırmada alt problemlerin genelinde sabit uzunluk sonlandırma kuralının tüm koşullarında ölçme kesinliği standart hata sonlandırma kuralının 0,385  $SE(\theta)$  ve 0,315  $SE(\theta)$  koşullarına kıyasla daha yüksek elde edilmiştir. Ayrıca sabit uzunluk sonlandırma kuralının 20 madde koşulunun kullanıldığı test durumunda,

theta yakınsama sonlandırma kuralının 0,05 koşulu ile karşılaştırıldığında da ölçme etkililiği daha yüksek elde edilmiştir. Literatürde de benzer sonuçları veren çalışmalar söz konusudur (Stocking,1987; Yi, Wang ve Ban, 2001).

En az 10 madde koşulunun standart hata ve theta yakınsama sonlandırma kuralları ile birlikte uygulanmasında özellikle ortalama 10 maddenin aşağısında testin sonlandırıldığı koşullara ölçme etkililiği bakımından avantaj sağladığı görülmektedir. Araştırmada standart hata ile minimum madde sayısının kombinasyonu ölçme kesinliği açısından standart hatanın tek başına kullanıldığı duruma göre avantaj sağlamıştır. Bu sonuçlar Babcock ve Weiss (2012)'in araştırma sonuçları ile paralellik göstermektedir.

Farklı sonlandırma kurallarının uyum katsayılarını önemli oranda etkilemediği gözlenmiştir. Uyum katsayısı değerleri RMSE, yanlılık değerleri ile genellikle benzerlik göstermektedir. En düşük uyum katsayısına 0,385 SE(theta) sonlandırma koşulunda rastlanmakta iken, en yüksek uyum katsayısına ise 0,02 theta yakınsama koşulunda rastlanmaktadır.

Ortalama test uzunluğu ve RMSE değerleri arasındaki ilişki incelendiğinde tüm alt problemlerde birbirine oldukça benzer olduğu görülmektedir. RMSE ve ortalama test uzunluğu arasında negatif bir ilişki söz konusudur. Her bir yetenek düzeyinde ortalama test uzunluklarına bakıldığında ise genel olarak her bir yetenek düzeyinde theta yakınsama kuralları ile daha çok madde uygulanırken, standart hata sonlandırma kuralları ile daha az madde uygulanmıştır. Belirli bir yetenek düzeyinde, madde sayısının az uygulandığı sonlandırma koşullarında ölçme kesinliği zayıf olacaktır (Choi, Graddy ve Dodd, 2011). En az madde koşulunun sonlandırma koşullarında birlikte kullanılması, hem testin geneli için RMSE değerini azaltmakta hem de yetenek düzeylerinde uygulanan madde sayısı da genellikle artış göstermektedir. Bu durumda en az madde koşulunun ölçme kesinliğinin artmasına katkıda bulunduğu yorumu yapılabilir. Daha iyi bir ölçme için değişen uzunluk sonlandırma kuralları, minimum madde sayısı ile desteklenerek daha fazla madde uygulanabilir. Bu durum yetenek seviyelerinin tahminlerinde istikrarı sağlayarak, ölçme kesinliği ve etkililiğini sağlayacaktır (Babcock ve Weiss, 2012).

## 5. SONUÇ VE ÖNERİLER

Bu bölümde araştırmanın sonuçları ve buna dayalı olarak yapılan öneriler yer almaktadır. Tüm koşulların toplu olarak karşılaştırıldığı ortalama RMSE, yanlılık, uyum katsayısı ve test uzunluğu değerlerine ait tablolar Ek-12, Ek-13, Ek-14, Ek-15'te verilmiştir.

### 5.1. Sonuçlar

Bu çalışmada 5 farklı sonlandırma kurallarına göre oluşturulan 12 koşul farklı madde havuzu büyüklüğü, başlama kuralı ve yetenek kestirim yöntemlerine göre incelenmiştir. Sonuçlar bu değişkenler göz önüne alınarak sonlandırma koşulları çerçevesinde verilmiştir.

#### 5.1.1. Madde Havuzu Büyüklüğünün 250 veya 500 Madde Olarak Belirlenmesi

- **Başlama kuralı  $b=0$  ve yetenek kestirimi MLE** seçildiğinde; belirlenen sonlandırma kurallarının çoğunluğunda 500 maddelik havuzda RMSE ve yanlılık değerleri daha düşük elde edilmiştir. Her iki havuz büyüklüğünde her bir sonlandırma kuralı için uygulanan madde sayılarına bakıldığında, çoğunluğunda çok önemli bir değişim olmadığı görülmüştür. Ancak 0,220 SE(theta) ve 0,220 SE(theta)-en az 10 madde koşullarının her ikisinde de RMSE ve yanlılık değerlerinde, 500 maddelik havuzda ortalama 3,1 madde daha az uygulanmasına rağmen çok önemli bir değişim olmadığı belirlenmiştir.
- **Başlama kuralı  $-1 < b < 1$  ve yetenek kestirimi MLE** seçildiğinde; RMSE ve yanlılık değerlerinin çoğunluğunu 500 maddelik havuzun lehine elde edilmiştir. Her iki havuz büyüklüğünde her bir sonlandırma kuralı için uygulanan madde sayılarında çoğu sonlandırma kuralı için önemli değişim olmadığı görülmüştür. Özellikle 0,220 SE(theta) ve 0,220 SE(theta)-en az 10 madde sonlandırma koşullarında, 500 maddelik havuzda ortalama 3 madde daha az uygulanmasına rağmen RMSE ve yanlılık değerlerinde çok önemli bir değişim olmadığı belirlenmiştir.

- **Başlama kuralı  $b=0$  ve yetenek kestirimi EAP** seçildiğinde; belirlenen sonlandırma kurallarının çoğunluğunda 500 maddelik havuzda RMSE değerleri daha düşük elde edilmiştir. Yanlılık değerlerine bakıldığında ise tüm sonlandırma koşullarında 500 maddelik havuzunun lehinedir. Her iki havuz büyüklüğünde her bir sonlandırma kuralı için uygulanan madde sayılarına bakıldığında, çoğunluğunda çok önemli değişim olmadığı görülmüştür. Geçerli koşullarda; göze çarpan en önemli nokta ise 0,220 SE(theta) ve 0,220 SE(theta)–en az 10 madde koşullarında, 500 maddelik havuzda ortalama 3,3 madde daha az uygulanmasına rağmen RMSE ve yanlılık değerlerinde çok önemli bir değişim olmadığıdır.
- **Başlama kuralı  $-1 < b < 1$  ve yetenek kestirimi EAP** seçildiğinde; belirlenen sonlandırma koşullarının çoğunluğunda 500 maddelik havuzda RMSE değerleri daha düşük elde edilmiştir. RMSE gibi yanlılık değerleri de sonlandırma koşullarının çoğunluğu için 500 maddelik havuzunun lehinedir. Her iki havuz büyüklüğünde her bir sonlandırma koşulu için uygulanan madde sayılarına bakıldığında, çoğunluğunda büyük madde havuzunda daha düşük madde sayısının uygulandığı görülmüştür. Geçerli koşullarda; 0,220 SE(theta) ve 0,220 SE(theta)–en az 10 madde koşullarında 500 maddelik havuzda ortalama 3,3 madde daha az uygulanarak, RMSE ve yanlılık değerlerinde çok önemli bir değişim olmadığı dikkat çekicidir.

Araştırmada, madde havuzu büyüklüğü ile RMSE değerleri arasında ters yönlü bir ilişki söz konusudur. Buna dayanarak, madde havuzu büyüklüğü arttıkça, ölçmenin kesinliğinin de arttığı söylenebilir. Choi, Graddy ve Dodd (2011) çalışmalarında 2 farklı özelliğe sahip, 2 farklı büyüklükte madde havuzlarında 3 farklı sonlandırma kuralını çoklu puanlanan maddeler için karşılaştırmışlardır. Aynı özelliğe sahip maddelerden oluşan havuzun büyüklüğü arttıkça RMSE değeri sonlandırma kurallarının hepsinde azalma göstermektedir. Ayrıca Ivie (2007); madde havuzu genişliğinin arttıkça daha yüksek ayırt edicilik ve daha düşük şans parametrelerinin, daha kararlı yetenek düzeyi tahmini sağladığını belirtmiştir.

### 5.1.2. Teste Başlama Kuralının $b=0$ veya $-1 < b < 1$ Belirlenmesi

- **Madde havuzu büyüklüğü 250 ve yetenek kestirimi MLE** seçildiğinde; RMSE değerleri sabit uzunluk sonlandırma kurallarında;  $-1 < b < 1$  başlangıç kuralının lehine olup, 0,385 ve 0,315 SE(theta) ile 0,05 theta yakınsama koşullarında ise  $b=0$  başlama kuralının lehinedir. Yanlılık değerleri karşılaştırıldığında çok önemli bir farklılık genel olarak gözlenmemiştir. Testte uygulanan madde sayıları arasında ise önemli bir farklılık göze çarpmamaktadır.
- **Madde havuzu büyüklüğü 500 ve yetenek kestirimi MLE** seçildiğinde; RMSE değerlerinde sabit uzunluk sonlandırma kurallarında başlangıç kurallarının seçimine dair bir farklılık gözlenmemiştir. 0,385 ve 0,315 SE(theta) koşullarında ise hem RMSE hem yanlılık değerleri  $b=0$  başlangıç kuralının lehinedir. 0,05 ve 0,02 theta yakınsama koşulunda ise  $b=0$  başlama kuralı seçiminde daha düşük RMSE değeri elde edilmiştir. Testte uygulanan madde sayıları arasında ise önemli bir farklılık göze çarpmamaktadır.
- **Madde havuzu büyüklüğü 250 ve yetenek kestirimi EAP** seçildiğinde; başlangıç kurallarının seçimine dair RMSE ve yanlılık değerleri birlikte değerlendirildiğinde önemli bir farklılık gözlenmemiştir. Yalnızca 0,385 SE(theta) koşulunun seçilmesi durumunda  $b=0$  başlangıç kuralında daha düşük RMSE değeri elde edilmiştir. Test uzunlukları açısından ise önemli bir farklılık görülmemiştir.
- **Madde havuzu büyüklüğü 500 ve yetenek kestirimi EAP** olarak belirlendiği durumda; başlangıç kurallarının  $b=0$  veya  $-1 < b < 1$  seçimine dair sonlandırma koşullarından elde edilen RMSE ve yanlılık değerlerinin çoğunda önemli bir farklılık gözlemlenmemiştir. Sonlandırma koşullarından 0,385 ve 0,315 SE(theta)da  $b=0$  başlangıç kuralında hem RMSE hem de yanlılık değerleri daha düşük elde edilmiştir. Test uzunlukları açısından ise önemli bir farklılık gözlenmemiştir.

Arařtırmada, bařlama kuralının  $b=0$  veya  $-1 < b < 1$  seřilmesinin ölçme kesinlięi ve test uzunluęu aęısından genellikle çok önemli bir farklılıęa neden olmadığı belirlenmiştir. RMSE ve yanlılık deęerlerinin dięer kořullara göre yüksek elde edildięi 0,385 ve 0,315 SE(theta) sonlandırma kořulunda farklı olarak,  $b=0$  bařlangıç yönteminin seřilmesi ile daha düşük RMSE ve yanlılık deęerleri elde edilmiştir.

### 5.1.3. Yetenek Kestirim Yöntemi Olarak MLE veya EAP Belirlenmesi

- **Madde havuzu büyüklüęü 250 ve bařlama kuralı  $b=0$**  olarak seřildięinde; RMSE deęerleri, 0,220 SE(theta) (tek başına ve en az 10 madde ile kullanımı) dıřındaki tüm kořullar için EAP yetenek kestirim yönteminin kullanıldıęı durumda daha düşük elde edilmiştir. Yanlılık deęeri ise tüm kořullarda EAP yetenek kestirim yönteminin kullanıldıęı durumda daha düşük elde edilmiştir. Test uzunlukları aęısından da çoęu sonlandırma kořullarında çok önemli farklılıklar gözlemlenmemiştir. 0,02 theta yakınsama sonlandırma kořulunun kullanımında EAP kestirim yönteminde özellikle RMSE deęeri, MLE yöntemine kıyasla oldukça düşüktür.
- **Madde havuzu büyüklüęü 250 ve bařlama kuralı  $-1 < b < 1$**  olarak seřildięinde; EAP kestirim yöntemi ile elde edilen RMSE ve yanlılık deęerleri daha düşük elde edilmiştir. Test uzunlukları aęısından da çok farklılık görülmemiştir. EAP ve MLE yöntemi ile elde edilen RMSE deęerleri arasındaki fark en yüksek deęerini 0,02 theta yakınsama sonlandırma kořulunda almıştır.
- **Madde havuzu büyüklüęü 500 ve bařlama kuralı  $b=0$**  olarak seřildięinde; EAP kestirim yöntemi ile elde edilen RMSE ve yanlılık deęerleri daha düşük elde edilmiştir. Test uzunlukları aęısından da çok farklılık görülmemiştir. Özellikle 0,02 theta yakınsama sonlandırma kořulunda EAP kestirim yöntemi kullanımında RMSE deęeri, MLE yöntemine kıyasla oldukça düşüktür.
- **Madde havuzu büyüklüęü 500 ve bařlama kuralı  $-1 < b < 1$**  olarak seřildięinde; EAP kestirim yöntemi ile elde edilen RMSE ve yanlılık

değerleri sonlandırma koşullarının çoğunluğunda daha düşük elde edilmiştir. Test uzunlukları açısından da çok farklılık gözlemlenmemiştir. EAP ve MLE yöntemi ile elde edilen RMSE değerleri arasındaki fark en yüksek değerini 0,02 theta yakınsama sonlandırma koşulunda almıştır.

Araştırmada EAP kestirim yönteminde MLE'ye kıyasla daha düşük RMSE ve yanlılık değerleri elde edilmiştir. Literatürde de MLE ve Bayes yöntemlerinin RMSE ve yanlılık değerlerinin karşılaştırılması ile ilgili, bu araştırmanın sonucunu destekleyecek çalışmalarda söz konusudur. Wang ve Vispoel (1998) BBT'de simülasyon çalışması yaparak MLE ve Bayes yöntemlerini (Oven's method, EAP ve MAP) karşılaştırmışlardır. MLE yöntemi ile daha düşük yanlılık, daha yüksek RMSE değerleri elde edilmiştir. İşeri (2002) çalışmasında Bayes yöntemlerinin ölçme kesinliği açısından daha iyi kestirim sağladığını belirtmiştir.

## 5.2. Öneriler

Araştırmanın sonuçlarından yola çıkarak aşağıdaki öneriler sunulabilir.

1. Alt problem 1 için geçerli olan koşullar söz konusu olduğunda; hem RMSE ve yanlılık hem de testin kullanılabilirliği bakımından yeteneğin tahmininde 0,220 SE(theta) ve 0,220 SE(theta)-en az 10 madde sonlandırma koşulları önerilebilir.
2. Alt problem 2 için geçerli olan koşullar söz konusu ise; hem RMSE ve yanlılık hem de testin kullanılabilirliği bakımından 0,220 SE(theta) ve 0,05 theta yakınsama-en az 10 madde sonlandırma koşullarının kullanılması önerilebilir
3. Alt problem 3 geçerli olan koşullar incelendiğinde; hem RMSE ve yanlılık hem de testin kullanılabilirliği bakımından 0,220 SE(theta) ve 0,05 theta yakınsama-en az 10 madde sonlandırma koşullarının kullanılması önerilebilir.
4. Alt problem 4 geçerli olan koşullar incelendiğinde; hem RMSE ve yanlılık hem de testin kullanılabilirliği bakımından 0,220 SE(theta) ve 0,05 theta

yakınsama-en az 10 madde sonlandırma koşullarının kullanılması önerilebilir.

5. Alt problem 5 geçerli olan koşullar incelendiğinde; hem RMSE ve yanlılık hem de testin kullanılabilirliği bakımından 0,220 SE(theta) ve 0,05 theta yakınsama-en az 10 madde sonlandırma koşullarının kullanılması önerilebilir.
6. Alt problem 6 geçerli olan koşullar incelendiğinde; hem RMSE ve yanlılık hem de testin kullanılabilirliği bakımından 0,220 SE(theta) sonlandırma koşulunun kullanılması önerilebilir.
7. Alt problem 7 geçerli olan koşullar incelendiğinde; hem RMSE ve yanlılık hem de testin kullanılabilirliği bakımından 0,220 SE(theta) sonlandırma koşulunun kullanılması önerilebilir.
8. Alt problem 8 geçerli olan koşullar incelendiğinde; hem RMSE ve yanlılık hem de testin kullanılabilirliği bakımından 0,220 SE(theta) sonlandırma koşulunun kullanılması önerilebilir.
9. RMSE ve yanlılık değerlerinin düşük elde edilmesi, yani yüksek ölçme kesinliği için büyük madde havuzları tercih edilebilir.
10. Hem ölçme kesinliği hem de testin ekonomik olması açısından değişen uzunluk sonlandırma kurallarından 0,220 SE(theta) büyük madde havuzlarında tercih edilebilir.
11. 0,385 SE(theta) ve 0,315 SE(theta) sonlandırma koşulları kullanıldığında daha yüksek ölçme kesinliği için başlama değerinin  $b=0$  seçilmesi tercih edilebilir.
12. Yetenek kestirim yöntemi olarak EAP kullanılması MLE ye kıyasla daha kullanışlıdır.
13. Özellikle theta yakınsama kuralında yetenek kestirim yönteminin EAP seçilmesi daha düşük RMSE ve yanlılık açısından tercih edilebilir.

Benzer çalışmalar için aşağıdaki öneriler sunulabilir.



1. Sonlandırma koşulları çok boyutlu testler için karşılaştırılabilir.
2. BBT uygulamalarında önemli bir yere sahip olan madde kullanım sıklığı ve içerik dengesinin incelenmesi bu araştırma kapsamı dışında tutulmuştur. Gelecek araştırmalarda bu iki durum sonlandırma koşulları ile birlikte çalışılabilir.
3. Minimum bilgi sonlandırma kurallarından elde edilen değerlerde diğer kurallardan elde edilenlerle karşılaştırılabilir.
4. Theta yakınsama kuralının standart hata sonlandırma kuralı ile birlikte kullanılmasının sonuçlara etkisi incelenebilir.
5. Madde havuzunun özellikleri değiştirilerek sonlandırma kuralları birbiri ile karşılaştırılabilir.
6. Diğer yetenek kestirim yöntemleri kullanılmasının sonlandırma kurallarından elde edilen sonuçlara etkisi incelenebilir.
7. Araştırmada madde seçme yöntemlerinden Maksimum Fisher Information yöntemi kullanılmıştır. Diğer madde seçme yöntemleri de benzer koşullarda karşılaştırılabilir.

## KAYNAKÇA

- Babcock, B. & Weiss, D.J. (2012). Termination criteria in Computerized Adaptive Tests: do variable-length CAT's provide efficient and effective measurement? *International Association for Computerized Adaptive Testing*, 1, 1-18.
- Blais, J. & Raiche, G. (2002). *Features of the sampling distribution of the ability estimate in Computerized Adaptive Testing according to two stopping rules*, International Objective Measurement Workshop, New Orleans, April 2002.
- Choi, S. W., Grady, M.W. & Dodd, B.G. (2011). A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurement*, 71, 37-53.
- Çıkrıkçı, Demirtaşlı, N. (1999). Psikometride yeni ufuklar: bilgisayar ortamında bireye uyarlanmış test. *Türk Psikoloji Bülteni*. 5(13), 31-36.
- Dodd, B.G., Koch, W.R. & de Ayala, R.J. (1993). Computerized Adaptive Testing using the partial credit model effects of item pool characteristics and different stopping rules. *Educational and Psychological Measurement*, 53, 61-77.
- Embretson, E. & Reise, S. P. (2000). *Item response theory for psychologist principles and application*. London: Lawrence Erlbaum Assc.
- Eggen,T. (2004). *Contributions to the theory and practice of Computerized Adaptive Testing*. Druk: Print Partners Ipskamp B.V., Enschede.
- Evans, J. J. (2010). *Comparability of examinee proficiency scores on Computer Adaptive Tests using real and simulated data*. Doctoral Dissertation. The State University of New Jersey.
- Flaugher, R. (2000). Item pools. In H. Wainer (Eds.), *Computerized Adaptive Testing*, 37-59. London: Lawrence Erlbaum Assc.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory. principles and application*. Boston: Kluwer-Nijhoff.
- Hambleton, R. K., Swaminathan, H. & Rogers,H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications.
- Han, K. C. (2011). *User's Manual: SimulCAT*. Graduate Management Admission Council.
- Ivei, J. L. (2007). *Test taking strategies in Computer Adaptive Testing That will improve your score: fact or fiction?* Doctoral Dissertation. University of Michigan.

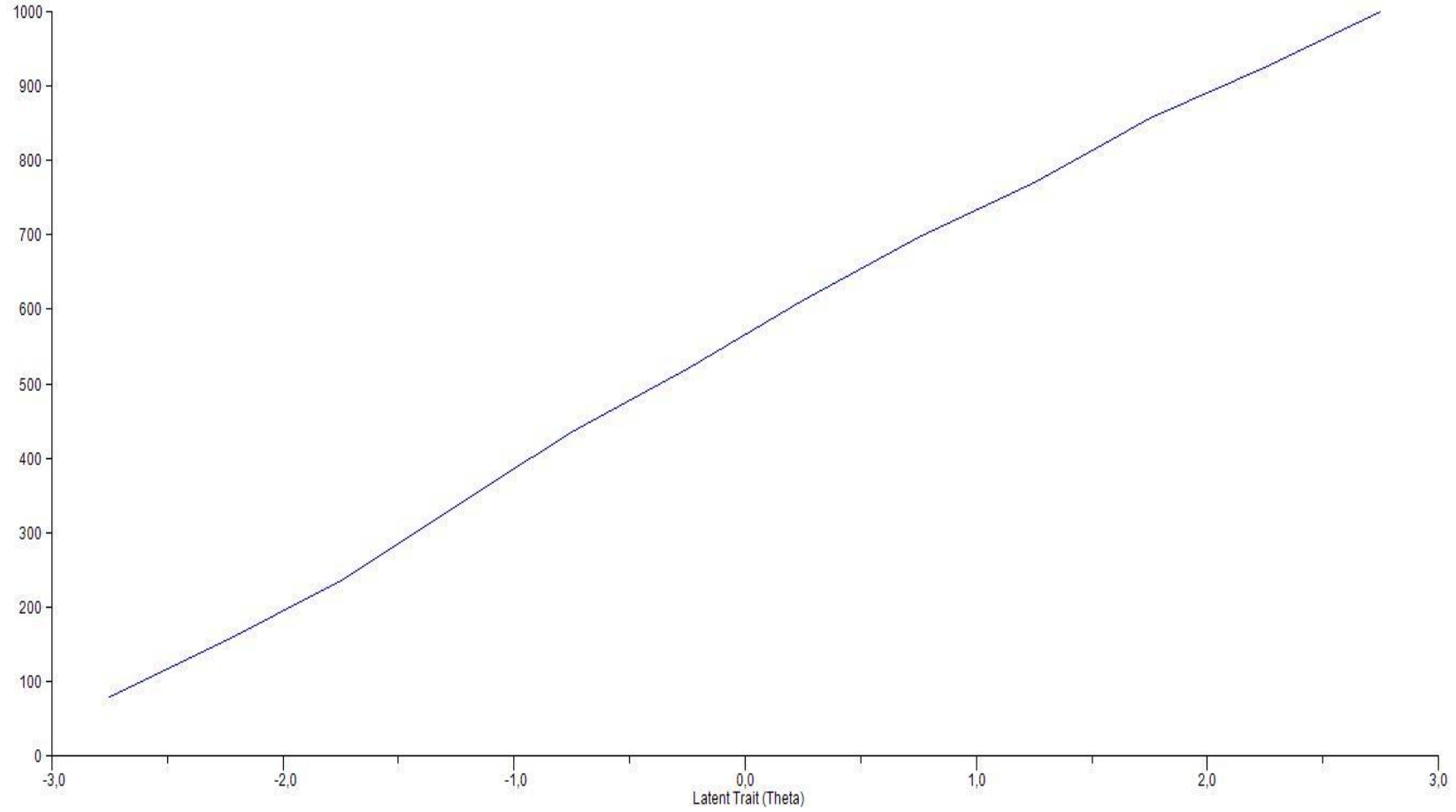
- İseri, A. I. (2002). *Assessment of students' mathematics achievement through Computer Adaptive Testing procedures*. Middle East Technical University.
- Kalender, İ. (2004). *Bilgisayar ortamında Bireyselleştirilmiş Testlerin eğitimde kullanımı*. XIII. Ulusal Eğitim Bilimleri Kurultayı. İnönü Üniversitesi, Eğitim Fakültesi, Malatya. 6-9 Temmuz 2004.
- Kalender, İ. (2011). *Effects of different Computerized Adaptive Testing strategies on recovery of ability*. Doctoral Dissertation. Middle East Technical University.
- Kaptan, F. (1993). *Yetenek kestiriminde Adaptive (Bireyselleştirilmiş) Test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması*. Doktora Tezi. Hacettepe Üniversitesi.
- Karasar, N. (2004). *Bilimsel araştırma yöntemi: kavramlar, ilkeler, teknikler* (13. Baskı). Ankara: Nobel Yayın Dağıtım.
- Köklü, N. (1990). *Klasik test teorisine göre geliştirilen tailored test ile grup testi arasında bir karşılaştırma*. Doktora Tezi. Hacettepe Üniversitesi.
- Linden, W. J & Glas, G. A. W. (2002). *Computerized Adaptive Testing: theory and practice*. USA: Kluwer Academic Publishers.
- Linecra, J. M. (2000). *Computer-Adaptive Testing: a methodology whose time has come*, [Çevrim-içi: <http://www.rasch.org/memo69.htm>], erişim tarihi: 02 Mayıs 2012.
- McBride, J. R. & Martin, J. T. (1983). Reliability and validity of adaptive ability tests in a military setting. In D. J. Weiss (Eds.), *New horizons in testing: Latent trait theory and computerized adaptive testing*, 223–226). New York: Academic Press.
- McBride, J. R., Wetzel, C.D. & Hetter, R. D. (2001). Preliminary psychometric research for CAT-ASVAB: selecting an adaptive testing strategy. In W. Sands, B. K. Waters, and J.R.McBride (Eds.). *Computerized Adaptive Testing: from inquiry to operation*, 83–95. Washington, DC: American Psychological Association.
- Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: a meta-analysis. *Psychological Bulletin*, 114, 449-458.
- Meijer, R. R. & Nerring, M. L. (2001). New development in the area of Computerized Testing. *Psychologie Francaise*, 46(3), 221-230.
- Mills, C. N. & Stocking, M.L. (1996). Practical issues in large-scale Computerized Adaptive Testing. *Applied Measurement in Education*, 9(4), 287-304.
- Pearson Assessment. (2012). [Çevrim-içi: [www.pearsonassessment.com](http://www.pearsonassessment.com)] Erişim tarihi: 02.05.12.

- Riley, B. B., Conrad, K. J., Bezruczko, N. & Dennis, M. (2007). Relative precision, efficiency and construct validity of different starting and stopping rules for a Computerized Adaptive Test: the GAIN substance problem scale. *Journal of Applied Measurement*, 8(1).
- Rudner, L. M. (1998). *Interactive Computer Adaptive Testing*, [Çevrim-içi: <http://EdRes.org/scripts/cat>], Erişim Tarihi: 02.05.12.
- Samejima, F. (1977). A method of estimating item characteristic functions using the maximum likelihood estimate of ability. *Psychometrika*, 42(2), 163-191.
- Segall, D. O. (2004). Computerized Adaptive Testing. In Kempf-Leanard (Eds.). *The Encyclopedia of Social Measurement*, 429-438. San Diego, CA: Academic Press.
- Simms, L. J. & Clark, L. A. (2005). Validation of a computerized adaptive version of the schedule for non-adaptive and adaptive personality (SNAP). *Psychological Assessment*, 17, 28-43.
- Stocking, M. L. (1987). Two Simulated feasibility studies in Computerized Adaptive Testing. *Applied Psychology: An International Review*, 36, 263-267.
- Thissen, D. & Mislevy, R. J. (2000). Testing algorithms. In H. Wainer (Eds.). *Computerized Adaptive Testing*, 101-135. London: Lawrence Erlbaum Assc.
- Wainer, H. (2000). *Computerized Adaptive Testing*. London: Lawrence Erlbaum Assc.
- Wainer, H. & Mislevy, R. J. (2000). Item response theory, item calibration and proficiency estimation. In H. Wainer (Eds.). *Computerized Adaptive Testing*. 65-102. London: Lawrence Erlbaum Assc.
- Wang, T., Hanson, B. A. & Lau, C. (1999). Reducing bias in CAT ability estimation: a comparison of approaches. *Applied Psychological Measurement*, 23, 263-278.
- Wang, T. & Vispoel, W. P. (1998). Properties of ability estimation methods Computerized Adaptive Testing. *Journal of Educational Measurement*, 35 (2), 109-135.
- Wang, S. & Wang, T. (2001). Precision of warm's weighted likelihood estimates for a polytomous model in Computerized Adaptive Testing. *Applied Psychological Measurement*, 25(4), 317-331.
- Weiss, D. J. (1982). Improving measurement quality and efficiency with Adaptive Testing. *Applied Psychological Measurement*, 6,473-492.
- Weiss, D. J. (1983). *New horizons in testing*. New York: Academic Press.
- 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.

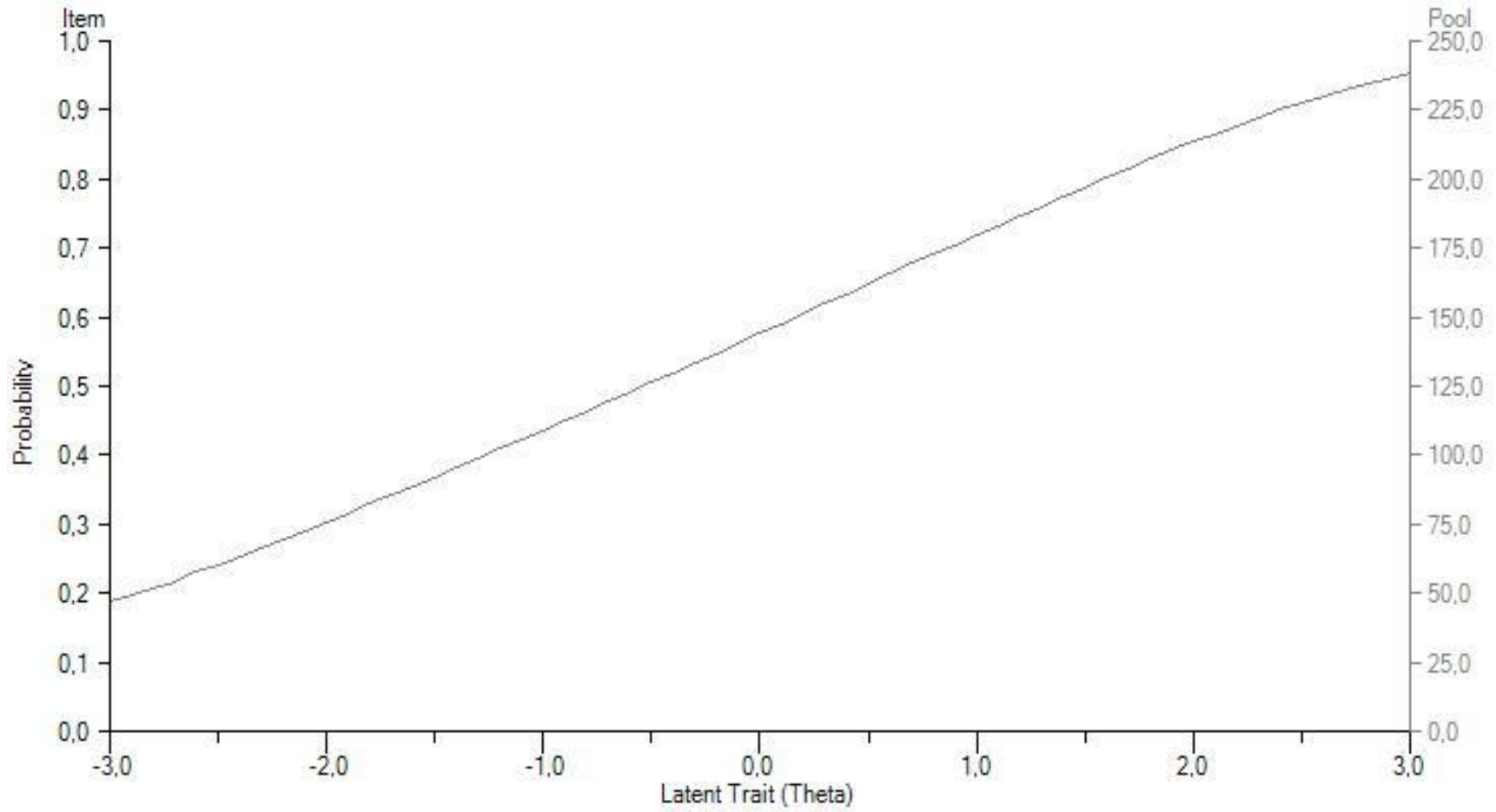
- Weiss, D. J. & Kingsbury, G. G. (1984). Application of Computerized Testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375.
- Yi, Q., Wang, T. & Ban, J. C. (2001). Effects of scale transformation and test termination rule on the precision of ability estimation in Computerized Adaptive Testing. *Journal of Educational Measurement*, 38, 267-292.
- Yoo, H. (2011). *Evaluating several Multidimensional Adaptive Testing procedures for diagnostic assessment*. Doctoral Dissertation. University of Massachusetts.

## **EKLER DİZİNİ**

**EK-1: Bireylerin Yetenek Parametrelerine Göre Dağılım Grafiği**

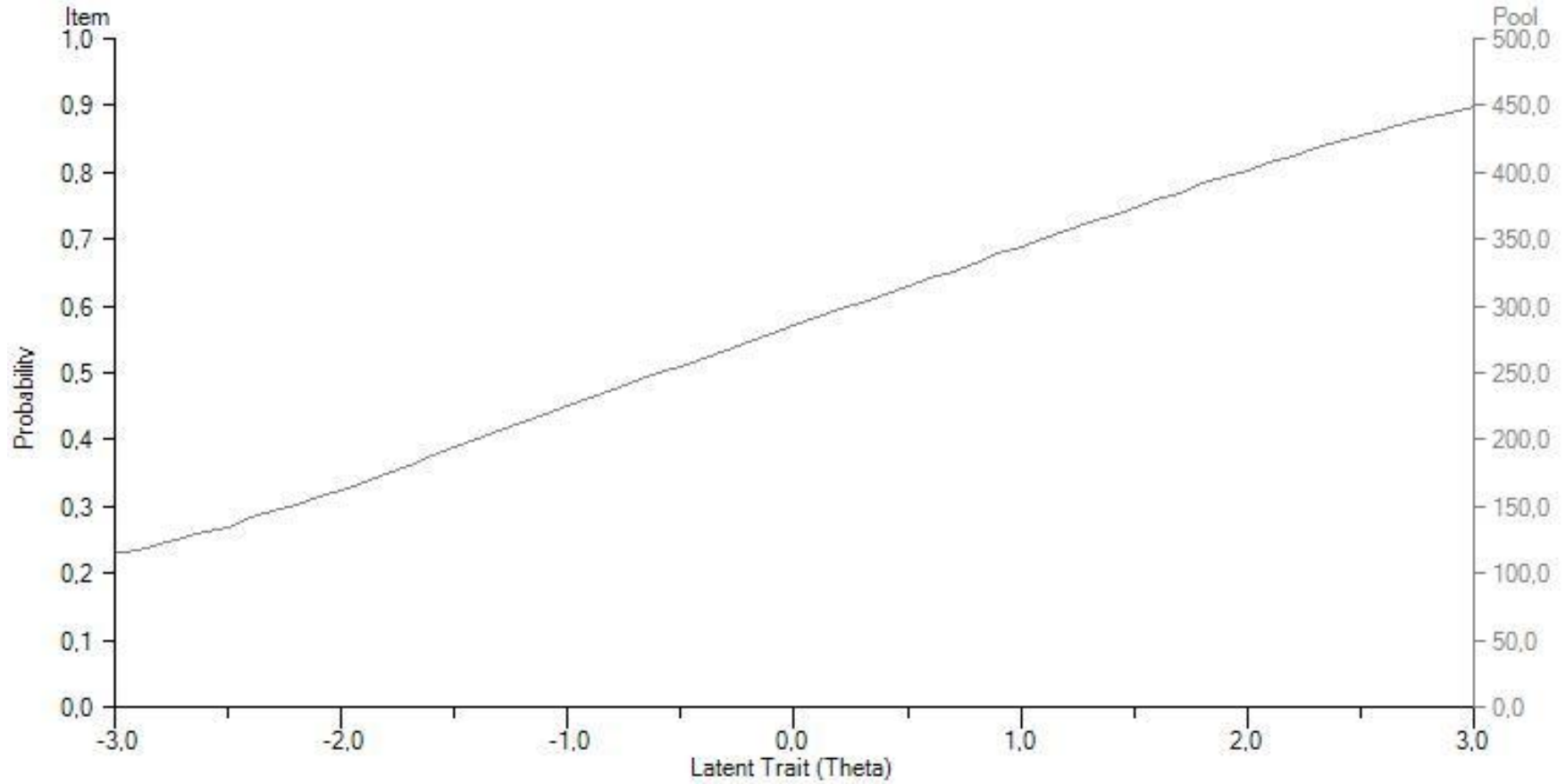


**EK-2: 250 Maddeden Oluşan Havuz Cevap Fonksiyonu**

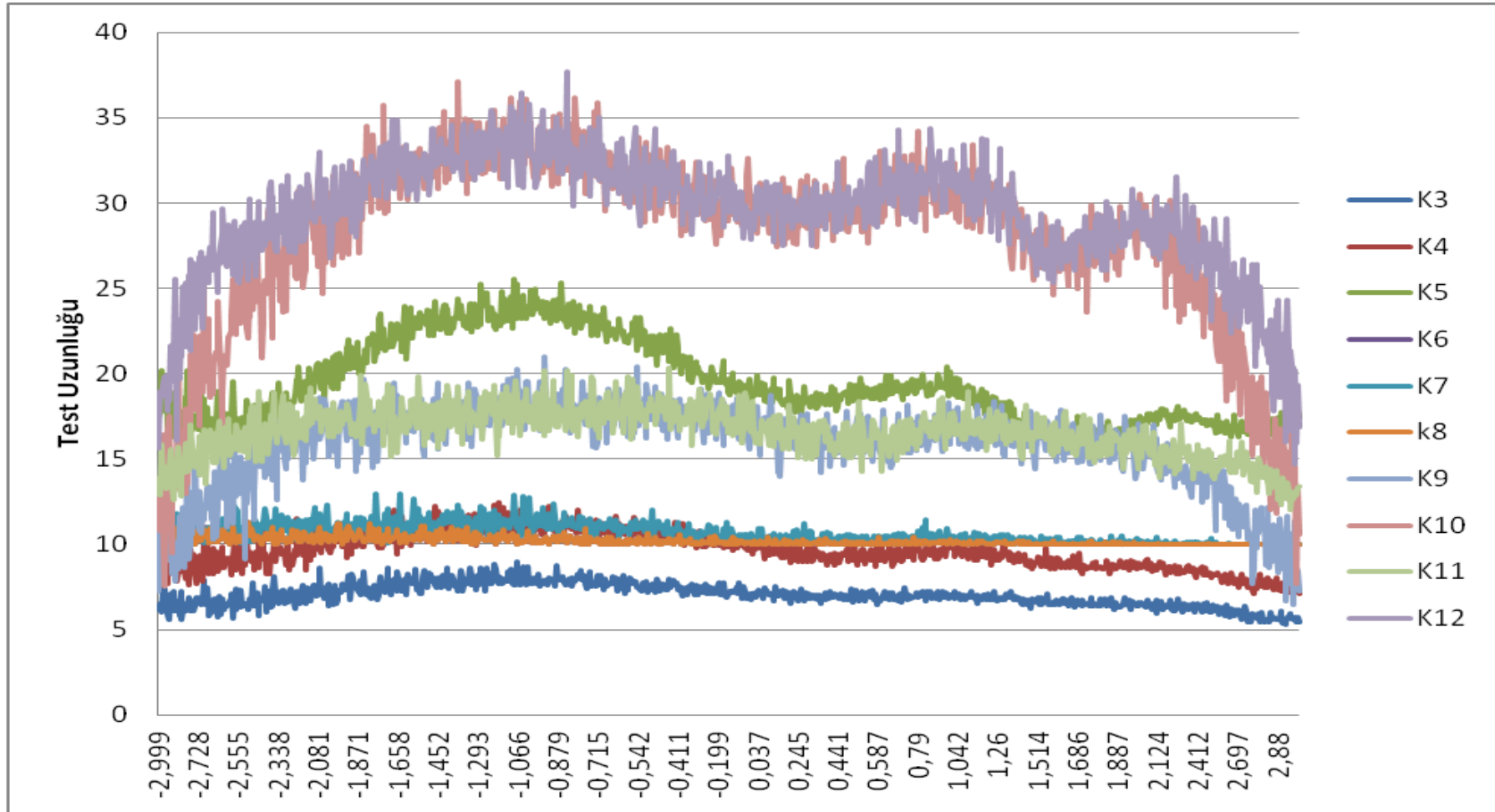




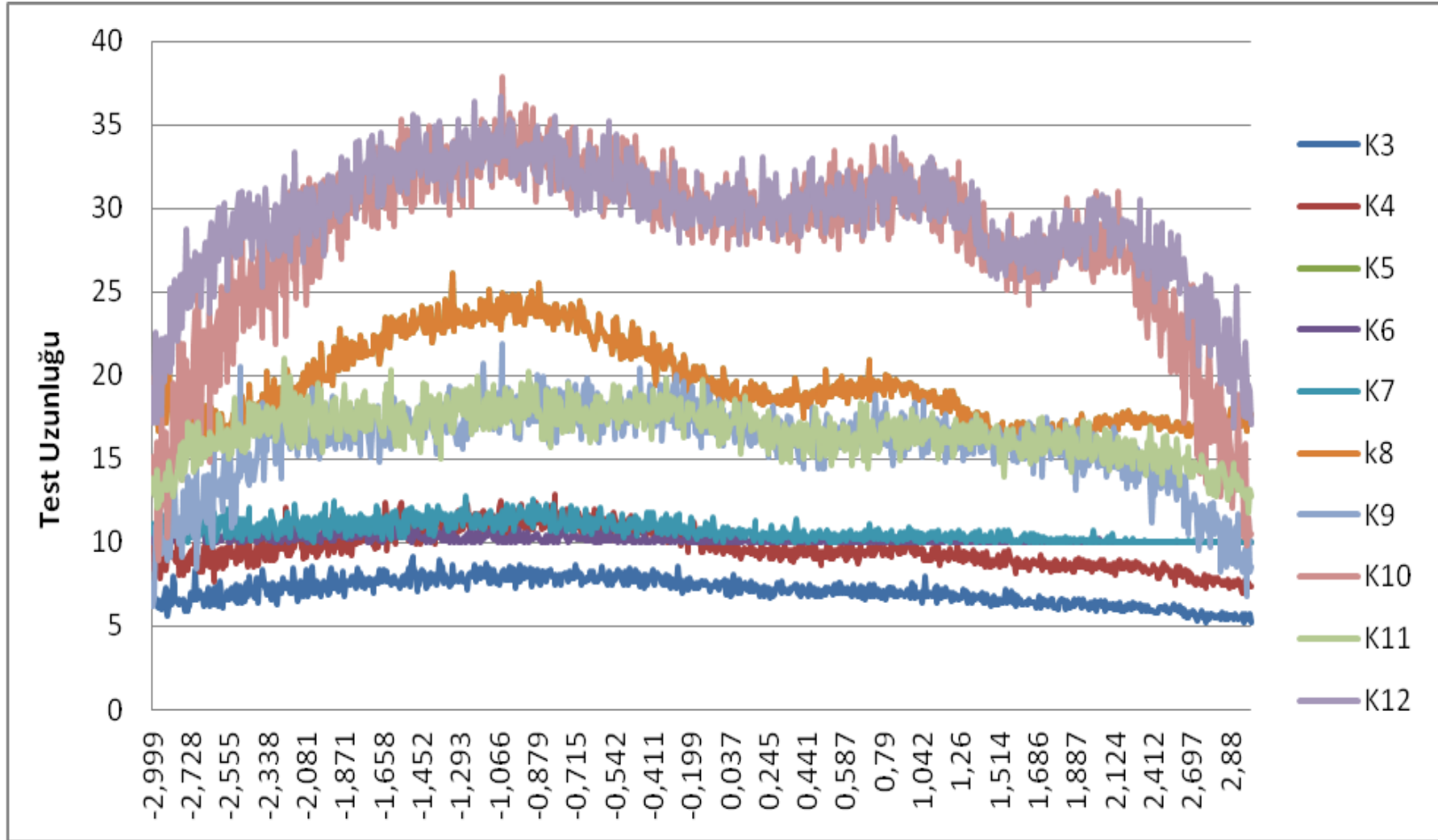
**EK-3: 500 Maddeden Oluşan Havuz Cevap Fonksiyonu**



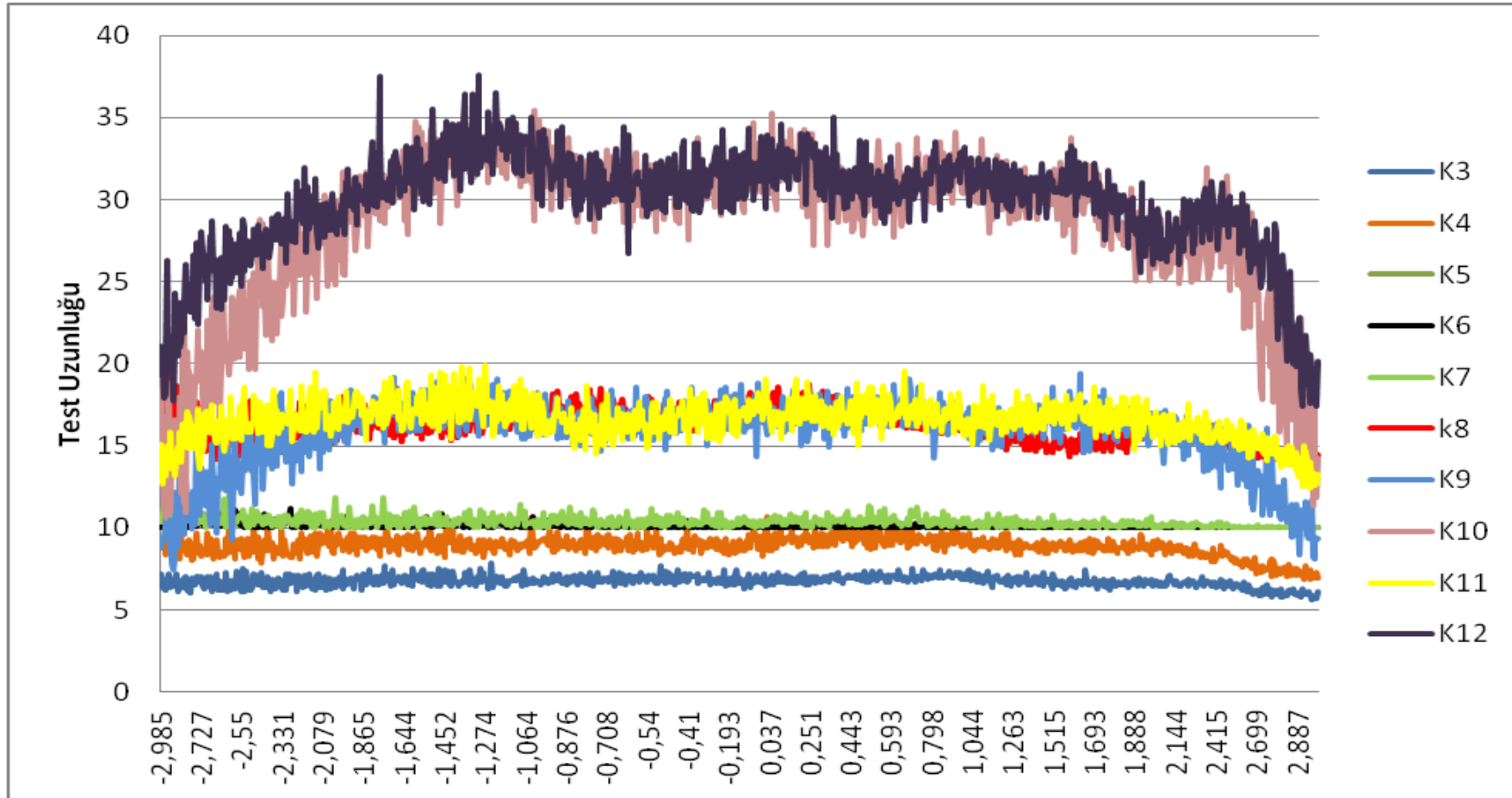
**EK-4: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 1**



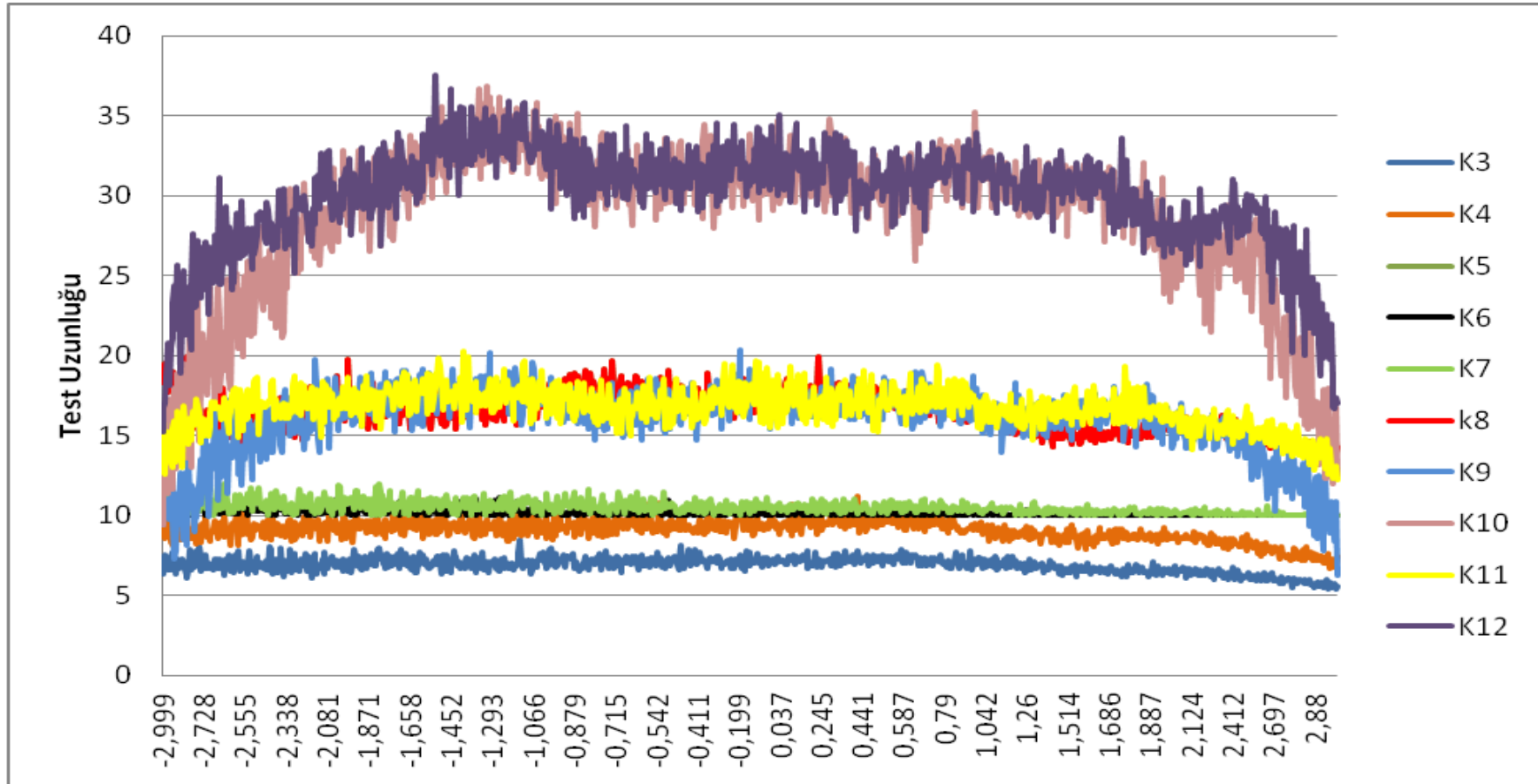
**EK-5: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 2**



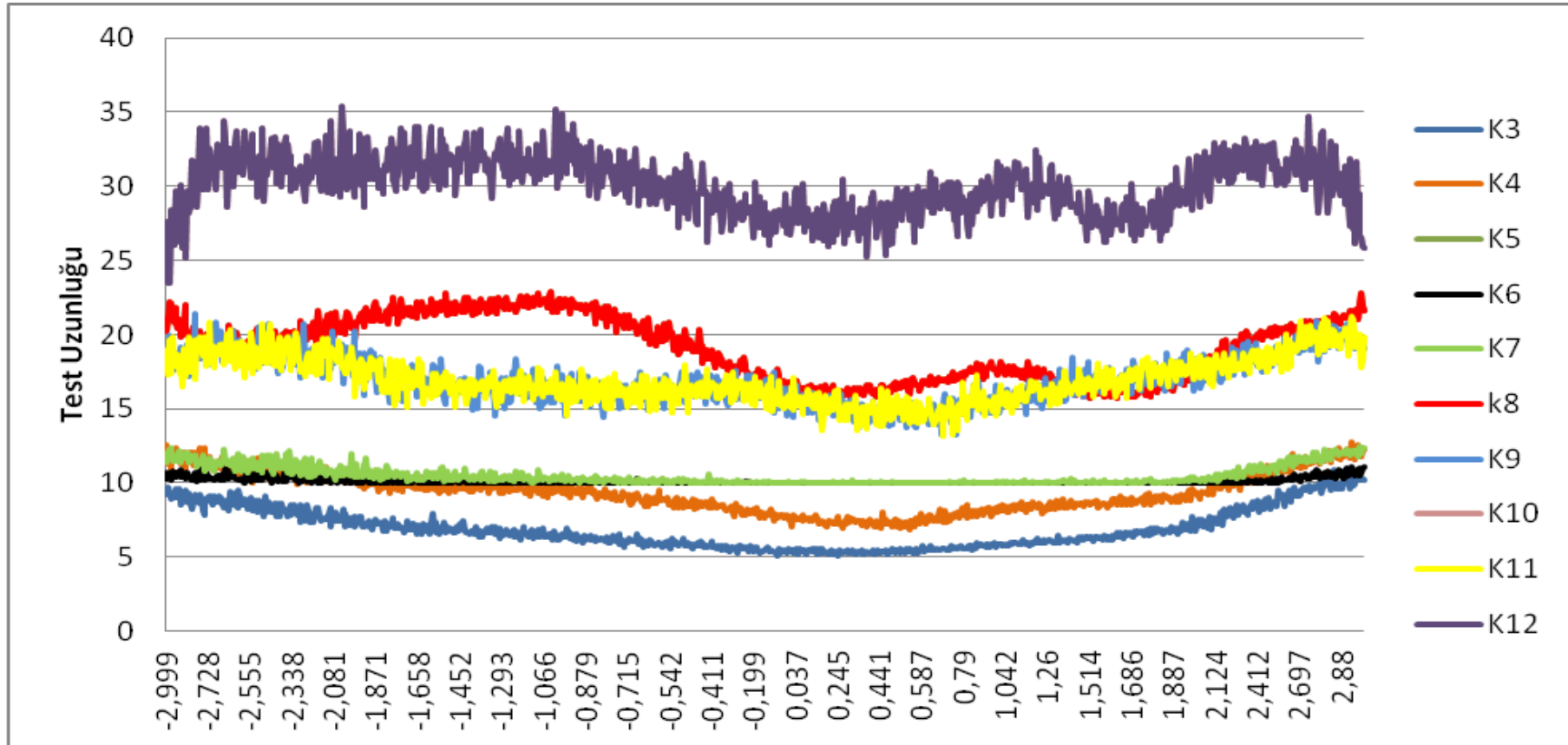
**EK-6: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 3**



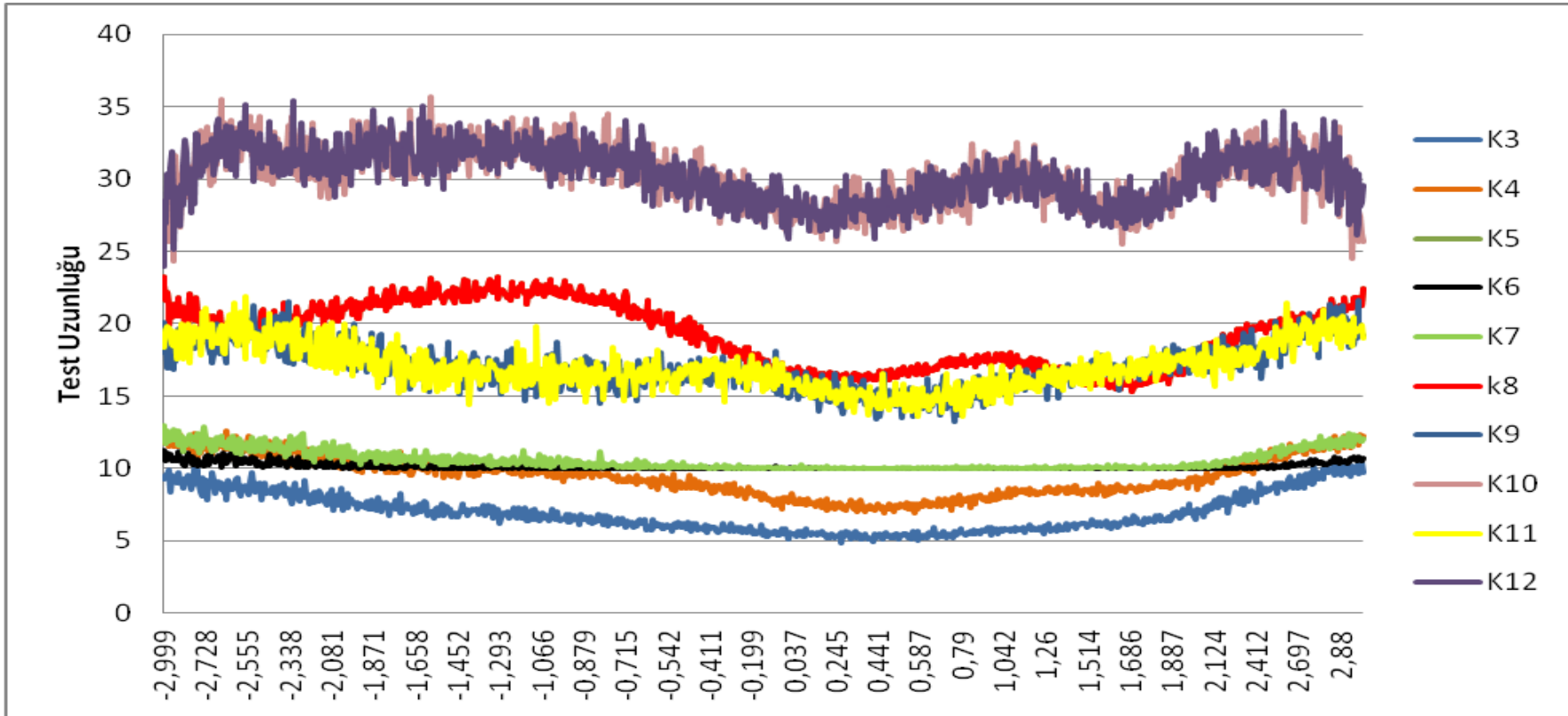
**EK-7: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 4**



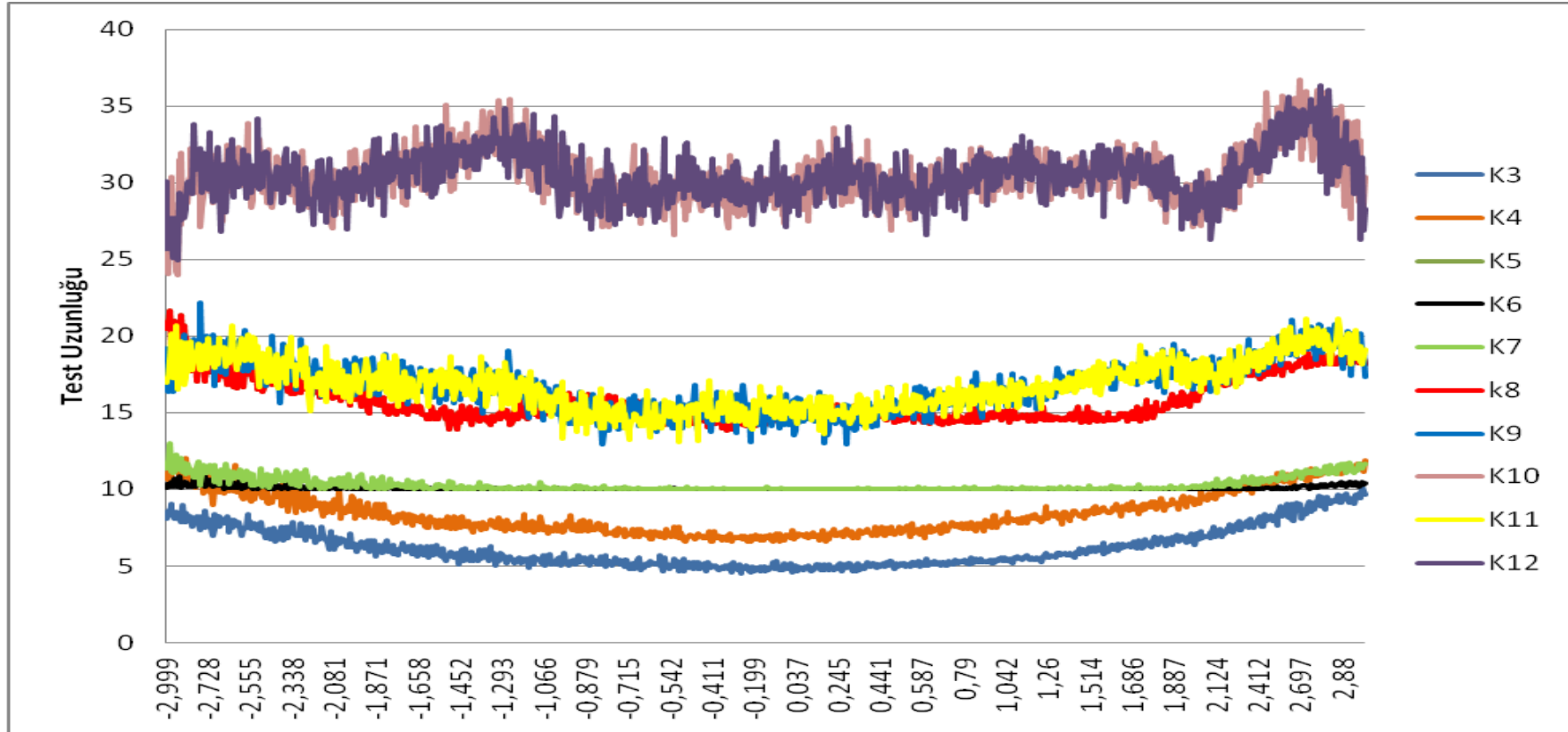
**EK-8: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 5**



**EK-9: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 6**

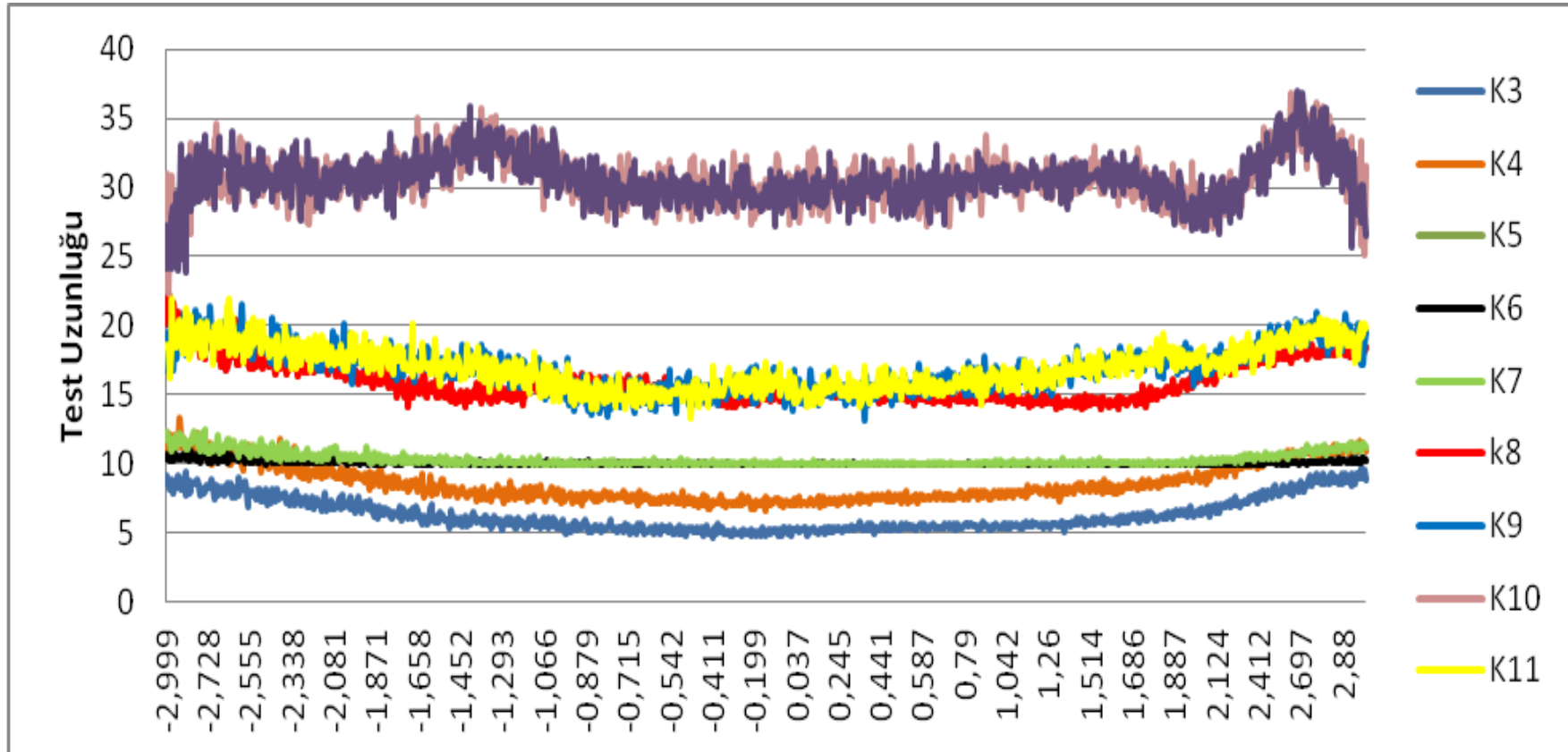


**EK-10: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 7**





**EK-11: Yetenek Düzeylerinde Alınan Test Uzunluğu Dağılımı-Alt Problem 8**



**EK-12: Her Bir Koşula Ait Elde Edilen Ortalama RMSE değerleri**

Sonlandırma Kuralları	Koşullar	250madde b=0	250madde -1< b<1	500madde b=0	500madde -1< b<1	250madde b=0	250madde -1< b<1	500madde b=0	500madde -1< b<1
		MLE	MLE	MLE	MLE	EAP	EAP	EAP	EAP
Sabit Uzunluk	15	0,302	0,297	0,273	0,272	0,269	0,270	0,247	0,248
	20	0,237	0,232	0,217	0,217	0,227	0,231	0,208	0,209
SE(THETA)	0,385	0,546	0,554	0,516	0,567	0,463	0,478	0,466	0,477
	0,315	0,386	0,403	0,382	0,403	0,363	0,362	0,363	0,370
	0,220	0,222	0,223	0,227	0,229	0,229	0,230	0,235	0,234
SE(THETA)- en az 10 madde	0,385-en az 10 madde	0,397	0,406	0,359	0,391	0,348	0,350	0,324	0,330
	0,315-en az 10 madde	0,356	0,359	0,341	0,362	0,337	0,336	0,313	0,319
	0,220- en az 10 madde	0,222	0,223	0,227	0,229	0,229	0,230	0,235	0,234
Theta Değeri Farkı	0,05	0,305	0,298	0,293	0,301	0,250	0,246	0,228	0,226
	0,02	0,266	0,265	0,262	0,268	0,189	0,190	0,168	0,168
Theta Değeri Farkı-en az 10 madde	0,05-en az 10 madde	0,246	0,250	0,228	0,231	0,245	0,245	0,227	0,227
	0,02-en az 10 madde	0,190	0,189	0,170	0,169	0,189	0,190	0,168	0,168

**EK-13: Her Bir Koşula Ait Elde Edilen Ortalama BIAS değerleri**

Sonlandırma Kuralları	Koşullar	250madde b=0	250madde -1< b<1	500madde b=0	500madde -1< b<1	250madde b=0	250madde -1< b<1	500madde b=0	500madde -1< b<1
		MLE	MLE	MLE	MLE	EAP	EAP	EAP	EAP
Sabit Uzunluk	15	0,026	0,026	0,021	0,021	0,010	0,009	0,004	0,005
	20	0,015	0,014	0,013	0,013	0,007	0,007	0,002	0,004
SE(THETA)	0,385	0,097	0,100	0,079	0,098	0,016	0,022	0,014	0,017
	0,315	0,044	0,050	0,042	0,050	0,009	0,005	0,009	0,012
	0,220	0,010	0,011	0,012	0,013	0,002	0,002	0,002	0,002
SE(THETA)- en az 10 madde	0,385-en az 10 madde	0,052	0,054	0,041	0,048	0,013	0,013	0,004	0,006
	0,315-en az 10 madde	0,040	0,038	0,032	0,041	0,011	0,011	0,005	0,006
	0,220- en az 10 madde	0,010	0,011	0,012	0,013	0,002	0,002	0,002	0,002
Theta Değeri Farkı	0,05	0,009	0,010	0,010	0,014	0,006	0,004	0,003	0,002
	0,02	0,006	0,005	0,007	0,009	0,003	0,001	0,001	0,001
Theta Değeri Farkı-en az 10 madde	0,05-en az 10 madde	0,012	0,013	0,009	0,009	0,007	0,007	0,003	0,003
	0,02-en az 10 madde	0,002	0,005	0,005	0,002	0,003	0,003	0,000	0,001

**EK-14:Her Bir Koşula Ait Ortalama Uyum Katsayıları**

Sonlandırma Kuralları	Koşullar	250madde b=0	250madde -1< b<1	500madde b=0	500madde -1< b<1	250madde b=0	250madde -1< b<1	500madde b=0	500madde -1< b<1
		MLE	MLE	MLE	MLE	EAP	EAP	EAP	EAP
Sabit Uzunluk	15	0,984	0,984	0,987	0,987	0,988	0,988	0,990	0,990
	20	0,990	0,990	0,992	0,992	0,991	0,991	0,993	0,992
SE(THETA)	0,385	0,949	0,947	0,954	0,945	0,965	0,962	0,964	0,963
	0,315	0,974	0,972	0,974	0,971	0,978	0,978	0,978	0,977
	0,220	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991
SE(THETA)- en az 10 madde	0,385-en az 10 madde	0,972	0,971	0,977	0,973	0,980	0,980	0,983	0,982
	0,315-en az 10 madde	0,978	0,977	0,979	0,977	0,981	0,981	0,984	0,983
	0,220- en az 10 madde	0,991	0,991	0,991	0,991	0,991	0,991	0,991	0,991
Theta Değeri Farkı	0,05	0,984	0,985	0,985	0,984	0,989	0,990	0,991	0,991
	0,02	0,988	0,988	0,988	0,988	0,994	0,994	0,995	0,995
Theta Değeri Farkı-en az 10 madde	0,05-en az 10 madde	0,989	0,989	0,991	0,991	0,990	0,990	0,991	0,991
	0,02-en az 10 madde	0,994	0,994	0,995	0,995	0,994	0,994	0,995	0,995

**EK-15:Her Bir Koşula Ait Elde Edilen Ortalama Test Uzunlukları**

Sonlandırma Kuralları	Koşullar	250madde b=0	250madde -1< b<1	500madde b=0	500madde -1< b<1	250madde b=0	250madde -1< b<1	500madde b=0	500madde -1< b<1
		MLE	MLE	MLE	MLE	EAP	EAP	EAP	EAP
Sabit Uzunluk	15	15,0	15,0	15,0	15,0	15,0	15,0	15,0	15,0
	20	20,0	20,0	20,0	20,0	20,0	20,0	20,0	20,0
SE(THETA)	0,385	7,0	7,1	6,8	6,9	6,9	6,9	6,2	6,3
	0,315	9,6	9,7	8,9	9,1	9,3	9,4	8,4	8,6
	0,220	19,4	19,5	16,3	16,5	19,1	19,2	15,8	15,9
SE(THETA)- en az 10 madde	0,385-en az 10 madde	10,2	10,2	10,1	10,2	10,1	10,1	10,1	10,1
	0,315-en az 10 madde	10,6	10,7	10,4	10,5	10,5	10,6	10,3	10,3
	0,220- en az 10 madde	19,4	19,5	16,3	16,5	19,1	19,2	15,8	15,9
Theta Değeri Farkı	0,05	15,9	15,9	15,8	16,0	16,9	17,0	16,7	16,9
	0,02	28,2	28,2	28,6	28,8	30,0	30,3	30,5	30,6
Theta Değeri Farkı-en az 10 madde	0,05-en az 10 madde	16,6	16,6	16,6	16,7	16,9	17,0	16,8	16,9
	0,02-en az 10 madde	29,4	29,5	29,9	30,0	30,0	30,2	30,5	30,6

## ÖZGEÇMİŞ

<b>Adı Soyadı</b>	Melek Gülşah EROĞLU
<b>e-posta</b>	mgulsah@gazi.edu.tr

### Eğitim ve Akademik Durumu

<b>Lisans</b>	Hacettepe Üniversitesi Eğitim Fakültesi OFMA Eğitimi Bölümü Fizik Öğretmenliği A.B.D.
<b>Yüksek Lisans</b>	Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme A.B.D.
<b>Yabancı Dil</b>	İngilizce, Almanca
<b>İş Deneyimi</b>	Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme A.B.D.