



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Eğitim Bilimleri Anabilim Dalı

Eğitimde Ölçme ve Değerlendirme Bilim Dalı

**BİREYSELLEŞTİRİLMİŞ BİLGİSAYARLI TEST
UYGULAMALARINDA KULLANILAN MADDE SEÇME
YÖNTEMLERİNİN KARŞILAŞTIRILMASI**

Sema SULAK

Doktora Tezi

Ankara, 2013

BİREYSELLEŐTİRİLMİŐ BİLGİSAYARLI TEST
UYGULAMALARINDA KULLANILAN MADDE SEÇME
YÖNTEMLERİNİN KARŐILAŐTIRILMASI

Sema SULAK

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü
Eğitim Bilimleri Anabilim Dalı
Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Doktora Tezi

Ankara, 2013

KABUL VE ONAY

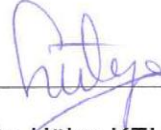
Sema SULAK tarafından hazırlanan "Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Kullanılan Madde Seçme Yöntemlerinin Karşılaştırılması" başlıklı bu çalışma 04.07.2013 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Doktora Tezi olarak kabul edilmiştir.



Prof. Dr. Şener BÜYÜKÖZTÜRK (Başkan)



Prof. Dr. Selahattin GELBAL



Doç. Dr. Hülya KELECİOĞLU (Danışman)



Doç. Dr. Nuri DOĞAN



Yrd. Doç. Dr. Burcu ATAR

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Yusuf ÇELİK

Enstitü Müdürü

BİLDİRİM

Hazırladığım tezin, tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- ☒ Tezimin tamamı her yerden erişime açılabilir.
- ☐ Tezim sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- ☐ Tezimin ... yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin tamamı her yerden erişime açılabilir.

04.07.2013



Sema SULAK

TEŞEKKÜR

Doktora eğitimim süresince kendisinden çok faydalandığım, emek ve zamanını harcayarak tez çalışmamı destekleyen danışmanım Doç. Dr. Hülya Kelecioğlu'na, değerli hocam Prof. Dr. Selahaddin Gelbal'a, görüş ve önerileriyle tez çalışmama katkı sağlayan jüri üyeleri, Prof. Dr. Şener Büyüköztürk'e, Doç. Dr. Nuri Doğan'a ve Yrd. Doç. Dr. Burcu Atar'a, akademik anlamda ufkumu açan sayın Yaşar Baykul'a, yurtdışı yayınlara erişim, akademik çeviri konusunda bana yardımcı olan ve hep yanımda olduğunu bildiğim kardeşim Bilge Sulak'a, varlıklarından güç aldığım sevgili kardeşlerim Esin Sulak, Hasibe Sulak, Mine Sulak ve Murat Sulak'a, manevi desteğini hep hissettiğim canım annem Emine Sulak'a, bir akademisyen olarak rol modelim olan canım babam Hidayet Sulak'a, yetişmemde büyük emekleri olan ve bana hayatı öğreten sayın Mevlüt Ağaoğlu'na ve Hüseyin Avcı'ya

Doktora çalışmamı Yurt İçi Doktora Burs programı ile destekleyen, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK)'na

sonsuz teşekkürlerimi sunarım.

ÖZET

SULAK, Sema, Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Kullanılan Madde Seçme Yöntemlerinin Karşılaştırılması, Doktora Tezi, Ankara, 2013

Bireyselleştirilmiş bilgisayarlı test uygulamalarının en önemli avantajı, cevaplayıcıların yetenek düzeyine uygun maddelerle karşılaşmasıdır. Bu avantajın sağlanması için testi alan kişilerin yeteneği hakkında en çok bilgi taşıyan maddelerin seçilmesi oldukça kritik bir öneme sahiptir. Bireyselleştirilmiş bilgisayarlı testte madde seçme yöntemi, testin nasıl başlayacağı, devam edeceği ve sonlanacağına hakim olan merkezi bir bileşendir.

Bu araştırmada BBT uygulamalarında kullanılan madde seçme yöntemleri, yetenek kestirim yöntemleri ve test durdurma kurallarına göre karşılaştırmalı olarak incelenmiştir. Bu amaçla 250 maddelik bir madde havuzu, ortalaması 0 ve standart sapması 1 olacak şekilde 2000 kişi simülatif olarak oluşturulmuştur. BBT koşulları, madde seçme yöntemleri (Maksimum Fisher Bilgisi, a-tabakalama, Olabilirlik Ağırlıklı Bilgi Ölçütü, Aşamalı Maksimum Bilgi Oranı, Kullbak-Leibler Bilgisi), yetenek kestirim yöntemleri (En Çok Olabilirlik Tahmini, Beklenen Sonsal Dağılım), test durdurma kuralları (40 madde, $SH < 0.2$ ve $SH < 0.4$) olmak üzere toplam otuz koşuldan oluşturulmuştur. Elde edilen bulguların analizinde; sabit test uzunluğuna dayalı durdurma kuralında tahminin standart hatası, sabit standart hataya dayalı durdurma kuralında ise ortalama madde sayısı kullanılmıştır. Madde seçme yöntemlerinin madde kullanım sıklıkları da incelenmiştir.

Sabit test uzunluğuna dayalı durdurma kuralına (40 madde) göre yapılan karşılaştırmalarda, En Çok Olabilirlik Tahmini yetenek kestirimi kullanıldığında elde edilen SH değerleri, Beklenen Sonsal Dağılım yetenek kestirimi

kullanıldığında elde edilen SH değerlerinden daha yüksek bulunmuştur. En Çok Olabilirlik Tahmini yetenek kestirimi koşullarında, test uzunluğu $n < 30$ iken, a-tabakalama; $n > 30$ iken Kullbak-Leibler madde seçme yöntemi en yüksek SH değerini göstermiştir. Beklenen Sonsal Dağılım yetenek kestirimi koşullarında ise bütün test uzunluklarında en yüksek SH değeri a-tabakalama madde seçme yönteminden elde edilmiştir. Madde havuzu kullanımında ise en iyi sonuç a-tabakalama madde seçme yönteminden elde edilmiştir.

Durdurma kuralının $SH < 0.2$ olduğu koşullarda, En Çok Olabilirlik Tahmini yetenek kestirimi kullanıldığında en düşük ve en yüksek madde sayısı ortalaması sırasıyla Aşamalı Maksimum Bilgi Oranı ve Maksimum Fisher Bilgisi madde seçme yönteminden; Beklenen Sonsal Dağılım yetenek kestirimi kullanıldığında ise, Kullbak-Leibler ve Olabilirlik Ağırlıklı Bilgi Ölçütü madde seçme yönteminden elde edilmiştir. Durdurma kuralının $SH < 0.4$ olduğu koşullarda, En Çok Olabilirlik Tahmini yetenek kestirimi kullanıldığında en düşük ve en yüksek madde sayısı ortalaması sırasıyla Maksimum Fisher Bilgisi ve Kullbak-Leibler madde seçme yönteminden; Beklenen Sonsal Dağılım yetenek kestirimi kullanıldığında ise Maksimum Fisher Bilgisi ve a-tabakalama madde seçme yönteminden elde edilmiştir. Durdurma kuralının $SH < 0.2$ ve $SH < 0.4$ olduğu koşullarda, bütün madde seçme yöntemlerinde, En Çok Olabilirlik Tahmini yetenek kestirimi kullanıldığında elde edilen ortalama madde sayısı, Beklenen Sonsal Dağılım yetenek kestirimi kullanıldığında elde edilen ortalama madde sayısından daha yüksek bulunmuştur. BBT uygulamalarında Beklenen Sonsal Dağılım yetenek kestiriminin test uzunluğunu kısalttığı sonucuna varılmıştır. Araştırmaya alınan bütün madde seçme yöntemlerinin; madde havuzu kullanımına ilişkin iyi bir denge göstermediği, yüksek a-parametresine sahip maddeleri daha çok kullandığı sonucuna varılmıştır.

Anahtar Sözcük: Bireyselleştirilmiş bilgisayarlı test uygulaması, madde seçme yöntemleri, Maksimum Fisher Bilgisi, a-tabakalama, Olabilirlik Ağırlıklı Bilgi Ölçütü, Aşamalı Maksimum Bilgi Oranı, Kullbak-Leibler

ABSTRACT

SULAK, Sema, Comparision of Item Selection Methods in Computerized Adaptive Testing, Ph.D. Dissertation, Ankara, 2013

The advantage of computer adaptive tests (CAT) is that the test takers encounter items matched to an examinees' ability level selecting the most informative items based on test takers' abilities is so critical in order to provide this advantage. Item selection can be seen as a core component of CAT, because it dominates how the CAT starts, continues, and stops for an examinee.

In this research, item selection methods were investigated in terms of test termination rule and ability estimation. For this aim, 250 items and 2000 people, who has mean of zero and standard deviation of 1, were created simulatively. Simulated CAT conditions were created based on item selection methods (Maksimum Fisher Information, a-stratification, Likelihood Weight Information Criterion, Gradual Maksimum Information Ratio, Kullbak-Leibler), abilitiy estimation methods (Maksimum Likelihood Estimation, Expected a Posteriori) and test termination rule (40 items, $SE < 0.2$ and $SE < 0.4$).

According to the results that compare fixed-length termination rule (40 items); SE value when used Maksimum Likelihood Estimation ability estimation method was bigger than SE value when used Expected a Posteriori ability estimation method. In Maksimum Likelihood Estimation ability estimation conditions, when test length is $n < 30$ a-stratification and is $n > 30$, Kullbak-Leibler item selection method had the highest SE value. On the other hand, during Expected a Posteriori ability estimation conditions for all test lengths the highest SE value was obtained by a-tabakalama item selection method. According to the results that compare item pool utilization; the most accurate results were gathered by a-stratification item selection method.

It is found that when termination rule was $SE < 0.2$ and Maksimum Likelihood Estimation was utilized as ability estimation method, the lowest and the highest item averages were for Gradual Maksimum Information Ratio and Maksimum Fisher Information item selection methods; on the other hand, when Expected a Posteriori was utilized, they were Kullbak-Leibler and Likelihood Weight Information Criterion item selection methods. Under the conditions where termination rule was $SE < 0.4$ and Maksimum Likelihood Estimation was used as ability estimation method, Maksimum Fisher Information had the highest and Kullbak-Leibler had the lowest item number mean; when Expected a Posteriori was utilized Maksimum Fisher Information and a-tabakalama methods had the highest and the lowest means. Overall, it was found that when termination rule was $SE < 0.2$ and $SE < 0.4$ and by all item selection methods, item number average is higher for Maksimum Likelihood Estimation in comparison to Expected a Posteriori ability estimation. It was found that Expected a Posteriori ability estimation had an impact on shortening the test length. According to the results items that had higher a-parameters were used more during simulation process that means the item selection methods were not balanced for item pool.

Keywords: Adaptive testing, item selection, Maksimum Fisher Information, a-stratification, Likelihood Weight Information Criterion, Gradual Maksimum Information Ratio, Kullbak-Leibler

İÇİNDEKİLER

KABUL VE ONAY.....	i
BİLDİRİM.....	ii
TEŞEKKÜR.....	iii
ÖZET.....	iv
ABSTRACT.....	vi
İÇİNDEKİLER.....	viii
KISALTMALAR.....	xi
TABLolar.....	xii
ŞEKİLLER.....	xiii
1. BÖLÜM.....	1
GİRİŞ.....	1
1.1. Bireyselleştirilmiş Bilgisayarlı Testin Temel Yaklaşımı ve İşleyişi.....	5
1.2. Bireyselleştirilmiş Bilgisayarlı Testin Unsurları.....	7
1.2.1. Madde Tepki Kuramı.....	8
1.2.2. Madde Havuzu.....	10
1.2.3. Test Başlama Süreci.....	11
1.2.4. Yetenek Kestirim Yöntemleri.....	11
1.2.4.1. En Çok Olabilirlik Tahmini Yetenek Kestirimi.....	11
1.2.4.2. Bayesian Yetenek Kestirimi.....	12
1.2.5. Madde Seçme Yöntemleri.....	14
1.2.5.1. Maksimum Fisher Bilgisi.....	14
1.2.5.2. Kullbak-Leibler Bilgisi.....	15
1.2.5.3. Aralık Bilgisi Ölçütü.....	16
1.2.5.4. Olabilirlik Ağırlıklı Bilgi Ölçütü.....	17

1.2.5.5. a-tabakalama.....	17
1.2.5.6. Aşamalı Maksimum Bilgi Oranı.....	18
1.2.5.7. En Uygun b Değeri.....	19
1.2.5. Test Durdurma Kuralı.....	19
1.3. Problem.....	21
1.3.1. Alt Problemler.....	21
1.4. Sınırlılıklar.....	22
1.5. Araştırmanın Amacı.....	22
1.6. Araştırmanın Önemi.....	24
1.7. İlgili Araştırmalar.....	25
2. BÖLÜM.....	33
YÖNTEM.....	33
2.1. Araştırmanın Türü.....	33
2.2. Verinin Elde Edilmesi.....	32
2.3. Verinin Analizi.....	33
3. BÖLÜM.....	38
BULGULAR VE YORUM.....	38
3.1. Birinci Alt Probleme İlişkin Bulgular.....	38
3.2. İkinci Alt Probleme İlişkin Bulgular.....	43
3.3. Üçüncü Alt Probleme İlişkin Bulgular.....	47
3.4. Dördüncü Alt Probleme İlişkin Bulgular.....	50
4. BÖLÜM.....	57
SONUÇ VE ÖNERİLER.....	57
4.1. Birinci Alt Probleme İlişkin Sonuçlar.....	57
4.2. İkinci Alt Probleme İlişkin Sonuçlar.....	58
4.3. Üçüncü Alt Probleme İlişkin Sonuçlar.....	58
4.4. Dördüncü Alt Probleme İlişkin Sonuçlar.....	59
4.5. Öneriler.....	60

5. KAYNAKÇA.....	62
EKLER.....	68
Ek 1. Araştırmanın Yürütüldüğü Grup Dağılım Grafiği.....	69
Ek 2. SimulCAT Bilgisayar Programı.....	70
Ek 3. Madde Havuzu Parametre Değerleri.....	72

KISALTMALAR

a: Madde ayırt edicilik parametresi

ABÖ: Aralık Bilgisi Ölçütü

AMBO: Aşamalı Maksimum Bilgi Oranı

b: Madde güçlük parametresi

BBT: Bireyselleştirilmiş Bilgisayarlı Test

BSD: Beklenen Sonsal Dağılım

c: Şans parametresi

EOT: En Çok Olabilirlik Tahmini

GMAT: Graduate Management Admission Test

GRE: The Graduate Record Exam

MFB: Maksimum Fisher Bilgisi

MTK: Madde Tepki Kuramı

OAB: Olabilirlik Ağırlıklı Bilgi Ölçütü

OSH: Ortalama Standart Hata

SH: Tahminin Standart Hatası

TABLOLAR

Tablo 1: Kağıt Kalem Testi ve Bireyselleştirilmiş Test Arasındaki Farklar.....	2
Tablo 2: Madde Havuzu Betimsel İstatistikleri.....	34
Tablo 3: BBT Koşullarının Özellikleri.....	35
Tablo 4: Test durdurma kuralının 40 madde olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında, test uzunluğuna göre madde seçme yöntemlerine ait istatistikler.....	39
Tablo 5: Test durdurma kuralının 40 madde olarak belirlendiği ve BSD yetenek kestiriminin kullanıldığı BBT koşullarında, test uzunluğuna göre madde seçme yöntemlerine ait istatistikler	41
Tablo 6: Testi durdurma kuralının sabit standart hataya dayalı olarak belirlendiği BBT koşullarında, yetenek kestirim ve madde seçme yöntemlerine ait istatistikler.....	48

ŞEKİLLER

Şekil 1: BBT Uygulamasının Şematik Gösterimi	6
Şekil 2: Test durdurma kuralının 40 madde olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları.....	44
Şekil 3: Test durdurma kuralının 40 madde olarak belirlendiği ve BSD yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları.....	46
Şekil 4: Test durdurma kuralının $SH < 0.2$ olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları.....	51
Şekil 5: Test durdurma kuralının $SH < 0.2$ olarak belirlendiği ve BSD yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları.....	52
Şekil 6: Test durdurma kuralının $SH < 0.4$ olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları.....	54
Şekil 7: Test durdurma kuralının $SH < 0.4$ olarak belirlendiği ve BSD yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları.....	55

1. BÖLÜM

GİRİŞ

Bireysel farklılıkların dikkate alındığı bir öğrenme ortamında, bireysel değişimlerin ölçülmesi de anlamlı ve önemli bir hal almaktadır. Ölçme değerlendirme alanında, bireysel farklılıkların dikkate alınması “tailored” veya “adaptive” yani bireyselleştirilmiş test kavramını gündeme getirmektedir.

Bireyselleştirilmiş testler için üç önemli özellik; cevaplayıcıya sorulacak ilk maddenin cevaplayıcı hakkında daha önceden bilinen bilgilerin kullanılarak seçilmesi, daha sonraki maddelerin kişinin verdiği cevaplar doğrultusunda seçilerek sorulması ve yeterli sayıda madde uygulandığında teste son verilmesidir (Weiss, 1976). Bu özellik klasik testler ile bireyselleştirilmiş testler arasındaki önemli bir farkı ortaya koymaktadır.

Bireyselleştirilmiş testlerde birbirinden farklı cevaplayıcılar farklı test formları alır. Yüksek yetenek düzeyindeki bir cevaplayıcı ile düşük yetenek düzeyindeki bir cevaplayıcı testte sadece birkaç ortak soruyla karşılaşabildiği gibi; bazen hiçbir ortak soruyla karşılaşmazlar. Bu durum herkesin aynı formu aldığı geleneksel kağıt kalem testlerinde görülmez. Kağıt kalem testlerinde bütün cevaplayıcılar belirli bir test formundaki bütün maddelerle kendi yetenekleri dikkate alınmaksızın karşılaşır ve bu testlerde “doğru cevap sayısı” ölçüsü işlemektedir (Wainer,2000).

Tian, Miao, Zhu, ve Gong (2007) yaptıkları araştırmada, kağıt kalem testleri ile bireyselleştirilmiş test uygulamaları arasındaki farkı Tablo 1’de verildiği şekilde özetlemişlerdir:

Tablo 1: Kağıt Kalem Testi ve Bireyselleştirilmiş Test Arasındaki Farklar

	Kağıt Kalem Testi	Bireyselleştirilmiş Test
Test düzeni	Her birey aynı testi alır	Her birey farklı bir test alır
Testin gücü	Orta güçlükteki adayı hedef alır	Bireysel olarak adayı hedef alır
Test uzunluğu	Her birey için aynıdır. Test uzundur	Her birey için farklıdır. Test daha kısadır
Test anı	Bütün bireylerin test edildiği sabit bir an	Her öğrenciye uygun herhangi bir an
Test organizasyonu	Çok zaman alır	Daha az zaman alır
Hemen sonuç verme	Hayır	Evet

Tablo 1’de görüldüğü üzere, bireyselleştirilmiş testin hem psikometrik hem de pratik olarak faydası vardır. Psikometrik olarak faydası tek bir kelimedede özetlenebilir “etkililik”. İyi yapılandırılmış bir bireyselleştirilmiş test, güçlük düzeyi bakımından uygun olmayan ve yeteneğin kestirilmesinde katkı sağlamayan maddeleri bireye uygulamaz. Böylece, etkili ve hassas yetenek tahminleri sağlanmış olur ve aynı zamanda test uzunluğunu kısalar (McBride, 1985; Wainer, 2000).

Bireyselleştirilmiş testlerin gelişim sürecine bakıldığında; Alfred Binet’in 1905 yılında zeka testleri ile yaptığı çalışmalar, çok basit gibi görünse de bireyselleştirilmiş test uygulamalarının temelini oluşturmaktadır. Binet, bir testin, maddeleri zorluklarına göre kademelendirerek, testi alan her kişiye göre uyarlanabileceğini fark etmiştir. Böylece Binet, grubun özelliğini belirlemek yerine bir bireyin özelliğini belirlemeye çalışmıştır (Linacre, 2000; Van Der

Linden ve Glas, 2010). Binet'in test uygulaması bireyselleştirilmiş teste iyi bir örnektir. Testte önceden kalibre edilmiş maddelerden oluşturulan bir havuz kullanılır. Testin değişken bir başlama noktası vardır. Testi alan kişinin genellikle kronolojik yaşı dikkate alınarak test başlatılabildiği gibi, o kişiye ait önceden sahip olunan bilgi varsa daha düşük veya daha yüksek bir düzeyden de test başlatılabilir. Testi alan kişinin verdiği cevaplar test uygulayıcısı tarafından o anda puanlanır ve değerlendirilir. Yeni sorulacak maddelerin seçimi ise testi alan kişinin önceki madde setlerinde gösterdiği performansına dayanır. Eğer cevaplayıcı maddelerin çoğunu cevaplamışsa daha yüksek yaş düzeyindeki zor maddeler, aksi durumda ise daha düşük yaş düzeyindeki kolay maddeler uygulanır. Testi alan kişinin tüm maddeleri yanlış yanıtladığı yaş düzeyinde ise test sonlandırılır (Weiss, <http://iacat.org/node/442>). Hambleton, Swaminathan ve Rogers (1991) ve Wainer (1990) yaptıkları araştırmalarda Binet'in bu test uygulamasından sonra, 1950'lerin başına kadar bireyselleştirilmiş test konusunun gündem dışı kaldığını belirtmişlerdir.

Bireyselleştirilmiş test uygulaması 1952 yılında Lord'un bir yorumuyla yeniden gündeme gelmiştir: Lord, "gözlenen puanlar testten bağımsız olduğunda, yetenek puanlarının da testten bağımsız" olacağını savunmuştur (Antal, Eros, Imre, 2010). Bir diğer önemli katkı 1960 yılında Rasch tarafından gerçekleştirilmiştir. Rasch, yayınladığı kitabında madde tepki modelini açıklamıştır ve bu model psikometri alanını etkilediği gibi bireyselleştirilmiş test uygulamalarına da farklı bir yön kazandırmıştır.

Lord'un, 1960'ların sonunda ABD' nin test kurumu olan Educational Testing Service (ETS)'de başlattığı çalışmalar da geliştirici olmuştur. Lord yaptığı araştırmalarda soru sayısının sabit olduğu testlerin yüksek ve düşük yetenek düzeyindeki cevaplayıcılar için uygun olmayacağını, bireye uyarlanmış sorular kullanıldığında aynı testin hiçbir bilgi kaybı olmaksızın daha kısa sürede uygulanabileceğini belirtmiştir (Weiss, 1983).

Lord 1970 yılında bireyselleştirilmiş testlerin teorik temellerini açıklamıştır ve bu ilk çalışmasında madde tepki kuramıyla, testi alan kişilerin bireysel yetenek düzeylerini teste uyarlamanın (tailor) mümkün olduğunu savunmuştur (Slater,2001). Bu öngörü, 1980'lere kadar bilgisayar teknolojisine dayalı programlama gücüne olan gereklilikten dolayı uygulanabilir olmamıştır.

Bilgisayar teknolojisinin gelişmesi ve kullanımının yaygınlaşması sonucunda bireyselleştirilmiş test uygulamalarında bilgisayarlardan faydalanılmaya başlanmış ve bu durum pek çok istatistiksel ve psikometrik anlamda yeniliğin uygulanma avantajını artırmıştır. Bu bilgisayar teknolojisi, madde tepki kuramı ile birleşince Lord'un öngörüsünü gerçeğe dönüştürmüştür ve günümüzde "bireyselleştirilmiş bilgisayarlı test (BBT)" olarak bilinen test uygulaması, teorik bir araştırma konusu olmaktan çıkıp operasyonel test programlarında yaygın bir şekilde kullanılmaya başlanmıştır.

Reckase 1974 yılında bilgisayar ortamında bireyselleştirilmiş testlerin ilk uygulamalarını gerçekleştirmiştir (Kingsbury, 2000). 1993 yılında, ETS, ilk geniş ölçekli BBT uygulamalarını GRE-CAT (Graduate Record Examination Computer Adaptive Test) ile gerçekleştirmiştir. 1997 yılında ETS işletme programlarında yüksek öğrenim için GMAT (Graduate Management Admission Test) BBT versiyonunu uygulamaya koymuştur. 1998 yılında ise TOEFL, BBT ile düzenlenmeye başlamış ve bu uygulama 2001 yılı itibariyle kağıt-kalem testlerinin yerini almaya başlamıştır (Slater, 2001; Weiss, 2003).

Görüldüğü üzere, son yıllarda ölçme alanında bireyselleştirilmiş bilgisayarlı teste olan ilgi artmıştır ve bu testin kullanımı yaygınlaşmıştır.

1.1. Bireyselleştirilmiş Bilgisayarlı Testin Temel Yaklaşımı ve İşleyişi

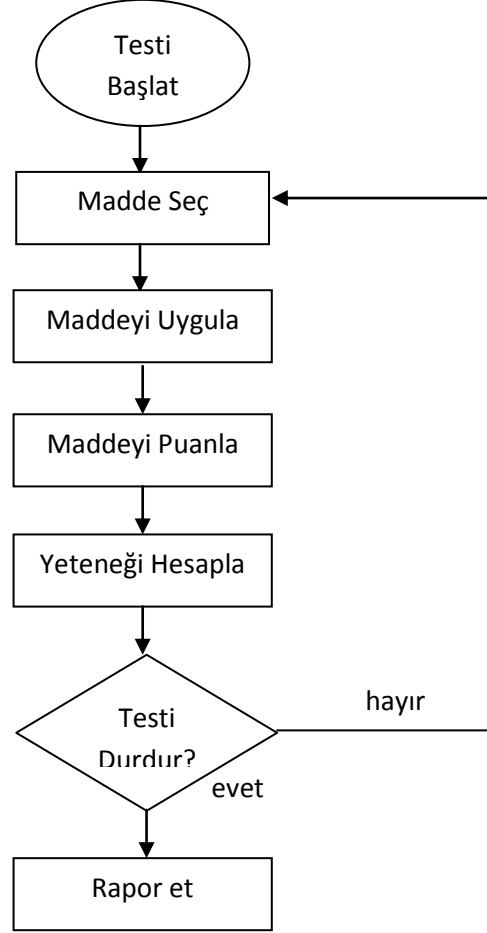
Bireyselleştirilmiş bilgisayarlı testlerde, testin yapısı ve uygulaması bilgisayara dayalı ve bireyselleştirilmiştir. Testi alan her kişi için, verdiği cevaplara bağlı olacak şekilde kestirilen yetenek düzeyine uygun maddeler seçilerek birbirinden farklı testler inşa edilmektedir. Böylece testi alan her kişi kompozisyonu kendisine optimize edilmiş farklı bir test almaktadır (Eggen, 2004).

Bireyselleştirilmiş bilgisayarlı test uygulaması seçilen maddelerin cevaplayıcıya sunulması ve verilen cevaplar yoluyla cevaplayıcının yetenek düzeyinin kestirilmesinin iteratif bir sürecidir. Yeni seçilecek madde, cevaplayıcının mevcut yetenek düzeyine uygun olarak belirlenir ve verilen cevabın ardından bu düzey tekrar hesaplanır. Test etme süreci, durdurma kuralları (belirli madde sayısına veya standart hata değerine ulaşıldığında vb) yerine gelinceye kadar devam eder (Orcutt, 2002; Thissen ve Mislevy, 2000; Wainer, 2000; Weiss, 1983).

Bireyselleştirilmiş bilgisayarlı test sürecinin algoritması testin nasıl başlayacağı, devam edeceği ve sonlandırılacağını değerlendiren kurallar setidir. Eggen (2004) bireyselleştirilmiş bilgisayarlı test algoritmasını şematik olarak Şekil 1'deki gibi açıklamıştır.

Şekil 1'de şematik olarak gösterilen bireyselleştirilmiş bilgisayarlı testin, bireyselleştirme kısmı pek çok yazılım ve donanım özelliğinin karşılanmasını gerektirir. Bireyselleştirilmiş bilgisayarlı test uygulaması madde havuzundan seçilen bir madde ile başlar. Testi alan cevaplayıcı standart bir klavye vb yardımıyla bilgisayar ekranında görünen maddeleri cevaplar. Cevaplayıcı bir madde üzerinde çalışırken bilgisayar yeni sorulacak maddeyi (madde doğru cevaplandığında veya yanlış cevaplandığında sorulmak üzere iki olası madde) hazırlamakla meşguldür.

Şekil 1: BBT uygulamasının şematik gösterimi



Testi alan kişi maddeyi okurken bilgisayar programı madde yanlış cevaplandığında veya doğru cevaplandığında yetenek kestirimini ne olacağını hesaplar ve madde havuzundan bu tahminler arasından bir madde seçer. Kişi cevaplamaı gerçekleştirir. Bilgisayar programı verilen cevabın doğru olup olmadığını kontrol eder, daha önceden hesapladığı tahminlere göre yetenek kestirimini günceller. Yeni madde seçilerek sorulur. Bütün bu işlemler, belirli bir doğrulukla yetenek kestirimi gerçekleştirilene kadar sürdürülür. Testin sonunda, bilgisayar, testin nasıl uygulandığıyla ilgili kısa bir rapor hazırlar. Bu raporlar daha sonra madde kalibrasyonları için kullanılmak üzere saklanır.

O halde bireyselleştirilmiş bilgisayarlı testler için anahtar sorular şunlardır (Wainer,2000):

- Teste başlamak için madde seçimi nasıl yapılır?
- Testi alan kişi maddeleri cevapladıkça yeni sorulacak maddelerin seçimi ve yetenek kestirimi nasıl yapılır?
- Testin ne zaman sonlanacağı nasıl belirlenir?

Yukarıda belirtilen sorular, bireyselleştirilmiş bilgisayarlı test unsurlarının çok iyi planlanmasını gerektirir.

1.2. Bireyselleştirilmiş Bilgisayarlı Testin Unsurları

Etkili BBT uygulamaları için BBT ilgilendiren konuların çok iyi planlanması gerekmektedir. BBT uygulamalarında cevaplayıcının yetenek düzeyine uygun test alması esas olduğundan, testte yer alacak maddelerin özellikleri ve bu maddelerin seçilmesi, kullanılacak ölçme modeli, her adımda yetenek kestiriminin hesaplanması konuları gündeme gelmektedir. BBT uygulamalarını oluşturan unsurlar şöyle belirlenmiştir (Kingsbury ve Zara, 1989; Orcutt, 2002; Wainer, 2000; Weiss ve Kingsbury, 1984):

- Madde Tepki Kuramı Modeli
- Madde Havuzu
- Test Başlama Süreci
- Madde Seçme Yöntemleri
- Yetenek Kestirim Yöntemleri
- Test Durdurma Kuralları

1.2.1. Madde Tepki Kuramı

Madde Tepki Kuramı modelleri BBT uygulamalarının merkezidir. Madde Tepki Kuramı, cevaplayıcı bir madde ile karşılaştığında ne olduğunu açıklayan matematiksel tanımlar ailesidir. MTK'nın temel varsayımı, bir ya da birden fazla özellik sayesinde, bir kişinin bir testte göstereceği performansın kestirilebileceğidir. Bu kestirim belirlenirken, bireyin gözlenebilir performansı ile, bu performansı belirlediği düşünülen yetenekler ve özellikler arasındaki ilişki matematiksel olarak ortaya konulur. Bu matematiksel tanım madde karakteristiğinin bir fonksiyonudur ve madde karakteristik eğrisi ($P_i(\theta)$) olarak ifade edilir. Madde karakteristik eğrisi, belirli bir yetenek düzeyindeki kişinin, bir maddeyi doğru yanıtlama olasılığı hakkında bilgi verir (Hambleton ve Swaminathan, 1985). Bu şartlı olasılığın fonksiyonu, "normal ogive" veya "lojistik ogive" olmaktadır. Eggen (2004) iki kategorili olarak puanlanmış maddeler için en popüler olarak kullanılan modellerin, lojistik tipteki modeller olduğunu belirtmiştir. Bu modellerde, bir kişinin yetenek düzeyi θ ile i maddesini doğru cevaplama olasılığı arasındaki ilişki, maddelerin karakteristikleri ya da parametreleri tarafından saptanır. İki kategorili MTK modellerinde karşılaşılan madde parametreleri şöyledir (Hambleton ve Swaminathan, 1985) :

a madde ayrıcalık parametresi, madde karakteristik eğrisinin eğimidir. b madde güçlük parametresi, bir maddenin %50 olasılıkla doğru yanıtlanması için gerekli yetenek düzeyini gösterir. c parametresi, şans parametresi olarak adlandırılır ve en düşük yetenek düzeyine sahip cevaplayıcıların maddeyi doğru cevaplama olasılığını gösterir.

MTK'da madde karakteristik fonksiyonlarının matematiksel yapılarına ve ölçülen değişkenin özelliğine göre farklı modeller oluşturulmuştur. Eldeki araştırmada iki kategorili veriler kullanıldığından diğer MTK modelleri açıklanmamıştır. İki kategorili MTK modelleri kullanılan parametre sayısına göre adlandırılmaktadır.

En basit madde tepki kuramı modeli, b ve θ arasındaki etkileşimi lojistik fonksiyonla birleştirir. Bu model, Danimarkalı matematikçi George Rasch tarafından geliştirilmiştir ve bir parametrelili lojistik model olarak adlandırılmıştır. İki parametrelili lojistik model, madde güçlüğü yanında, madde ayrıcalığını da fonksiyona dahil eder. Üç parametrelili lojistik modelde ise madde güçlüğü ve madde ayrıcalığının yanına şans parametresi c eklenir. Üç parametrelili lojistik model matematiksel olarak şu şekilde ifade edilir:

$$P_{ij}(x_i = 1 | \theta_j) = c_i + (1 - c_i) \left[\frac{\exp(a_i(\theta_j - b_i))}{1 + \exp(a_i(\theta_j - b_i))} \right]$$

Eşitlikte, θ_j yetenek düzeyi için j kişinin, i maddesini doğru cevaplama olasılığı ifade edilmiştir.

MTK, testi alan kişiler farklı sorularla karşılaşsa bile, farklı kişilerin yeteneklerini kestirmek için standart bir çerçeve sunar (Hambleton, Swaminathan ve Rogers, 1991). MTK, aynı zamanda testi alan kişinin yeteneğini ve maddenin güçlüğü aynı ölçekte tanımlar. Madde güçlüğü ile yetenek düzeyini eşleştirmek istenen bir durumdur. Çünkü, test maddeleri orta güçlükte olduğunda daha efektif olarak ölçüm yapılabilmektedir. Test süreci boyunca testi alan kişinin tahmini yeteneğini belirleyerek, tahmin edilen yetenekle eşleşen maddeleri uygulamak mümkün hale gelmektedir (örneğin, testi alan kişi tarafından %50 olasılıkla doğru cevaplanabilecek maddeler) ve testin etkililiği arttırılmaktadır (Slater, 2001). Böylece MTK, bireyselleştirilmiş test uygulamalarında kullanışlı bir hal almaktadır. MTK, her yetenek kestirim düzeyi için bireysel standart hatayı değerlendirme imkanı sunar ki bu durum BBT uygulamalarında testi durdurma kuralı olarak kullanılabilmektedir.

1.2.2. Madde havuzu

Madde havuzu, cevaplayıcıların yeteneklerini iyi tanımlayacak şekilde inşa edilmiş maddelerin yapılandırılmış bir koleksiyonudur (Eggen, 2004). Madde havuzu ne kadar iyi olursa BBT test algoritması o kadar iyi sonuç verir. İyi bir madde havuzunun, pek çok farklı yetenek düzeyi için yüksek kalitede maddelerden oluşturulması gerekmektedir.

BBT uygulamalarında madde istatistikleri hesaplanmaz, önceden geliştirilmiş ve istatistikleri hesaplanmış maddelerden yararlanılır. Eğer madde iyi yapılandırılmamışsa hiçbir istatistiksel yöntem veya elektronik beceri yardımcı olamaz (Wainer, 2000). Öncelikli olarak; madde havuzunun çok sayıda madde içermesi gerekir. Urry (1977) yaptığı araştırmada, bireyselleştirilmiş bilgisayarlı testte daha etkili yetenek tahminleri yapmak için en az 100 soruluk bir madde havuzuna ihtiyaç olduğunu bulmuştur. Ancak 100 maddelik bir havuz yeterli değildir. BBT için gerekli madde sayısını Kingsbury ve Zara (1989) “ne kadar çok olursa iyidir” şeklinde açıklamıştır. Madde havuzunun, test uzunluğunun altı ila on iki kat fazlası olacak şekilde madde içermesi gerektiğini belirten araştırmalar da vardır (Stocking, 1992). Madde havuzuyla ilgili olarak dikkat edilmesi gereken hususlardan birisi de madde kullanım sıklığıdır.

Madde kullanım sıklığı (item exposure), testi alan kişilerin madde havuzunda yer alan bazı maddelerle sıklıkla karşılaşması sorunudur. Farklı cevaplayıcılara aynı maddelerin tekrar edilmesi, maddenin ezberlenmesine olanak sağlayabilir ve bu durum test güvenliğini tehlikeye sokabilir. Bunun önüne geçmek için farklı yöntemler araştırılmaktadır. Bu durumu aşmak için geniş bir madde havuzu oluşturmak ve sürekli olarak madde havuzuna yeni maddeler eklemek gerekir. Madde havuzuna yeni maddelerin eklenmesi pretesting olarak adlandırılmaktadır.

1.2.3. Test Başlama Süreci

Testin başlangıcında seçilen ilk madde ölçme için kritik değildir ancak bireyin psikolojik durumunu etkilemesi bakımından kritiktir. Madde çok zor olursa, testi alan kişi başarısızlık duygusuyla diğer maddeleri cevaplamada isteksiz olabilir, madde çok kolay olduğunda ise testi alan kişi testi ciddiye almayıp maddeleri cevaplamada daha dikkatsiz davranabilir (Slater, 2001; Wainer, 2000).

Teste başlamak için ilk maddenin seçilmesinde farklı yöntemler vardır. Cevaplayıcı hakkında önceden sahip olunan bilgi (önceki testlerden aldığı puanlar, karne notu vb) ilk maddenin seçilmesinde kullanılabilir. BBT uygulamalarına başlamadan önce cevaplayıcıların nihai test puanlarına etki etmeyecek madde setleri, tüm cevaplayıcılara uygulanır ve elde edilen yetenek düzeyi ilk maddenin seçilmesinde kullanılabilir (Slater, 2001; Sireci, 2003).

1.2.4. Yetenek Kestirim Yöntemleri

BBT uygulamalarında yaygın olarak kullanılan yetenek kestirim yöntemleri, En Çok Olabilirlik ve Bayesian kestirime dayalı olan yöntemlerdir.

1.2.4.1. En Çok Olabilirlik Tahmini Yetenek Kestirimi

En Çok Olabilirlik Tahmini (EOT) yetenek kestirim yöntemi, olabilirlik fonksiyonunu maksimum yapan θ değerini elde etmeye çalışır. Bu olabilirlik fonksiyonu aşağıdaki şekilde ifade edilmektedir.

$$L(u|\theta) = \prod_{i=1}^n P_i(u_i|\theta)$$

Eşitlikte u cevap vektörünü, $P_i(u_i|\theta)$, testi alan kişinin θ yeteneğinde i maddesine vereceği cevap örüntüsü olarak ifade edilen u_i 'yi elde etmenin

olasılığını, n ise madde sayısını göstermektedir. Cevaplayıcının θ yeteneğinin maksimum olabilirlik tahmini gösteren $\hat{\theta}$, bu olabilirlik fonksiyonu maksimum yapan değerdir. Bu değer, yukarıdaki eşitlikte verilen olabilirlik fonksiyonun türevlerinin alınarak sıfıra eşitlenmesinin çözümünden elde edilir. Bu eşitlikleri çözmek için iteratif yöntemlere başvurulmaktadır ve bu amaçla kullanılan en yaygın yöntem Newton-Rahpson yöntemidir (Wang ve Vispoel,1998).

EOT yetenek kestirimi BBT uygulamalarında yaygın olarak kullanılmaktadır. Çünkü, etkili bir kestiricidir ve yansızdır. EOT yetenek kestiriminin bir dezavantajı ise, testte yer alan maddelerin tümü doğru ya da yanlış olduğunda kestirimde bulunamaz, madde sayısı az olduğunda uygun bir çözüm vermeyebilir.

1.2.4.2. Bayesian Yetenek Kestirimi

Bayesian yönteminde, EOT'den farklı olarak yetenek kestirimine, yetenek dağılımları varsayımı dahil edilmektedir. Bayesian yöntemi, cevap örüntüsünün gözlenmesinden önceki yetenek düzeyi bilgisinden faydalanmaktadır. Bayesian yöntemleri, maddelerin doğru veya yanlış cevaplanmasından elde edilen olabilirlik bilgisi ile yetenek dağılımlarını birleştirerek yetenek kestirimlerini ortaya koyar. Başlangıçta, yetenek dağılımlarının ortalaması sıfır ve standart sapması bir olan normal bir dağılımda olduğu varsayılır. Bu başlangıçtaki dağılım önsel dağılım olarak adlandırılır. Testi alan kişi ilk maddeyi cevapladıktan sonra, verilen cevapla ilişkili olan olabilirlik, önsel dağılımla birleşerek; sonsal dağılım olarak adlandırılan yetenek dağılımını oluşturur. Elde edilen sonsal dağılım ikinci maddenin cevaplanmasında önsel dağılım olarak kullanılır. Bu süreç test sonlanıncaya kadar devam eder (Wang, Vispoel, 1998).

Bayesian yöntemler, θ düzeyine ilişkin bir tahmin yapmak için en azından iki farklı kategoride yanıt gereksinim duyan en çok olabilirlik yönteminin aksine, çok az sayıda maddeden sonra tahmin edebilme avantajına sahiptir. Bayesian

yöntemlerin önemli bir dezavantajı ise, seçilen önsel olasılıktan kaynaklanan bir yanlılık göstermesidir. Bayesian yöntemi altında önsel ve sonsal dağılımların formlarına ve sonsal dağılımın karakteristiklerine göre farklı varsayımlarda yöntemler geliştirilmiştir. Bunlar, Owen yetenek kestirim, Maksimum Sonsal Dağılım yetenek kestirim ve Beklenen Sonsal Dağılım yetenek kestirim yöntemleridir. Araştırmada Beklenen Sonsal Dağılım (BSD) yetenek kestirim yöntemi kullanıldığından diğer yöntemlere değinilmemiştir.

1.2.4.2.1. Beklenen Sonsal Dağılım (Expected a Posteriori) Yetenek Kestirimi

Beklenen Sonsal Dağılım (BSD) yetenek kestirim yöntemi, yetenek kestiriminde sonsal dağılımın ortalamasından faydalanır ve her adımda normallik varsayımını dikkate almaz. Bock ve Aitkin (1981) BSD yetenek kestirimini aşağıdaki eşitlikle ifade etmiştir:

$$EAP(\theta) = E(\theta|u) = \int_{-\infty}^{\infty} P(\theta|u)\theta d\theta$$

$E(\theta|u)$, sonsal dağılım ortalamasını; $P(\theta|u)$, sonsal dağılımı; $\int_{-\infty}^{\infty} \theta d\theta$ ise bütün θ değerlerinin üstündeki alanı göstermektedir (Ho,2010).

BSD yetenek kestiriminin avantajlarından birisi, karışık iteratif matematiksel hesaplamalara gerek duymamasıdır. BSD yetenek kestirimi düşük standart hata değeri gösterir ve EOT yetenek kestiriminin aksine, testi alan kişinin hiç doğru test maddelerinin hiçbiri doğru olmadığı veya hepsi doğru olduğu durumlarda da yetenek kestirimini gerçekleştirir (Hambleton, Swaminathan ve Rogers, 1991)

1.2.5. Madde Seçme Yöntemleri

Madde seçme yöntemi, bireyselleştirilmiş bilgisayarlı testin önemli bileşenlerinden birisidir. BBT uygulamalarında kullanılan belli başlı madde seçme yöntemleri şunlardır:

1.2.5.1. Maksimum Fisher Bilgisi

Bireyselleştirilmiş bilgisayarlı testte yaygın olarak kullanılan madde seçme yöntemi, Maksimum Fisher Bilgisi (MFB)dir. Bu yöntem, her maddeye verilen cevaptan sonra yapılan yetenek kestirimi (interim yetenek kestirimi) $\hat{\theta}$ da o ana kadar uygulanan $m-1$ madde için, $I[\hat{\theta}_{m-1}]$ 'i maksimum yapan x maddesini bulmayı amaçlar. Madde karakteristiklerinin 3 parametrelili lojistik modelde tanımlandığı çoktan seçmeli bir madde havuzunda MFB'ya dayanan madde seçimi aşağıdaki eşitlikle belirlenmektedir:

$$I_i[\hat{\theta}_{m-1}] = \frac{(Da_i)^2(1-c_i)}{\left[c_i + e^{Da_i(\hat{\theta}_{m-1}-b_i)} \right] \left[1 + e^{-Da_i(\hat{\theta}_{m-1}-b_i)} \right]^2}$$

Yukarıdaki eşitlikte, a_i , b_i ve c_i ; ayrıcalık, güçlük ve şans madde parametrelerini; D ise 1.702 değerindeki ölçme sabitini göstermektedir (Han,2009). MFB popüler bir yöntemdir çünkü; BBT, her birey için maksimum test bilgisiyle sonuçlanır. Ancak bu yöntem BBT başlangıcında (örneğin 5 veya daha az madde uygulandığında) interim yetenek tahmininde yetersiz kalmaktadır ve beklendiği kadar bilgi sağlayamamaktadır. Tahmin edilen yetenek düzeyi, cevaplayıcının gerçek yetenek düzeyine yakın olmadığı zaman, tahmin edilen yetenek düzeyine yakın maksimum local bilgiye sahip olan maddeyi seçmek çok mümkün olmayabilir. Bu durum, BBT uygulamalarının başında meydana gelmektedir (Linda, 1996). Bir diğer problem ise, bu yöntemin düşük a parametre değerine sahip maddelerden daha çok yüksek a

parametresi değerine sahip maddeleri seçmesidir. Ho (2010) MFB madde seçme yönteminde Fisher bilgisinin, madde ayrılcılığının bir fonksiyonu olduğunu belirtmiştir. Yani, yüksek ayrılcılıktaki bir madde düşük ayrılcılıktaki bir maddeden daha çok bilgi sağlamaktadır. Bu yüzden MFB madde seçme yöntemi, yüksek ayrılcılıktaki maddeleri seçmeye meyillidir. Bu durum, BBT uygulamalarında madde havuzunun yanlış kullanılmasına yol açmakta ve madde havuzu kullanımında (maintenance) problem oluşturmaktadır. Bahsedilen problemlerin önüne geçmek için farklı madde seçme yöntemleri araştırılmıştır.

1.2.5.2. Kullbak-Leibler Bilgisi

Chang ve Ying (1996), global bilgi yaklaşımına dayanan Kullbak-Leibler bilgisi madde seçme yöntemini geliştirmişlerdir. Bir i maddesi için Kullbak-Leibler (KL) bilgisi şöyle tanımlanmıştır:

$$K_i(\theta \parallel \theta_0) = P_i(\theta_0) \log \left[\frac{P_i(\theta_0)}{P_i(\theta)} \right] + [1 - P_i(\theta_0)] \log \left[\frac{1 - P_i(\theta_0)}{1 - P_i(\theta)} \right]$$

KL bilgisi, bu iki değişkenin (θ ve θ_0) bir fonksiyonudur ve üç boyutlu uzayda bir yüzeydir. Bu iki θ düzeyinin bir fonksiyonu olarak, KL bilgisi, iki θ düzeyi arasında bir maddenin değişme kapasitesini karakterize eder.

Fisher bilgisi, belirli bir θ 'da bir maddenin farklılaşma gücünü gösterir ve theta'nın civarındaki bilginin ölçüsünü yani "yerel bilgi"yi kullanır. KL bilgisi ise, θ düzeylerinin daha geniş ranji boyunca farklılaşma gücünü gösterir ve bunun için "global bilgiyi" kullanır.

Gerçek yetenek düzeyi θ_0 bilinmediğinden ve θ parametresi belirsiz olduğundan KL bilgisi direkt olarak hesaplanamaz. Chang ve Ying (1996), θ tahminlerinin

optimal aralık değerlerinin ortalamalarını alarak KL bilgisini, BBT uygulamalarına uygulamıştır:

$$K_i(\theta_0) = \int_{\theta_0-\delta}^{\theta_0+\delta} K_i(\theta|\theta_0) d\theta$$

Yerel bilgiyi kullanan MFB madde seçme yöntemi, θ düzeyinin konumu hakkında yeterli bilgiye sahip olduğu zaman BBT uygulamalarının sonraki aşamalarında kullanılabilirken; global bilgiyi kullanan KL madde seçme yöntemi, θ düzeyinin konumu hakkında yeterli bilgiye sahip olunmasa da, BBT uygulamalarının başlangıç aşamalarında kullanılabilir (Ho,2010). Chang ve Ying (1996) MFB yerine KL kriterinin kullanılmasının test uzunluğu kısa olduğunda (özellikle madde sayısı 30'dan az olduğu durumlarda) ve testin başlangıcında yetenek kestiriminin hatalarını ve yanlılık düzeylerini düşürdüğünü belirtmişlerdir (Akt: Han, 2009).

1.2.5.3. Aralık Bilgisi Ölçütü

Veerkamp ve Berger(1997), tarafından MFB madde seçme yöntemine alternatif olarak geliştirilmiştir. Aralık Bilgisi Ölçütü (ABÖ)'nde bilgi fonksiyonu interim yetenek kestiriminin güvenli aralığı boyunca ortalanmıştır. Bir i maddesi için ABÖ'nün matematiksel gösterimi şöyledir:

$$\int_{\hat{\theta}_L}^{\hat{\theta}_R} I_i[\theta] d\theta$$

Yukarıdaki formülde, $\hat{\theta}_L$ ve $\hat{\theta}_R$, θ 'nın güven aralıklarını göstermektedir. ABÖ'nin gerçek ortalama değeri, güven aralığının (confidence interval) uzunluğu ile bölünmüş olan yukarıdaki formüldür. Fakat, bunun ortalamasını almak madde seçim yönteminin amacı için gereksizdir.

1.2.5.4. Olabilirlik Ağırlıklı Bilgi Ölçütü

Veerkamp ve Berger tarafından ikinci bir alternatif olarak Olabilirlik Ağırlıklı Bilgi Ölçütü (OAB) yöntemi geliştirilmiştir. Bu yöntemde, bilgi fonksiyonu, θ ölçeği boyunca toplanır ve madde uygulamalarından sonra olabilirlik fonksiyonuyla ağırlıklandırılır. OAB kriterinde seçilecek i maddesi aşağıdaki eşitliğin değerini maksimum yapacak maddenin seçilmesiyle belirlenir:

$$\int_{\theta=-\infty}^{\infty} L(\theta; x_{m-1}) I_i[\theta] d\theta$$

Yukarıdaki formülde, $L(\theta; x_{m-1})$; $(m-1)$ inci madde uygulandıktan sonra cevaplayıcı vektörü x_{m-1} 'in olabilirlik fonksiyonunu göstermektedir. Veerkamp ve Berger (1997) yaptıkları simülasyon çalışmasında ABÖ ve OAB yöntemlerini MFB ile karşılaştırmıştır ve her iki yöntemin madde seçimi bakımından MFB'ya önemli bir üstünlüğü olmadığı sonucuna varmışlardır (Akt:Han, 2009).

Şimdiye kadar bahsedilen her üç yöntem de maddeleri, madde bilgi fonksiyonlarına dayanarak bir noktadaki θ tahmini yerine, bir θ ranjı boyunca değerlendirmiştir. Her üç yöntem de MFB'da görülen büyük standart hata problemini, tek bir noktadaki tahmin yerine tahmin ranjını kullanarak çözmeye çalışmıştır. Ancak bu yöntemler, MFB da karşılaşılan yüksek a parametre değerlerini seçme problemine bir çözüm önermemektedir (Han,2009).

1.2.5.5. a-tabakalama

Bu yöntemde test uygulamasının başlangıcında yüksek ayrıcılıktaki maddelerin seçilmesinin önüne geçilmeye çalışılmaktadır. Chang ve Ying (1999), madde havuzundaki maddeleri a parametre değerlerine göre tabakalandırmayı önermiştir. a-tabakalama adı verilen bu yöntemde; BBT uygulamalarının başında interim θ ya en yakın b parametresi değerine sahip madde en düşük a parametresine sahip maddelerden başlayarak madde tabaklarından seçilir. Bu

metod, BBT uygulaması sürdükçe madde havuzu tabaklarından daha yüksek a parametre değerine sahip maddeler seçilerek işletilir.

Chang ve Ying (1999) yaptıkları simülasyon çalışmasında a -tabakalama metodunun MFB metoduna kıyasla madde havuzunun kararsız kullanımının, tahminin standart hatası ve yanlılıklarının yükselmesinin önüne geçmede daha iyi sonuç verdiğini bulmuştur (Han,2009). a -tabakalama methodunun da bazı dezavantajları vardır. Madde havuzunu tabakalandırmak madde havuzunun kapasitesinin az olduğu durumlarda veya içerik dengesinin kontrol edilmesi gerektiği durumlarda maddelerin ulaşılabilirliğini sınırlandırır ve madde kullanım sıklığı (item overexposure) problemine yol açabilir. Han (2009) a ve b parametreleri arasında korelasyonel bir ilişki bulunmasının olağan dışı olduğunu; eğer maddeler a parametre değerlerine göre tabakalandırılırsa madde tabakalarının b parametresi bakımından birbirine eşit olmama durumunun doğabileceğini belirtmiştir. Test uzunluğu sabit olmadığında bu yöntemle gerçekleştirilen madde seçim performansı beklendiği kadar etkili olmayabilir.

1.2.5.6. Aşamalı Maksimum Bilgi Oranı

Bu yöntem Han (2009) tarafından geliştirilmiştir. Han, madde havuzu kullanımını geliştirmek için MFB yönteminin yerine beklenen madde etkililiğine dayanan bir yöntem önermiştir. Beklenen madde etkililiği, interim $\hat{\theta}$ 'daki maddeye ait potansiyel bilginin gerçekleşme düzeyi olarak tanımlanmıştır. Bir i maddesi, interim θ_i^* 'de maksimum potansiyel bilgisine ulaştığında, interim $\hat{\theta}$ 'da beklenen madde etkililiği şöyle hesaplanır;

$$\frac{I_i[\hat{\theta}_{m-1}]}{I_i[\theta_i^*]}$$

Üç parametrelili lojistik model kullanıldığında θ_i^* , Birnbaum eşitliği kullanılarak hesaplanmaktadır:

$$\theta_i^* = b_i + \frac{1}{Da_i} \log\left(\frac{1 + \sqrt{1 + 8c_i}}{2}\right)$$

Han(2009) madde etkililiğini (beklenen madde bilgisi) madde yeterliği üstünde hesaba katmayı önermiştir. Böylece, bu yöntemin aşağıdaki ölçütü maksimum yapan maddeyi aradığı belirtmiştir

$$\frac{I_i[\hat{\theta}_{m-1}]}{I_i[\theta_i^*]} \left(1 - \frac{m}{M}\right) + I_i[\hat{\theta}_{m-1}] \frac{m}{M}$$

Yukarıda geçen M test uzunluğunu, m ise uygulanan madde sayısının 1 fazlasını göstermektedir. Eşitliğin ilk kısmı madde yeterliğini, ikinci kısmı ise madde etkililiğini açıklamaktadır. Han(2009) yaptığı sümülasyon çalışmalarında MFB yöntemine kıyasla Aşamalı Maksimum Bilgi Oranı (AMBO) yönteminin madde havuzu kullanım özelliklerini geliştirdiği sonucuna varmıştır.

1.2.5.7. En uygun b değeri

Bu yöntem a -tabakalama yönteminin tek bir tabakalandırma yapılarak özelleştirilmiş bir uygulamasıdır. a ve c parametrelerinin yerine interim theta tahminine en yakın olan b değerindeki maddeyi seçmeyi amaçlar.

1.2.6. Test Durdurma Kuralı

BBT uygulamalarında testi durdurmak için; sabit test uzunluğu ve değişken test uzunluğu olmak üzere iki yöntem vardır (Weiss ve Kingsbury, 1984; Sireci, 2003). Sabit test uzunluğu durdurma kuralında; belirli sayıdaki madde uygulandığında test sonlandırılmaktadır. Değişken uzunluklu durdurma

kuralında ise, sabit bir standart hata değerine veya belirli bir kesme puanına ulaşıldığında teste son verilmektedir.

Yukarıda bahsedilen hususlar pek çok araştırmaya konu edilmiştir ve BBT uygulamalarının daha iyi gerçekleştirilmesinde katkı sağlamıştır. BBT uygulamalarının en belirgin avantajı, testin, cevaplayıcının yetenek düzeyine uygun şekilde düzenlenebilmesidir. BBT uygulamalarından etkili sonuçlar elde etmek için, testi alan kişiyle ilgili test bilgisini maksimum yapan maddeyi seçme kritik bir öneme sahiptir. BBT uygulamalarında MFB madde seçme yaygın olarak kullanılmaktadır. Ancak, daha önce de belirtildiği gibi bu yöntem yüksek a parametresine sahip maddeleri kullanmaktadır ve özellikle testin başlangıcında yetenek kestiriminde yetersiz kalmaktadır. Veldkamp (2012) bu sınırlıkları gideren madde seçme yöntemlerinin araştırılmasının üzerinde durulması gereken önemli bir konu olduğunu belirtmiştir. Eldeki araştırmada, MFB'ye alternatif olarak geliştirilen madde seçme yöntemleri ele alınmıştır. a-tabakalama madde seçme yönteminin, MFB'de görülen yüksek a parametre değerine sahip maddeleri seçme sorununu giderdiği yönünde araştırmalar (Chang ve Ying; 1999, Chang ve Deng, 2010) vardır ve bu sebeple eldeki araştırmaya alınmıştır. Kullbak-Leibler madde seçme yönteminin MFB'den daha iyi yetenek kestirim sağladığı yönündeki araştırmalar (Eggen, 1999) ve OAB, AMBO yetenek kestirim yöntemlerinin MFB ile karşılaştırmalı olarak incelenmesi (Han, 2009) yönündeki araştırmalardan elde edilen sonuçlar değerlendirilerek eldeki çalışmaya alınacak madde seçme yöntemleri belirlenmiştir. Weissman (2003) yaptığı araştırmada, yetenek kestirim yöntemlerinin, madde seçme yöntemlerini etkilediğini belirtmiştir. Bock ve Mislevy (1982), BSD yetenek kestiriminin EOT yetenek kestiriminden daha iyi olduğunu, Wang ve Visposel (1998) ise BSD yetenek kestiriminin daha yanlı olduğunu belirtmişlerdir. Yapılan araştırma sonuçları dikkate alınarak EOT ve BSD yetenek kestirim yöntemlerinin madde seçme yöntemlerini nasıl etkilediği araştırma konusu yapılmıştır. Testin başlangıcında özellikle $n < 5$ olduğunda MFB madde seçme yönteminin yetersiz kaldığı yönünde araştırmalar yapılmıştır (Deng, 2002; Han, 2009; Weissman, 2003). BBT uygulamalarının durdurma kuralı 40 madde

olarak belirlenmiştir ve 5, 10, 20, 30 ve 40 madde uzunluklarında madde seçme yöntemlerinin nasıl işlediği araştırılmıştır. BBT uygulamalarının popüleritesinin sebeplerinden birisi ölçüm kestiriminde kayıp olmadan test uzunluklarını kısaltmasıdır. Wainer(1990) ve Weiss(1983), test uzunluğunun BBT ile yarı yarıya düşebileceğini iddia etmiştir. Bu durumu değerlendirmek amacıyla sabit standart hataya dayalı test durdurma kuralları 0.2 ve 0.4 seçilerek; test uzunluklarına göre karşılaştırmalar yapılmıştır. Eldeki araştırmada madde seçme yöntemleri, yetenek kestirim yöntemleri ve durdurma kuralları bir arada incelenmektedir ve bu amaçla aşağıdaki sorulara cevap aranmaktadır.

1.3. Problem

Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde seçmede kullanılan yöntemlere (MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler) ilişkin standart hatalar ve madde kullanım sıklığı; yetenek kestirim yöntemlerine (EOT ve BSD) ve test durdurma kuralına (40 madde, $SH < 0.2$ ve $SH < 0.4$) göre nasıl değişmektedir?

1.3.1. Alt Problemler

- 1) Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde seçmede kullanılan yöntemlere (MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler) ilişkin standart hatalar,
 - a) Test uzunluğuna (5, 10, 20, 30 ve 40 madde) ve
 - b) Yetenek kestirim yöntemlerine (EOT ve BSD) göre nasıl değişmektedir?
- 2) Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı, sabit test uzunluğuna(40 madde) dayalı test durdurma kuralında; madde seçme yöntemlerine (MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler) ve yetenek kestirim yöntemlerine (EOT ve BSD) göre nasıl değişmektedir?

- 3) Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde seçmede kullanılan yöntemlere (MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler) ilişkin ortalama madde sayısı,
 - a) Test durdurma kuralı ($SH < 0.2$ ve $SH < 0.4$) ve
 - b) Yetenek kestirim yöntemlerine (EOT ve BSD) göre nasıl değişmektedir?
- 4) Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı, sabit standart hataya ($SH < 0.2$ ve $SH < 0.4$) dayalı test durdurma kuralında; madde seçmede kullanılan yöntemlere (MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler) ve yetenek kestirim yöntemlerine göre (EOT ve BSD) nasıl değişmektedir?

1.4. Sınırlılıklar

Araştırma,

- 1) MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler madde seçme yöntemleri
- 2) EOT ve BSD yetenek kestirim yöntemleri
- 3) Standart hataya dayalı test durdurma kuralı 0.2 ve 0.4 değerleri

İle sınırlandırılmıştır

1.5. Araştırmanın Amacı

Madde seçme yöntemi, BBT uygulamalarında merkezi bir unsurdur. BBT uygulamalarının nasıl başlayacağı, devam edeceği ve sonlanacağına madde seçme yöntemi hakim olmaktadır. Maksimum Fisher Bilgisi madde seçme yöntemi BBT uygulamalarında en çok kullanılan yöntem olmasına karşın; test uygulamalarının başında özellikle 10 ve daha az madde uygulandığında interim yetenek kestiriminde yetersiz kalmaktadır. Bir diğer problem ise MFB yüksek a parametre değerine sahip maddeleri seçme eğilimindedir ki bu durum madde

kullanım sıklığı sorununa yol açmaktadır ve dolayısıyla test güvenliğini düşürmektedir (Wainer,2000; Weiss,1984; Van der Linden ve Glas,2010). Bu problemlerin önüne geçmek için farklı madde seçme yöntemleri geliştirilmiştir. Eldeki araştırmada, a-tabakalama, OAB, AMBO ve Kullbak-Leibler madde seçme yöntemlerinin MFB madde seçme yönteminin kısıtlarını gidermede nasıl sonuç verdiği karşılaştırmalı olarak incelenmektedir. BBT uygulamalarında yetenek kestirimi ile madde seçme yöntemi arasında sıkı bir ilişki vardır. Madde seçme sürecinin anahtar noktası, cevaplayıcının yeteneği ile madde güçlüğünü eşleştirmektir. Şöyle ki; BBT uygulamalarında her madde cevaplandıktan sonra yetenek kestirimi yapılmaktadır ve bu yetenek kestiriminin sonucu madde seçiminde kullanılmaktadır. Yetenek kestirim yöntemlerinden EOT ve BSD araştırmaya dahil edilerek madde seçme yöntemlerini nasıl etkilediği belirlenmeye çalışılmıştır. BBT uygulamalarının başında (özellikle test uzunluğu beş maddeden küçük olduğunda) madde seçme yöntemlerinin yetersiz kaldığı yönünde araştırmalar mevcuttur (Linda,1996; Van der Linden,2000; Han,2009). Test uzunluğuna bağlı olarak madde seçme yöntemlerinin nasıl farklılaştığını belirlemek için farklı test uzunlukları (5, 10, 20, 30 ve 40 madde) da bir değişken olarak alınmıştır. Testi durdurma kuralının sabit standart hataya bağlı olarak belirlendiği BBT uygulamalarında madde seçme yöntemlerini karşılaştırmak için ise, standart hatanın 0.2 ve 0.4 olduğu koşullar oluşturulmuştur. Eldeki araştırmanın amacı yetenek kestirim yöntemi, sabit madde sayısı ve standart hataya dayalı durdurma kuralının madde seçme yöntemlerini nasıl etkilediğini belirlemektir.

BBT uygulamalarında bazı maddeler testi alan kişilere çok sıklıkla uygulandığında, bu maddelere ait bilgiler ezberlenebilir ve ileride testi alacak kişiler tarafından öğrenilebilir. Test puanları olumsuz etkilenir ve testin güvenliği sağlanamaz. Bu durum, madde seçme yöntemlerinin madde havuzu kullanımı açısından da ele alınması gerekliliğini doğurur (Sireci, 2003). Eldeki araştırmada da oluşturulan BBT koşullarının hepsinde madde seçme yöntemlerinin, madde havuzundaki maddeleri ne sıklıkta kullandığı araştırılmıştır.

1.6. Araştırmanın Önemi

Bilgisayar ortamında bireyselleştirilmiş test uygulamalarının psikometri alanında giderek kullanımı artmaktadır. Yapılan araştırmalarda madde seçme yöntemleri karşılaştırmalı olarak incelenmiştir ancak çalışmalarda çoğunlukla iki veya daha fazla madde seçme yöntemi karşılaştırılmıştır (Linda,1996, Weissman, 2003, Han, 2009).

BBT'in en önemli güdüsü (gerekçesi) etkili ölçümlerdir (Eggen,2004). Bu etkililiği sağlayan, her bir cevaplayıcının bireysel yeteneğine göre test edilmesidir. Testi alan cevaplayıcıların yetenekleri hakkında en çok bilgiyi taşıyan maddelerle karşılaşmasının yanında güçlük düzeyi bakımından uygun olmayan ve cevaplayıcının yeteneğinin kestirilmesinde az katkı sağlayan maddelerin de seçilmemesi gerekmektedir. Bu durum, BBT uygulamalarında madde seçme yöntemlerinin önemini ortaya koymaktadır.

BBT uygulamalarına getirilen önemli eleştirilenlerden birisi de test gizliliğinin düşmesi ihtimalidir. Bu araştırmayla madde havuzu kullanımında en iyi dengenin hangi madde seçme yöntemiyle elde edileceği araştırılmaktadır. Elde edilen sonuçlar, operasyonel BBT uygulamalarının düzenlenmesine katkı sağlaması bakımından önemlidir.

BBT uygulama teknikleri geliştirme ve uygulama konusu üzerinde yapılan araştırmaların arttığı, çeşitlendiği bir alan olurken, ülkemizde bu konunun akademik düzeyde ele alınması oldukça yeni ve sınırlı düzeydedir. Yapılan araştırmanın bu yönüyle de önemli olacağı ve alan yazına katkı sağlayacağı umulmaktadır.

1.7. İlgili Araştırmalar

Kaptan (1993), kağıt kalem test ile bireyselleştirilmiş bilgisayarlı testi yetenek kestirimi bakımından karşılaştırmıştır. Kağıt kalem testi 50 maddeden, bireyselleştirilmiş bilgisayarlı test ise 14 maddeden oluşturulmuştur. Yetenek kestiriminde EOT yetenek kestirim yöntemi kullanılmıştır. Her ik yöntemden elde edilen sonuçlarda anlamlı bir farklılık bulunmamıştır. Ancak BBT uygulamalarının test uzunluğunu kısalttığı sonucuna varılmıştır.

Linda (1996), yaptığı araştırmada BBT uygulamalarında, MFB madde seçme yöntemi ile Kullbak-Leibler madde seçme yöntemlerinin performanslarını karşılaştırmalı olarak incelemiştir. Araştırmada, TOEFL testinden alınan 500 maddelik gerçek bir data alınmıştır ve yine 500 maddelik simülatif bir data türetilmiştir. Kullbak-Leibler madde seçme yönteminin δ indeksinin farklı değerlerinde, MFB madde seçme yöntemine eşit ya da daha küçük yanlılığa sahip olduğu görülmüştür. θ yetenek düzeyinin -3, -2, -1 ve 2 olduğu durumlarda, Kullbak-Leibler madde seçme yönteminin ortalama standart hata(OSH) değeri, MFB ile elde edilen OSH değerinden daha küçük bulunmuştur. Simülatif datadan elde edilen analizlerde ise; $\theta = -3$ iken MFB madde seçme yöntemi, Kullbak-Leibler madde seçme yönteminden daha düşük yanlılık ve OSH göstermiştir. $\theta = 2$ iken, her iki yöntem için yanlılık ve OSH değerleri birbirine yakın bulunmuştur. Test uzunluğu ve yetenek düzeyleri birlikte incelendiğinde, Kullbak-Leibler madde seçme yönteminin; $-2 < \theta < 2$ yetenek düzeyinde ve test uzunluğunun 10 ile 30 madde sayısı aralığı boyunca MFB'dan daha düşük yanlılık ve OSH değerine sahip olduğu gözlenmiştir. Araştırmada, Kullbak-Leibler madde seçme yönteminde MFB'ya benzer ya da daha iyi sonuçlar elde edildiği sonucuna varılmıştır. Kullbak- Leibler madde seçme yönteminin farklı test uzunluklarında ve BBT uygulama koşullarında etkililiğinin araştırılması önerilmiştir.

Veerkamp ve Berger (1997) ABÖ ve OAB yöntemlerini MFB ile karşılaştırmıştır ve bu yöntemlerin MFB'ya olmazsa olmaz bir üstünlüğü olmadığı sonucuna varmışlardır.

Eggen (1999), BBT uygulamalarında Kullbak-Leibler ve MFB madde seçme yöntemlerini karşılaştırmıştır. Simülatif olarak gerçekleştirilen araştırma için 250 maddelik madde havuzu oluşturulmuştur. Test uzunluğu 40 madde olarak belirlenmiştir. Tahminin standart hatası ve bias değerleri hesaplanmıştır. Kullbak-Leibler madde seçme yönteminin, MFB madde seçme yönteminden daha iyi performans gösterdiği sonucuna varılmıştır.

Wen, Chang ve Hau (2000), MFB madde seçme yöntemi ile a-tabakalama madde seçme yöntemini farklı BBT koşullarında simülatif olarak karşılaştırmıştır. Eldeki araştırmayı ilgilendiren yöntem ve bulgular şöyledir. Ortalaması 0, standart sapması 1 olacak şekilde 5000 cevaplayıcı ve 400 maddeden oluşan bir madde havuzu oluşturulmuştur. a-tabakalama madde seçme yöntemi, her bir tabakada 100 madde ve a parametre değerleri sırasıyla 0.5, 1, 1.5 ve 2 olacak şekilde 4 tabakaya ayrılmıştır. Test uzunluğu 15 madde ve yetenek kestirimi EOT olarak belirlenmiştir. Sonuçların değerlendirilmesinde tahminin standart hatası hesaplanmıştır. MFB madde seçme yönteminin a-tabakalama madde seçme yönteminden daha etkili sonuçlar verdiği sonucuna varılmıştır. Ancak, a-tabakalama yönteminin madde havuzu kullanımında MFB'dan daha iyi sonuç verdiği görülmüştür. Özellikle BBT koşullarının başlangıcında a-tabakalama madde seçme yönteminin düşük a parametresine sahip maddeleri, test ilerledikçe daha yüksek a-parametresine sahip maddeleri kullandığını ve madde havuzu kullanımında MFB madde seçme yönteminden daha iyi bir dengeye sahip olduğu sonucuna varılmıştır.

İşeri (2002) Ortaöğretim Kurumları Seçme Yerleştirme Sınavı ve Özel Okullar Sınavının matematik kısımlarından oluşturduğu soru bankasını kullanarak matematik başarısının bilgisayar ortamında bireyselleştirilmiş test uygulamasıyla ölçülmesi konusundaki çalışmasında şu sonuçları elde etmiştir:

Bilgisayar ortamında bireyselleştirilmiş testler yetenek düzeyini daha az soru kullanarak güvenilir biçimde ölçmektedir. Bayesian yetenek kestirim yöntemi daha iyi kestirim sağlamıştır. Sabit madde sayısına bağlı durdurma kuralı ve sabit standart hataya bağlı durdurma yöntemlerinin her ikisinden de iyi sonuçlar elde edilmiştir.

Yi ve Chang (2003), *a*-tabakalama, *b* parametresine dayalı *a*-tabakalama ve MFB madde seçme yöntemlerini, belirli bir içeriğe bağlı denge ve madde kullanım sıklığı bakımından incelemiştir. Madde kullanım sıklığı yöntemi olarak Sympson-Hetter kullanılmıştır. *a*-tabakalama madde seçme yönteminin BSD yetenek kestirimi altında, madde havuzundaki maddelerin kullanım sıklığı indeksini düşürdüğü ve madde havuzu kullanımında denge sağladığı sonucuna varmışlardır.

Weissman (2003), BBT uygulamalarında madde seçme yöntemlerinin etkililiğini araştırmıştır. Araştırma simülatif olarak gerçekleştirilmiştir; iki farklı madde seçme yöntemi olan Maksimum Fisher Bilgisi (MFB) ve Maksimum Fisher Aralık Bilgisi (MFAB) yöntemlerinin etkililiği; En Çok Olabilirlik Tahmini, Modal Sonsal Dağılım (modal *a posteriori*), Golden Section Search, Alternatif En Çok Olabilirlik Tahmini ve Alternatif Modal Sonsal Dağılım olmak üzere beş farklı yetenek kestirim yöntemi; 5, 10, 15 ve 25 olmak üzere dört farklı madde uzunluğu durumları için karşılaştırılmıştır. Araştırmada kullanılan madde havuzu, matematik yeteneğini ölçen büyük ölçekli bir sınavdan elde edilen üç parametrelili lojistik model kullanılarak parametreleri belirlenmiş 367 çoktan seçmeli test maddesinden oluşturulmuştur. Deneysel çalışmada dört faktör ele alınmıştır. (1) madde seçme yöntemi (MFB ve MFAB) (2) yetenek kestirim yöntemi (EOT, MSD, GSS, EOT/Alt, MSD/Alt) (3) yetenek düzeyi ($\theta = \{-2, -1, 0, +1, +2\}$) ve (4) test uzunluğu (5, 10, 15 ve 25 madde).

ML yetenek kestirimi kullanıldığında MFB madde seçme yöntemine ait etkililik orta yetenek düzeylerinde ($\theta = \{-1, 0, +1\}$) ve bütün madde uzunluklarında (5, 10, 15 ve 25) MSD yetenek kestirimi kullanıldığında elde edilen etkililikten daha

yüksek bulunmuştur. EOT ve MSD yetenek kestirimlerine dayalı MFB madde seçme yöntemleri içinde en iyi performans; orta seviyedeki yetenek düzeylerinde MSD yetenek kestirimi kullanılarak yapılan madde seçme yönteminde gözlenmiştir. Etkililikteki en büyük fark, uç yetenek düzeylerinde MSD yetenek kestirimi ile yapılan MFB madde seçme yöntemi ile bütün yetenek düzeylerinde EOT yetenek kestirimi ile yapılan MFB madde seçme yönteminde bulunmuştur. EOT ve MSD yetenek kestirimlerine dayalı olarak yapılan MFAB madde seçme yönteminin etkililik farkı; düşük yetenek düzeylerinde, 5 ve 10 maddelik test uzunluklarında daralmıştır. Test etkililiğindeki en yüksek artış, EOT/Alt, MSD/Alt ve GSS yetenek kestirim yöntemleri kullanıldığında elde edilmiştir. Etkililikteki kazanç daha çok, test uzunluğu kısa olduğu durumlarda belirginleşmiştir. Araştırmanın sonunda yetenek kestirim yönteminin, madde seçmenin etkililiğini, madde seçme yönteminden daha çok etkilediği sonucuna varılmıştır. Madde seçme yöntemi ve yetenek kestirim yöntemi arasındaki ilişkinin daha geniş ölçekli uygulamalarla değerlendirilmesini önermişlerdir. Ayrıca, madde seçme yöntemlerinin etkililiğinin reel bir BBT uygulaması yapılarak değerlendirilmesini araştırmacılara tavsiye edilmiştir.

Han (2009) yaptığı araştırmada BBT uygulamasında random seçimli MFB, fade-away seçimli MFB, AMBO ve fade-away seçimli AMBO madde seçme yöntemlerinin etkililiğini simülatif olarak incelemiştir. Madde havuzu, GMAT sınavından elde edilen 500 maddelik çoktan seçmeli test maddesinden oluşturulmuştur. Dağılımlarının ortalaması 0, standart sapması 1 olan 10.000 kişilik cevaplayıcının her gün iki kere olmak üzere 20 gün süren BBT uygulaması simüle edilmiştir. Bir BBT uygulamasında 250 cevaplayıcıya 40 madde uzunluğunda bir test uygulanmıştır. Araştırma sonuçlarının değerlendirilmesi için, madde kullanım sıklığı oranı, test bilgisi, madde havuzu indeksi, θ yetenek kestiriminin yanlılığı ve hataları hesaplanmıştır. Beş madde seçme yönteminin değerlendirilmesi iki önemli noktada gerçekleşmiştir (1) θ yetenek kestiriminin performansı ve (2) madde havuzu kullanımı. θ yetenek kestirimin standart hatası; ayrıca, yanlılık ve ortalama mutlak hatalar da hesaplanmıştır. Hangi madde seçme yönteminin optimal madde havuzu

kullanımı gösterdiğini değerlendirmek için madde kullanım sıklık oranları analiz edilmiştir. Araştırmanın sonunda MFB ve AMBO madde seçme yöntemlerinin theta ölçüsü boyunca en düşük SH ile, random seçimli madde seçme yönteminin ise en yüksek SH ile sonuçlandığı görülmüştür. Tahminin yanlılıkları değerlendirildiğinde, madde seçme yöntemleri arasında anlamlı bir farklılık olmadığı görülmüştür. Madde havuzu kullanımının etkililiği için 500 maddenin, her bir madde seçme yönteminde kullanılma sıklığı oranları hesaplanmıştır. Random metodunun hiçbir maddeyi 0.2 madde kullanım sıklığı oranından yüksek olacak şekilde kullanmadığı görülmüştür. MFB metodunun maddelerin kullanım dengesinde en yüksek uyumsuzluk gösterdiği; yani, çok büyük bir madde grubu 0.2 oranında toplanırken, çok büyük bir madde grubunun da hiç kullanılmadığı gözlenmiştir. AMBO madde seçme yönteminin MFB yöntemine benzer bir madde havuzu kullanımı sergilediği bulunmuştur. En iyi balansın, fade away metoduyla gerçekleştirilen AMBO madde seçme yönteminde gerçekleştiği görülmüştür. Bu yöntemle, ne en yüksek madde kullanım sıklığı ne de hiç kullanılmayan maddeler gözlenmiştir, dengeli bir dağılım elde edilmiştir. Madde havuzunun yaklaşık %90'ı, 0.12 madde kullanım sıklığı oranının altında yer almıştır. Bu durum, BBT uygulamalarında aynı maddelerin sıklıkla yer almasının ve cevaplayıcıların aynı maddelerle karşılaşarak test güvenliğinin tehlikeye girmesinin önüne geçmektedir. MFB madde seçme yönteminin BBT uygulamalarında yaygın kullanılan bir yöntem olmasına karşın, bu çalışmada; AMBO madde seçme yöntemi BBT uygulamaların başında daha düşük SH göstermiştir ve madde havuzu kullanımında da daha iyi bir denge ortaya koymuştur. Araştırmacı, AMBO ve MFB madde seçme yöntemlerinin farklı test koşulları altında karşılaştırılmasını gelecek araştırmalar için önermiştir.

Costa, Karino, Moura ve Andrade (2009) yaptıkları çalışmada BBT uygulamalarında MFB, Kullbak-Leibler Bilgisi ve Maksimum Beklenen Bilgi (MBB) madde seçme yöntemlerinin performansını değerlendirmişlerdir. Bu üç yöntemi karşılaştırmak için Brasiliyas Üniversitesi İngilizce Testi'nden alınan 246 maddelik bir havuz oluşturularak beş farklı simülasyon çalışması yapılmıştır. Yetenek kestirim yöntemi olarak BSD kullanılmıştır. Madde seçme

yöntemlerinin performanslarını değerlendirmek için yanlılık ve hata karesi ortalaması hesaplanmıştır. Birinci simülasyon çalışması testi durdurma kuralı hatanın 0.4 ve 0.2 olduğu durumlar için yapılmıştır ve kaç maddeye ihtiyaç duyulduğu belirlenmeye çalışılmıştır. Durdurma kuralı 0.4 olduğunda her üç yöntemin de 20 maddeye; 0.2 olduğunda ise 40'dan fazla maddeye ihtiyaç duyduğu bulunmuştur. İkinci simülasyon çalışmasında, -4 ve 4 arasında değişen θ düzeylerinde, test uzunluğu 25 madde ve başlangıç düzeyi 0 olan bir BBT uygulaması gerçekleştirilmiştir. Her üç yöntemin, 25 maddelik bir test için θ 'yı benzer etkililikte kestirmiştir. Üçüncü simülasyon çalışması, test başlama yetenek düzeyi $\theta_0 = -1.50; \theta_0 = 0.0; \theta_0 = 1.50$ değerleri seçilerek gerçekleştirilmiştir ve madde seçme yöntemleri karşılaştırılmıştır. $\theta_0 = -1.50$ iken her üç yöntemde de standart hatanın %50'si 0.28'e eşit veya daha küçük bulunmuştur. $\theta_0 = 0.0$ iken MBB madde seçme yöntemi en yüksek ranja sahip bulunmuştur. Standart hatalar ise her üç yöntem için 0.24 ile 0.30 arasında değişmiştir. $\theta_0 = 1.50$ iken her üç yöntem gerçek θ değerine yakın kestirimler ortaya koymuştur. Beşinci simülasyon çalışmasında on farklı θ değeri için yanlılık ve ortalama standart hata 30 maddelik bir testten hesaplanmıştır. Her üç madde seçme yönteminde de test uzunluğu arttığında yanlılık ve ortalama standart hata değerlerinin düştüğü görülmüştür. Yapılan analizler sonucunda araştırmaya konu edilen madde seçme yöntemlerinden benzer sonuçlar elde edildiği, bir farklılık olmadığı sonucuna varılmıştır. Araştırmacılar, çalışmada kullandıkları madde havuzunun yetersiz kaldığını ve sonuçların bundan etkilenmiş olabileceğini belirtmişlerdir ve daha geniş madde havuzunda araştırma yapılmasını; ayrıca, madde kullanım sıklığı oranı, testi durdurma kuralları vb özelliklerin madde seçme yöntemleri ile ilişkisinin araştırılmasını önermişlerdir.

Deng, Ansley ve Chang (2010), MFB, a-tabakalama ve refine edilmiş a-tabakalama madde seçme yöntemlerini simülatif olarak karşılaştırmışlardır. Refine edilmiş a-tabakalama madde seçme yöntemi, yüksek a parametre değerine sahip tabakadan daha çok madde seçilmesi, düşük a parametre

değerine sahip tabakadan ise daha az madde seçilmesine dayanmaktadır. Araştırma kapsamında 300 maddelik bir madde havuzu ve 3000 cevaplayıcı simüle edilmiştir. Test uzunluğu 15 Araştırmada elde edilen bulguların yorumlanması için tahminin standart hatası, yetenek kestirimlerinin güvenilirliği ve madde havuzu kullanım düzeyleri hesaplanmıştır. Araştırmada, MFB madde seçme yönteminin diğer iki yönteme nazaran yetenek kestiriminde daha etkili olduğu ancak madde havuzu kullanım düzeyinde yetersiz kaldığı sonucuna varılmıştır.

Han (2010), BBT koşullarında kullanılan 5 madde seçme yöntemi, a-tabakalama, Aralık Bilgisi Ölçütü (ABÖ), Olabilirlik Ağırlıklı Bilgi Ölçütü (OAB), Kullbak-Leibler Bilgisi ve Aşamalı Maksimum Bilgi Oranı (AMBO) yöntemlerini karşılaştırmıştır. Araştırmada, madde havuzunun kullanımında etkililik ve performans dengesini en iyi sağlayan madde seçme yöntemi saptanmaya çalışılmıştır. Han yaptığı araştırmada MFB ve 5 farklı madde seçme yönteminin karşılaştırmasını simülatif olarak gerçekleştirmiştir. Simülasyon çalışmalarında SimulCAT programı kullanılmıştır. Madde havuzu Graduate Management Admission Test (GMAT)'de yer alan çoktan seçmeli 500 maddeden oluşturulmuştur. Test uzunluğu 10, 20 ve 40 madde olmak üzere üç farklı durumda; testi alan cevaplayıcıların sayısı ise 80.000, 40.000 ve 20.000 olarak simüle edilmiştir. Madde seçme yöntemlerinin değerlendirilmesi iki önemli noktada gerçekleştirilmiştir 1) yetenek kestiriminin performansı ve 2) madde havuzu kullanımı. Araştırmanın sonucunda; test uzunluğundaki artışa bağlı olarak bütün madde seçme yöntemlerinde SH değerlerinde düşüş gözlenmiştir. MFB, KL ve AMBO yöntemleri için diğer üç yönteme kıyasla düşük SH değeri hesaplanmıştır. a-tabakalama madde seçme yönteminin θ alanlarının çoğunda en yüksek SH değerine sahip olduğu görülmüştür. OAB yöntemi, düşük θ alanlarında yüksek SH değerine sahip olurken $\theta > -0.5$ olduğunda daha düşük SH göstermiştir. Yanlılıklar ise, madde seçme yöntemlerinin hepsinde test uzunluğu arttıkça daha düşük sonuçlar göstermiştir. Genel olarak; MFB, Kullbak-Leibler Bilgisi ve AMBO madde seçme yöntemlerinden özellikle test

uzunluğunun daha kısa olduğu durumlarda diğer madde seçme yöntemlerine göre daha iyi sonuçlar elde edildiği görülmüştür.

Kalender (2011), BBT yöntemi ile elde edilen yetenek kestirimlerini farklı yetenek kestirim ve test sonlandırma kurallarını dikkate alarak Öğrenci Seçme Sınavı (ÖSS) fen alt testinin kağıt kalem formatı sonuçlarını karşılaştırmalı olarak incelemiştir. Yetenek kestirim yöntemleri olarak; En Çok Olabilirlik Tahmini (EOT) ve Beklenen Sonsal Dağılım (BSD); madde seçme yöntemi olarak MFB kullanılmıştır. Test sonlandırma kuralları ise standart hata eşik değeri ve sabit soru sayısı olarak belirlenmiştir. 242 maddelik bir madde havuzu oluşturulmuştur. Eldeki araştırmayı ilgilendiren bulgular şöyledir. BSD yetenek kestirimi, sabit soru sayısına bağlı durdurma kuralının uygulandığı BBT koşullarında EOT'den daha iyi sonuç vermiştir. Standart hata eşik değerine bağlı durdurma kuralı uygulandığında, EOT yetenek kestiriminin kullanıldığı BBT koşulları için gerekli ortalama madde sayısı, BSD yetenek kestiriminin kullanıldığı koşullardan daha az bulunmuştur. Standart hata eşik değerinin 0.10 olarak belirlendiği BBT koşullarında ise gerekli madde sayısı ortalamasında kağıt kalem testlerine nazaran belirgin bir azalma gözlenmemiştir.

Yapılan araştırmalar incelendiğinde, BBT uygulamalarında kullanılan madde seçme yöntemlerinin farklı koşullarda güçlü ve zayıf yönlerinin olduğu belirlenmiştir. Araştırmalarda (Deng, Ansley ve Chang, 2010; Wen, Chang, Hau, 2000; Eggen, 2009; Yui ve Chang, 2003) genellikle iki madde seçme yöntemi karşılaştırmalı olarak ele alınmıştır. İki'den çok madde seçme yöntemini inceleyen araştırmalarda (Han, 2010; Weissman, 2003) ise durdurma kuralları ve yetenek kestirim yöntemleri bir arada ele alınmamıştır. Türkiye'de yapılan araştırmalarda (İşeri, 2002; Kalender, 2011; Kaptan, 1993) ise kağıt kalem testleri ile BBT uygulamaları karşılaştırmalı olarak incelenmiştir. Yetenek kestirim yöntemleri EOT ve BSD ele alınırken, durdurma kuralları da bir değişken olarak incelenmiştir.

2. BÖLÜM

YÖNTEM

Bu bölümde araştırmanın türü, verinin elde edilmesi ve verinin analizi açıklanmıştır.

2.1. Araştırmanın Türü

Bu araştırmada, Bireyselleştirilmiş Bilgisayarlı Test uygulamalarında kullanılan madde seçme yöntemleri, test durdurma kuralları ve yetenek kestirim yöntemlerine göre incelenmiştir. Bu yönüyle çalışma, temel araştırma niteliğindedir.

2.2. Verinin Elde Edilmesi

Araştırmanın verisi, Han (2009) tarafından geliştirilen SimulCAT bilgisayar programıyla simülatif olarak elde edilmiştir. Veri elde etme aşamalarında öncelikle araştırmanın yürütüldüğü grup, daha sonra madde havuzu ve BBT koşulları oluşturulmuştur.

2.2.1. Araştırmanın Yürütüldüğü Grup

BBT uygulamasına katılan cevaplayıcılar araştırmacı tarafından SimulCAT bilgisayar programıyla simüle edilmiştir. Bu amaçla, ortalaması 0, standart sapması 1 olacak şekilde 2000 kişi oluşturulmuştur. Araştırmanın yürütüldüğü gruba ait özellikler Ek.1'de sunulmuştur.

2.2.2. Madde Havuzu

BBT uygulamaları için gerekli olan madde havuzu, SimulCAT bilgisayar programıyla üç parametrelili lojistik modele göre 250 maddeden oluşturulmuştur. Madde parametreleri uniform olarak araştırmacı tarafından belirlenmiştir. a parametresi için minimum 0.80, maksimum 1.5; b parametresi için minimum -3 , maksimum 3 ve c parametresi için minimum 0.05, maksimum 0.15 değerleri saptanmıştır. Bilgisayar programı, madde parametre değerlerine göre madde havuzunu oluşturmuştur, elde edilen madde havuzu özellikleri Tablo 2’de sunulmuştur.

Tablo 2: Madde Havuzu Betimsel İstatistikleri

Madde Parametreleri	Minimum	Maksimum	Ortalama	Standart sapma
a	0.80	1.50	1.15	0.22
b	-3	3	0.14	1.73
c	0.05	0.15	0.09	0.03

Madde havuzunda yer alan her bir maddenin a, b ve c parametre değerleri Ek.3’de verilmiştir.

2.2.3. BBT Koşulları

BBT koşulları, madde seçme yöntemleri, yetenek kestirim yöntemleri ve durdurma kurallarından oluşturulmuştur. Araştırmaya alınan 5 madde seçme yöntemi, 2 yetenek kestirim yöntemi ve 3 durdurma kuralı için toplam 30 koşul belirlenmiştir. BBT koşulları Tablo 3’de sunulmuştur. Tasarlanan bütün koşullar için 30 farklı simülasyon çalışması gerçekleştirilmiştir.

Tablo 3: BBT Koşullarının Özellikleri

Koşul	Madde Seçme Yöntemi	Yetenek Kestirim Yöntemi	Test Durdurma Kuralı
Koşul 1	MFB	EOT	40 madde
Koşul 2	a-tabakalama	EOT	40 madde
Koşul 3	OAB	EOT	40 madde
Koşul 4	AMBO	EOT	40 madde
Koşul 5	Kullbak-Leibler	EOT	40 madde
Koşul 6	MFB	BSD	40 madde
Koşul 7	a-tabakalama	BSD	40 madde
Koşul 8	OAB	BSD	40 madde
Koşul 9	AMBO	BSD	40 madde
Koşul 10	Kullbak-Leibler	BSD	40 madde
Koşul 11	MFB	EOT	SH<0.2
Koşul 12	a-tabakalama	EOT	SH<0.2
Koşul 13	OAB	EOT	SH<0.2
Koşul 14	AMBO	EOT	SH<0.2
Koşul 15	Kullbak-Leibler	EOT	SH<0.2
Koşul 16	MFB	BSD	SH<0.2
Koşul 17	a-tabakalama	BSD	SH<0.2
Koşul 18	OAB	BSD	SH<0.2
Koşul 19	AMBO	BSD	SH<0.2
Koşul 20	Kullbak-Leibler	BSD	SH<0.2
Koşul 21	MFB	EOT	SH<0.4
Koşul 22	a-tabakalama	EOT	SH<0.4
Koşul 23	OAB	EOT	SH<0.4
Koşul 24	AMBO	EOT	SH<0.4
Koşul 25	Kullbak-Leibler	EOT	SH<0.4
Koşul 26	MFB	BSD	SH<0.4
Koşul 27	a-tabakalama	BSD	SH<0.4
Koşul 28	OAB	BSD	SH<0.4
Koşul 29	AMBO	BSD	SH<0.4
Koşul 30	Kullbak-Leibler	BSD	SH<0.4

2.2.3.1. Simülasyon Çalışmaları

BBT koşulları, SimulCAT bilgisayar programıyla simüle edilmiştir. Simülasyon çalışmaları üç adımda gerçekleştirilmiştir (Ek.2). Birinci adımda, araştırmacı tarafından oluşturulan cevaplayıcı ve madde havuzu dosyaları bilgisayar programına yüklenmiştir. İkinci adımda madde seçme yöntemleri ve test durdurma kuralları; üçüncü adımda ise, yetenek kestirim yöntemleri, test başlatma kuralı, replikasyon sayısı ve çıktı dosyaları seçilmiştir. Test başlatma kuralı $\theta=0.5$ olarak belirlenmiştir ve bütün simülasyon çalışmaları için 100 replikasyon yapılmıştır.

2.3. Verilerin Analizi

BBT koşullarının simülasyonundan elde edilen veriler, bilgisayar programı sonuç raporu olarak iki dosyada toplanmıştır. Birinci dosyada, cevaplayıcılara ait interim ve final θ ve SH değerleri; ikinci dosyada, madde havuzunda yer alan maddeler ve kullanım sıklıkları yer almıştır.

2.3.1. Durdurma Kuralına Dayalı Verilerin Analizi

Test durdurma kuralı 40 madde olarak belirlenen BBT koşullarında, test uzunluğuna göre madde seçme yöntemlerinin nasıl farklılaştığını bulmak amacıyla 5, 10, 20, 30 ve 40 madde için interim θ ve tahminin standart hatası(SH) hesaplanmıştır. Tahminin standart hatası,

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}}$$

eşitliği ile hesaplanmıştır.

Test durdurma kuralı $SH < 0.2$ ve $SH < 0.4$ olan BBT koşullarında, madde seçme yöntemleri, ortalama madde sayısına göre değerlendirilmiştir. Şöyle ki, BBT uygulaması belirli standart hata değerinde sonlanacağından, madde seçme yöntemlerinin bu standart hata değerine ulaşıldığında kullandıkları ortalama madde sayıları araştırılmıştır.

2.3.2. Madde Kullanım Sıklığına Dayalı Verilerin Analizi

Madde kullanım sıklığı, BBT koşullarında yer alan maddelerin kullanılma düzeyini ifade etmektedir. İyi bir madde seçme yönteminin, madde havuzundaki maddeleri eşit düzeyde kullanması beklenmektedir. Ayrıca, MFB madde seçme yönteminin yüksek a parametresine sahip maddeleri daha çok kullandığı pek çok araştırmada bu madde seçme yönteminin bir sınırlılığı olarak belirtilmiştir (Van der Linden ve Glas, 2010; Weiss, 1984; Wainer, 2000). Madde havuzunda yer alan maddeler a parametre değerine göre sıralanmıştır ve madde kullanım sıklıkları bar grafikleri ile gösterilmiştir. Madde seçme yöntemleri madde kullanma sıklıkları bakımından karşılaştırılmış; aynı zamanda, MFB madde seçme yönteminin sınırlılığı olan yüksek a -parametresine sahip maddelerin sıklıkla kullanılması sorunun diğer madde seçme yöntemlerinde ne ölçüde giderildiği değerlendirilmiştir.

3. BÖLÜM

BULGULAR VE YORUM

Yapılan analizlerinden elde edilen bulguların açıklanması, araştırmanın alt problemleri sırasınca sunulmuştur.

3.1 Birinci Alt Probleme İlişkin Bulgular

Araştırmanın birinci problemi şu şekilde belirlenmiştir:

“Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde seçmede kullanılan yöntemlere (MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler) ilişkin standart hatalar,

- a) Test uzunluğuna (5, 10, 20, 30 ve 40 madde) ve
- b) Yetenek kestirim yöntemlerine (EOT ve BSD) göre nasıl değişmektedir?”

Bu problemin analizi için, BBT durdurma kuralı 40 madde olarak belirlenmiştir. BBT uygulaması sonunda; beşinci, onuncu, yirminci, otuzuncu ve kırkıncı madde uygulandıktan sonra elde edilen interim yetenek kestirimlerinin ortalamaları ($\hat{\theta}$) ve tahminin standart hatası ($SH\theta$) bulguların analizinde kullanılmıştır. Madde seçme yöntemleri $SH\theta$ değerlerine göre karşılaştırılmıştır. Elde edilen bulgular, yetenek kestirim yöntemlerine göre sunulmuştur. EOT yetenek kestiriminin kullanıldığı BBT koşulları için elde edilen sonuçlar Tablo 4’de verilmiştir.

Tablo 4: Test durdurma kuralının 40 madde olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında, test uzunluğuna göre madde seçme yöntemlerine ait istatistikler

Madde Seçme Yöntemi	Test Uzunluğu									
	5		10		20		30		40	
	$\hat{\theta}$	SH	$\hat{\theta}$	SH	$\hat{\theta}$	SH	$\hat{\theta}$	SH	$\hat{\theta}$	SH
MFB	0.12	0.55	0.05	0.36	0.03	0.25	0.02	0.20	0.02	0.18
a-tabakalama	-1.55	0.78	-1.57	0.52	-1.39	0.31	-1.29	0.23	-1.19	0.19
OAB	-0.60	0.74	-0.28	0.38	-0.11	0.25	-0.62	0.21	-0.04	0.18
AMBO	-1.52	0.50	-1.28	0.35	-1.26	0.25	-1.06	0.21	-0.68	0.19
Kullbak-Leibler	-1.6	0.67	-1.20	0.37	-1.10	0.25	-0.57	0.22	-0.21	0.22

Tablo 4 incelendiğinde; test uzunluğu 5 madde olduğunda, en düşük SH değeri AMBO madde seçme yönteminden (0.50), en yüksek SH değeri a-tabakalama madde seçme yönteminden (0.78) elde edilmiştir. Test uzunluğu 10 madde iken, en düşük ve en yüksek SH değerleri sırasıyla AMBO (0.35) ve a-tabakalama (0.52) madde seçme yöntemlerinde saptanmıştır. Test uzunluğu 20 madde iken, en yüksek SH değeri a-tabakalama madde seçme yönteminden (0.31) elde edilirken, diğer madde seçme yöntemlerinin SH değerleri birbirine eşit (0.25) olarak hesaplanmıştır. Test uzunluğu 30 madde olduğunda da en yüksek SH değeri yine a-tabakalama madde seçme yönteminde (0.23); en düşük SH değeri ise OAB (0.21) ve AMBO (0.21) madde seçme yöntemlerinde belirlenmiştir. Test uzunluğunun 40 madde olduğu durumda, en düşük SH değeri OAB (0.18) ve MFB (0.18); en yüksek SH değeri Kullbak-Leibler (0.22) madde seçme yönteminden elde edilmiştir. Elde edilen bulgular incelendiğinde, test uzunluğunun 30 maddeden küçük olduğu durumlarda ($n < 30$) a-tabakalama

madde seçme yönteminin; test uzunluğunun otuz maddeden büyük olduğu durumlarda ise ($n > 30$) Kullbak-Leibler madde seçme yönteminin yüksek SH değeri gösterdiği görülmektedir. En yüksek SH değerinin a-tabakalama madde seçme yönteminden elde edilmesi, Han(2009)'ın yaptığı araştırmanın sonuçlarıyla paralellik gösterirken, Linda(1996)'nın Kullbak-Leibler madde seçme yönteminin MFB madde seçme yönteminden daha iyi sonuç verdiğini belirttiği araştırmasından farklılık göstermiştir.

Test uzunluklarına göre bütün madde seçme yöntemleri ele alındığında; 5 madde sonunda madde seçme yöntemlerinin SH değerleri arasında büyük farklılıklar olduğu, ancak 10 maddeden sonra SH değerleri farkının azaldığı ve birbirine yakın sonuçlar elde edildiği belirlenmiştir. Bu bulguya, interim θ değerlerinde rastlanmamıştır. BBT uygulamalarının başında ($n < 5$) MFB madde seçme yönteminin yetenek kestiriminde yetersiz kalması sınırlığı incelendiğinde ise; sadece AMBO madde seçme yönteminin MFB'den daha düşük SH değeri gösterdiği bulunmuştur. Elde edilen bu iki bulgu, araştırmaya alınan bütün madde seçme yöntemlerinin, BBT uygulamalarının başında yetenek kestiriminde zayıf olduklarını ve MFB madde seçme yöntemine belirgin üstünlükleri olmadığı şeklinde yorumlanmıştır.

BSD yetenek kestiriminin kullanıldığı BBT koşullardan elde edilen bulgular Tablo 5'de sunulmuştur.

Tablo 5: Test durdurma kuralının 40 madde olarak belirlendiği ve BSD yetenek kestiriminin kullanıldığı BBT koşullarında, test uzunluğuna göre madde seçme yöntemlerine ait istatistikler

Madde Seçme Yöntemi	Test Uzunluğu									
	5		10		20		30		40	
	$\hat{\theta}$	SH	$\hat{\theta}$	SH	$\hat{\theta}$	SH	$\hat{\theta}$	SH	$\hat{\theta}$	SH
MFB	0.01	0.47	0.02	0.33	0.02	0.23	0.02	0.20	0.02	0.18
a-tabakalama	0.01	0.70	0.01	0.49	0.02	0.31	0.02	0.23	0.02	0.18
OAB	0.01	0.55	0.02	0.35	0.02	0.24	0.02	0.20	0.02	0.18
AMBO	0.01	0.49	0.01	0.33	0.02	0.24	0.02	0.20	0.02	0.18
Kullbak-Leibler	0.01	0.47	0.02	0.33	0.02	0.24	0.02	0.20	0.02	0.18

Tablo 5 incelendiğinde; test uzunluğu 5 madde iken, en düşük SH değeri Kullbak-Leibler (0.47) ve MFB (0.47) madde seçme yönteminden, en yüksek SH değeri a-tabakalama madde seçme yönteminden (0.70) elde edilmiştir. Test uzunluğu 10 madde iken, Kullbak-Leibler (0.33), AMBO (0.33) ve MFB (0.33) madde seçme yöntemlerinin SH değeri en düşük; a-tabakalama (0.49) madde seçme yönteminin SH değeri en yüksek bulunmuştur. Test uzunluğu 20 madde olduğunda da benzer durum gözlenmiştir. Test uzunluğu 30 madde olduğunda en yüksek SH değeri a-tabakalama (0.23) madde seçme yönteminden elde edilirken diğer dört yöntemin SH değerleri (0.20) birbirine eşit bulunmuştur. Test uzunluğu 40 madde olduğunda bütün madde seçme yöntemlerinin SH değerleri (0.18) birbirine eşit olarak hesaplanmıştır. Görüldüğü üzere, tüm test uzunluklarında a-tabakalama madde seçme yöntemi en yüksek SH değerine sahip olmuştur. Elde edilen bulgular değerlendirildiğinde, test uzunluğu 5 madde iken madde seçme yöntemlerinin SH değerleri arasında büyük farklılıklar olduğu, ancak test uzunluğunun 10 madde ve üstü olarak belirlendiği

BBT kořullarda SH deęerleri farkının azaldığı ve birbirine yakın sonuçlar elde edildięi belirlenmiştir. Öyle ki, test uzunluğu 40 maddeye ulaştığında madde seçme yöntemlerinin SH deęerleri birbirine eşit olmuştur. BBT uygulamalarının başında ($n < 5$) bütün madde seçme yöntemlerinde görülen bu belirgin düşüş, MFB madde seçme yönteminin yetenek kestirimi bakımından yetersiz kalması sorununu gidermede Kullbak-Leibler madde seçme yöntemi hariç dięer madde seçme yöntemlerinin belirgin bir üstünlüğü olmadığı şeklinde yorumlanmıştır. Interim θ deęerleri incelendiğinde ise; BBT uygulamalarının başından itibaren bütün madde seçme yöntemlerinden elde edilen deęerler birbirine eşit bulunmuştur.

EOT ve BSD yetenek kestirim yöntemlerine göre madde seçme yöntemleri incelendiğinde; EOT ve BSD yetenek kestirim yöntemlerinin her ikisinde de en yüksek SH deęeri a-tabakalama madde seçme yönteminden elde edilmiştir. Genel olarak EOT yetenek kestirimi kullanıldığında elde edilen SH deęerleri, BSD yetenek kestirimi kullanıldığında elde edilen SH deęerlerinden daha yüksek bulunmuştur. Bu farklılık en belirgin test uzunluğu 5 madde olduğunda saptanmıştır. Örneğin Kullbak-Leibler madde seçme yönteminin EOT yetenek kestiriminde SH deęeri 0.67 iken, BSD yetenek kestiriminde SH deęeri 0.47 olarak hesaplanmıştır. Wang ve Visposel (1998), yetenek kestirim yöntemlerini karşılaştırdıkları arařtırmalarında BSD yetenek kestiriminin daha düşük SH deęeri gösterdiğini bulmuşlardır, Lord (1986) BSD yetenek kestiriminin EOT yetenek kestiriminden daha iyi sonuçlar verdięini belirtmiştir. Eldeki arařtırmada elde edilen bulgular da bu sonuçla paralellik göstermektedir. Elde edilen bu bulgu, BBT uygulamalarında özellikle testin başlangıcında BSD yetenek kestirim yönteminin öncelikle tercih edilmesi gerektięi şeklinde yorumlanmıştır.

EOT ve BSD yetenek kestiriminin kullanıldığı durumların her ikisinde de test uzunluğu 5 maddeden 10 maddeye ulaştığında SH deęerlerinde belirgin bir düşüş gözlenmiştir. Örneğin; EOT yetenek kestirimi kullanıldığında a-tabakalama madde seçme yönteminin SH deęeri 0.78'den 0.52 deęerine; BSD yetenek kestirimi kullanıldığında 0.70'den 0.49 deęerine düşmüştür. 10, 20, 30

ve 40 maddelik test uzunluklarında ise bu denli belirgin bir düşüşe rastlanmamıştır.

Farklı test uzunluklarında cevaplayıcıların interim theta değerleri incelendiğinde; BSD yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerinin interim θ değerleri birbirine yakın sonuçlar verirken; EOT yetenek kestiriminin kullanıldığı koşullarda, madde seçme yöntemlerinin interim θ değerleri birbirinden farklı bulunmuştur.

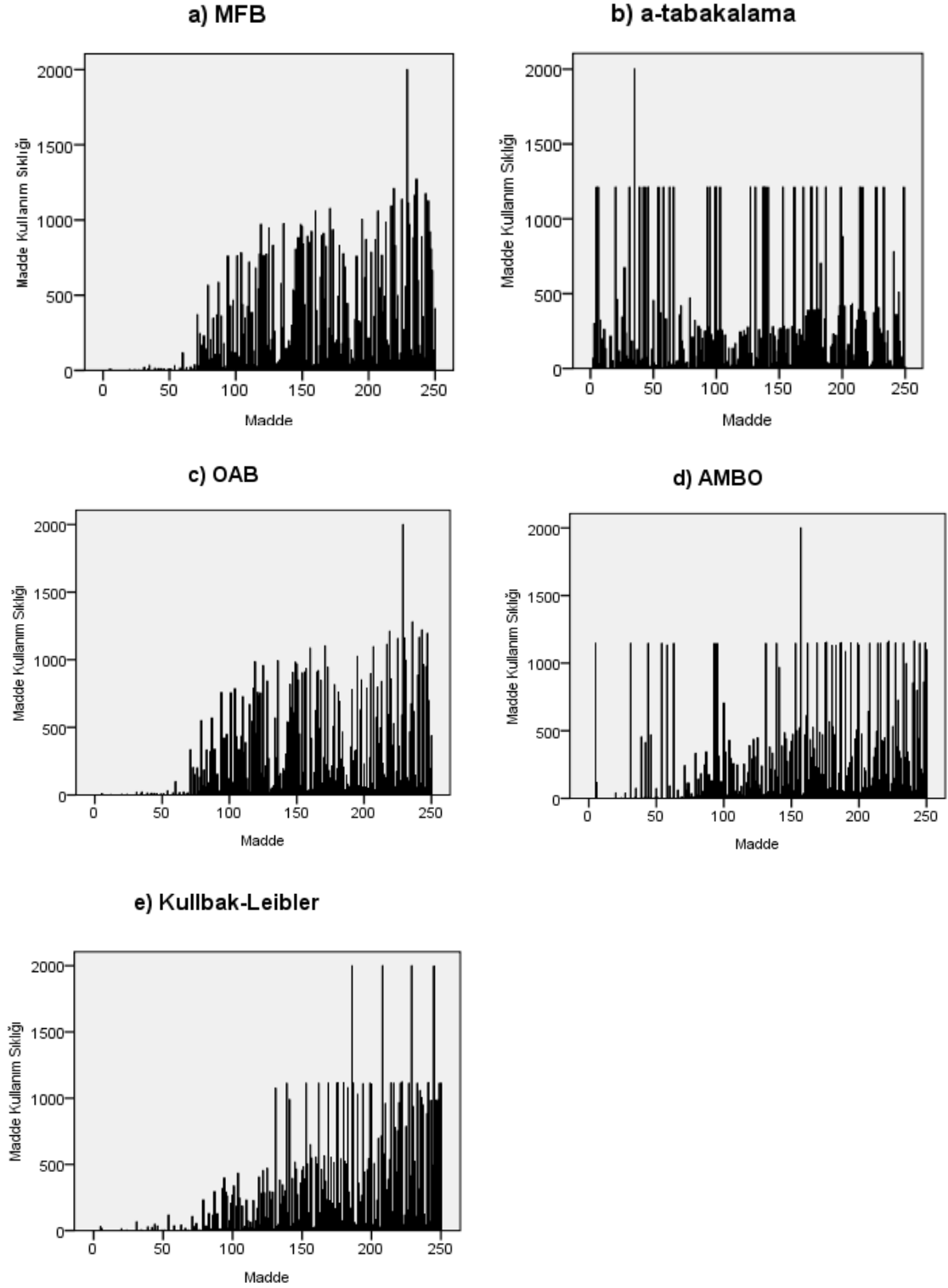
3.2. İkinci Alt Probleme İlişkin Bulgular

Araştırmanın ikinci alt problemi şöyle belirlenmiştir:

“Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı, sabit test uzunluğuna(40 madde) dayalı test durdurma kuralında; madde seçme yöntemlerine (MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler) ve yetenek kestirim yöntemlerine (EOT ve BSD) göre nasıl değişmektedir?”

Bu problemi analiz etmek için madde havuzunda yer alan maddeler a parametre değerlerine göre küçükten büyüğe sıralanmıştır ve her bir maddenin kullanım sıklığı bar grafiğinde gösterilmiştir. Grafikte x ekseninde yer alan birinci madde en düşük a parametre, iki yüz ellinci madde en yüksek a parametre değerine sahip maddeyi göstermektedir. y ekseninde ise, maddelerin kullanılma frekansları yer almaktadır. Grafik, madde kullanım sıklıklarının yanısıra yüksek a parametre değerini kullanma bakımından da yorumlanmıştır. Elde edilen sonuçlar yetenek kestirim yöntemlerine göre sunulmuştur. EOT yetenek kestiriminin kullanıldığı BBT koşullarından elde edilen bulgular Şekil 2’de sunulmuştur.

Şekil 2: Test durdurma kuralının 40 madde olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları

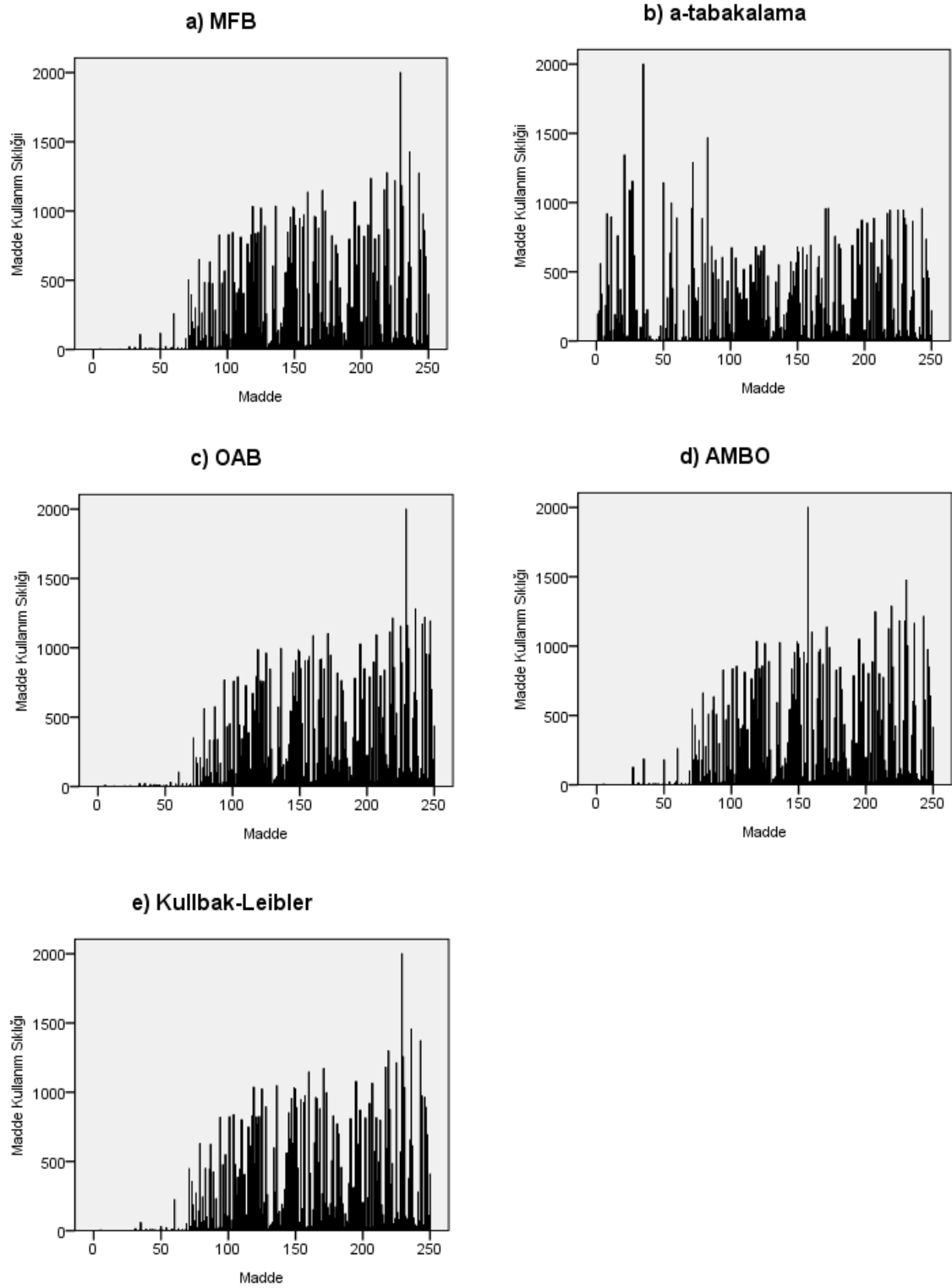


Şekil 2 incelendiğinde; MFB, OAB ve Kullbak-Leibler madde seçme yöntemlerinin yüksek α parametre değerine sahip maddeleri kullandıkları belirlenmiştir. AMBO ve a-tabakalama madde seçme yöntemlerinin diğer yöntemlere nazaran düşük α -parametresine sahip maddelere de yer verdiği ancak en iyi dengenin a-tabakalama madde seçme yönteminden elde edildiği belirlenmiştir.

BSD yetenek kestiriminin kullanıldığı koşullardan elde edilen bulgular Şekil 3'de sunulmuştur.

Şekil 3 incelendiğinde; a-tabakalama madde seçme yöntemi hariç diğer madde seçme yöntemlerinin yüksek α -parametresine sahip maddeleri kullandığı bulunmuştur. Yetenek kestirim yöntemlerine göre nasıl farklılık olduğu incelendiğinde, MFB ve OAB madde seçme yöntemlerinden elde edilen sonuçların değişmediği görülmüştür. AMBO madde seçme yöntemi, EOT yetenek kestirimine göre yapılan analiz sonucunda düşük α parametresine sahip maddeleri kullanırken, BSD yetenek kestirimine göre yapılan analiz sonucunda bu durumu terk ettiği gözlenmiştir.

Şekil 3: Test durdurma kuralının 40 madde olarak belirlendiği ve BSD yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları



3.3. Üçüncü Alt Probleme İlişkin Bulgular

Araştırmanın üçüncü alt problemi şöyle belirlenmiştir:

“Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde seçmede kullanılan yöntemlere (MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler) ilişkin ortalama madde sayısı,

- a) Test durdurma kuralı ($SH < 0.2$ ve $SH < 0.4$) ve
- b) Yetenek kestirim yöntemlerine (EOT ve BSD) göre nasıl değişmektedir?”

Üçüncü alt problemde test durdurma kuralı sabit standart hata 0.4 ve 0.2’den küçük olarak belirlenmiştir. Madde seçme yöntemlerinin karşılaştırılması için standart hata değerine ulaşıldığında kullanılan madde sayısı ortalaması hesaplanmıştır. Elde edilen bulgular Tablo 6’de sunulmuştur. Elde edilen bulgular durdurma kuralının $SH < 0.2$ ve $SH < 0.4$ değerleri sırasınca açıklanmıştır.

3.3.1. Test Durdurma Kuralı $SH < 0.2$ Olan BBT Koşullarına İlişkin Bulgular

Standart hatanın 0.2’den küçük olduğu ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında; en düşük ve en yüksek madde sayısı ortalaması sırasıyla AMBO ve MFB madde seçme yöntemlerinden elde edilmiştir. BSD yetenek kestiriminin kullanıldığı BBT uygulamalarında ise, en düşük ve en yüksek madde sayısı ortalaması Kullbak-Leibler ve OAB madde seçme yöntemlerinden elde edilmiştir.

Tablo 6: Testi durdurma kuralının sabit standart hataya dayalı olarak belirlendiği BBT koşullarında, yetenek kestirim ve madde seçme yöntemlerine ait istatistikler

Yetenek Kestirim Yöntemi	Madde Seçme Yöntemi	Durdurma Kuralı					
		SH< 0.2			SH<0.4		
		Minimum madde	Maksimum madde	Ortalama madde	Minimum madde	Maksimum madde	Ortalama madde
EOT	MFB	26	95	40.71	7	9	8.72
	a-tabakalama	-	-	-	13	16	14.65
	OAB	27	88	32.85	8	13	9.54
	AMBO	12	41	31.75	7	10	8.96
	Kullbak-Leibler	13	38	32.63	8	12	9.72
BSD	MFB	18	124	30.07	6	11	7.07
	a-tabakalama	-	-	-	12	17	12.54
	OAB	26	78	31.18	8	9	8.41
	AMBO	18	43	30.23	7	12	7.46
	Kullbak-Leibler	27	48	30.13	6	11	7.16

a-tabakalama madde seçme yöntemi, EOT ve BSD yetenek kestirimlerinin her ikisinde de sonuç vermemiştir. Bilgisayar programı madde havuzunda uygun madde bulunamadığı için simülasyonu tamamlayamamıştır. Bu durum, araştırmaya alınan madde havuzu büyüklüğünün yetersiz olmasından veya a-parametre değer aralığının küçük olmasından kaynaklanabileceği şeklinde yorumlanmıştır. a-tabakalama madde seçme yönteminde; madde havuzu, a parametre değerlerine göre tabakalandırılmaktadır ve eldeki araştırmada da madde havuzu üç tabakaya ayrılmıştır. Literatürde farklı büyüklükteki madde havuzları için araştırmalar yapılmıştır. Chang ve Hau(2001) yaptıkları araştırmada a parametre değeri 0.4 ile 2 arasında değişen, 360 maddelik bir madde havuzu ve a-tabakalama madde seçme yöntemi için dört tabaka belirlemiştir, diğer taraftan; Costa, Karino, Moura ve Andrade (2009), 246 maddelik bir madde havuzu kullanarak a-tabakalama madde seçme yöntemini 0.2 standart hata değeri için sonuçlandırabilmiştir. Yapılan araştırmalar ve eldeki bulgular değerlendirildiğinde; a- tabakalama madde seçme yönteminin standart hatanın 0.2'den küçük olması koşulu altında gerçekleştirilememesinin sebebi olarak madde havuzu büyüklüğünün de etkisi olmakla beraber; a parametre değerinin 0.80 ile 1.5 arasında tutulmasından kaynaklandığı düşünülmektedir.

Madde sayısı ortalamaları, yetenek kestirim yöntemleri bakımından incelenmiştir. EOT yetenek kestiriminin kullanıldığı BBT koşullarından elde edilen madde sayısı ortalamalarının, BSD yetenek kestiriminin kullanıldığı BBT koşullarından daha yüksek olduğu görülmüştür. Bu durum, BBT uygulamalarında daha kısa testler elde etmek için BSD yetenek kestiriminin tercih edilebileceği şeklinde yorumlanmıştır.

3.3.2. Test Durdurma Kuralı $SH < 0.4$ Olan BBT Koşullarına İlişkin Bulgular

Testi durdurma kuralı standart hatanın 0.4'den küçük olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında; madde sayısı ortalamasına

göre; en düşük ve en yüksek madde sayısı sırasıyla MFB ve a-tabakalama madde seçme yöntemlerinden elde edilmiştir. BSD yetenek kestiriminin kullanıldığı BBT koşullarında, madde sayısı ortalaması bakımından, MFB ve Kullbak-Leibler madde seçme yöntemi en düşük değere sahipken; a-tabakalama madde seçme yöntemi en yüksek değere sahip olarak bulunmuştur.

Yetenek kestirim yöntemlerinin her ikisinin kullanıldığı durumlarda da en düşük test uzunluğu MFB, en yüksek test uzunluğu a-tabakalama madde seçme yönteminden elde edilmiştir. a-tabakalama madde seçme yönteminin 0.4 standart hata değerine ulaşmak için en çok maddeye ihtiyaç duymasının nedeni, bu madde seçme yönteminin, maddeleri madde havuzunu tabakalandırarak seçmesinden kaynaklanabileceği şeklinde yorumlanmıştır.

3.4. Dördüncü Alt Probleme İlişkin Bulgular

Araştırmanın dördüncü alt problemi şu şekildedir:

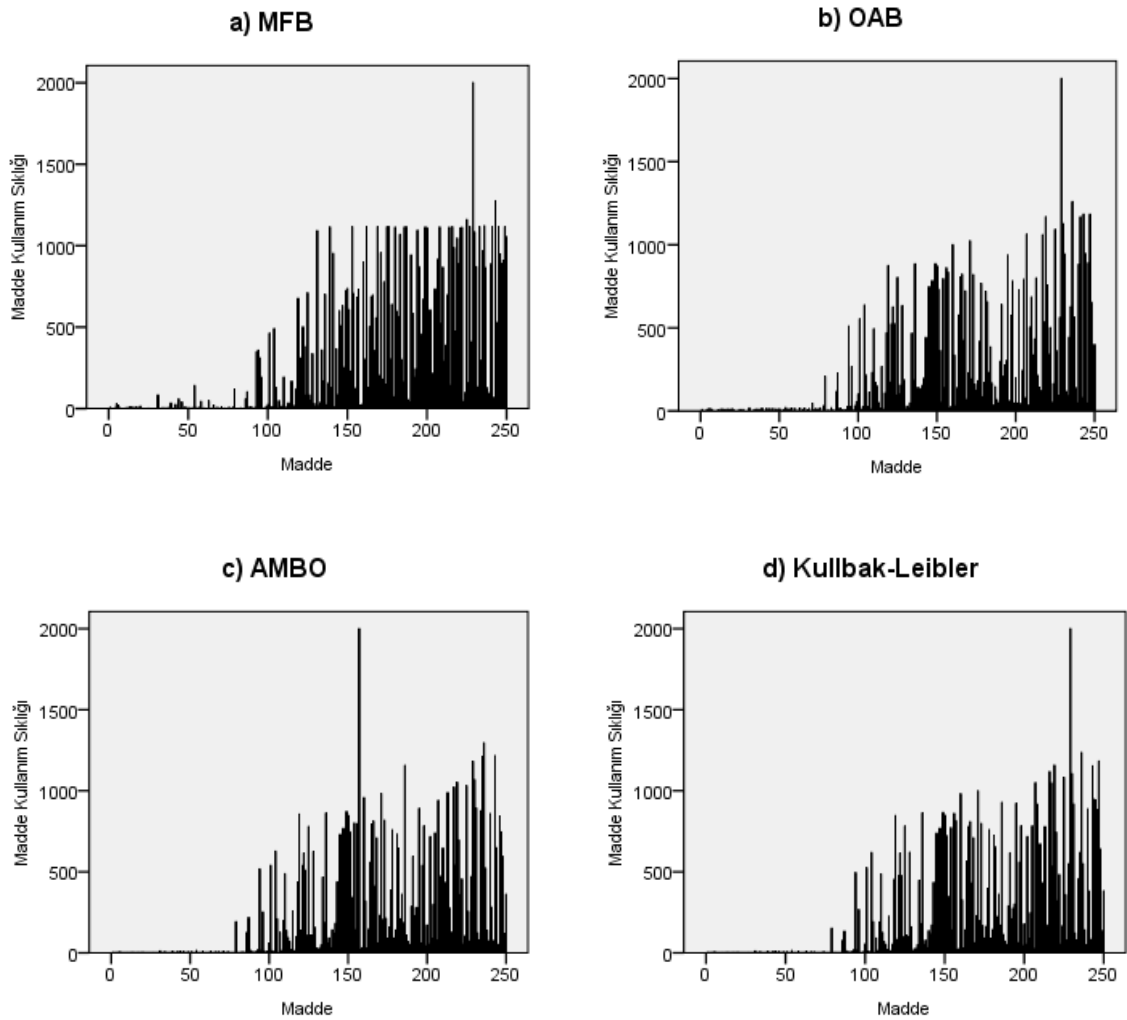
“Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı, sabit standart hataya ($SH < 0.2$ ve $SH < 0.4$) dayalı test durdurma kuralında; madde seçmede kullanılan yöntemlere (MFB, OAB, a-tabakalama, AMBO ve Kullbak-Leibler) ve yetenek kestirim yöntemlerine göre (EOT ve BSD) nasıl değişmektedir?”

Bu alt problemin analizi için; a parametre değerine göre küçükten büyüğe sıralanmış maddeler ile madde kullanım sıklıkları bar grafiğinde gösterilmiştir. Grafikler, madde havuzundaki maddelerin kullanım sıklıkları ve aynı zamanda a parametre değerlerine göre kullanım sıklıkları bakımından yorumlanmıştır. Elde edilen bulgular durdurma kuralının $SH < 0.2$ ve $SH < 0.4$ değerleri sırasınca açıklanmıştır.

3.4.1. Test Durdurma Kuralı $SH < 0.2$ Olan BBT Koşullarında Madde Kullanım Sıklıklarına İlişkin Bulgular

Testi durdurma kuralının SH değerinin 0.2'den küçük olarak belirlendiği ve EOT yetenek kestirim yönteminin kullanıldığı BBT koşullarında madde kullanım sıklıklarına ilişkin bulgular Şekil 4'de sunulmuştur.

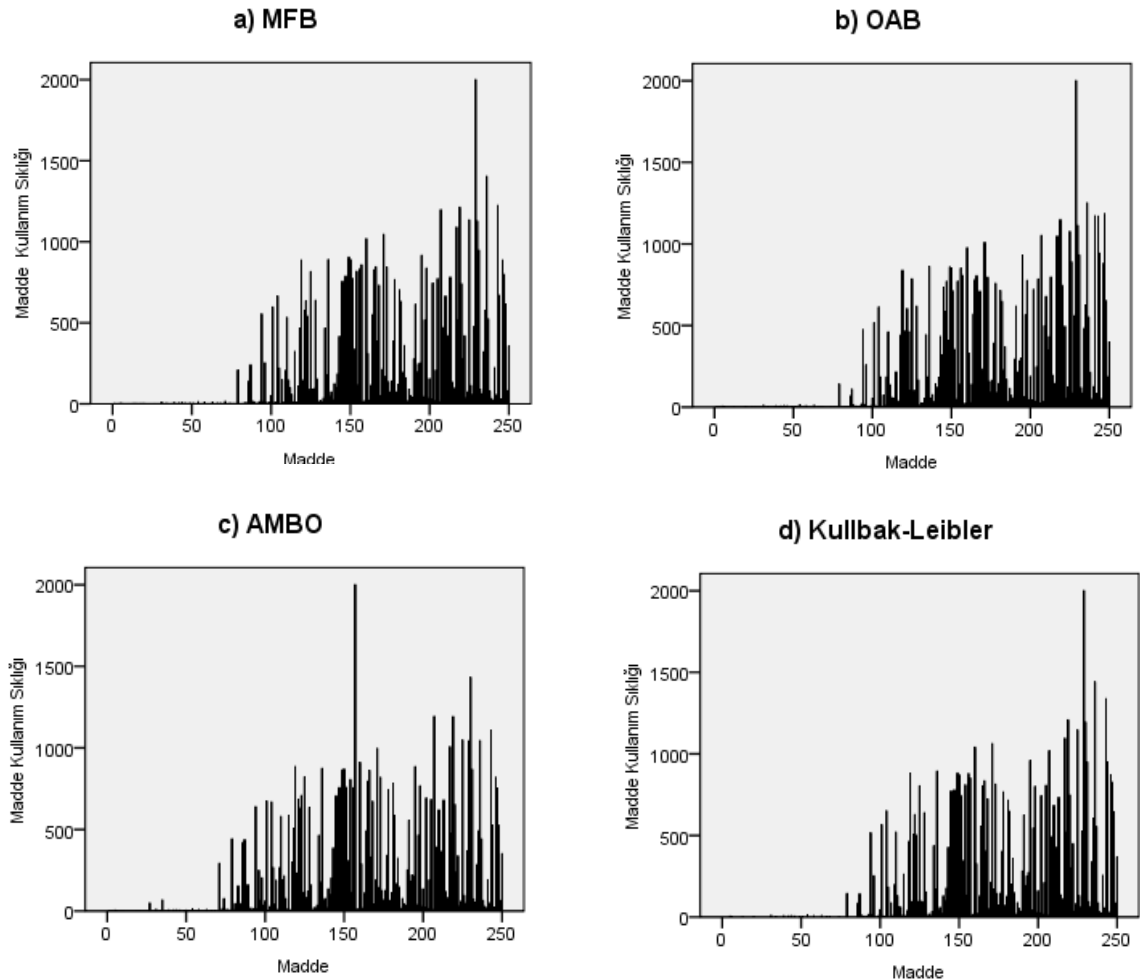
Şekil 4: Test durdurma kuralının $SH < 0.2$ olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları



Şekil 4 incelendiğinde; araştırmaya konu edilen bütün madde seçme yöntemlerinin yüksek α parametre değerine sahip maddeleri kullandıkları görülmektedir. MFB madde seçme yönteminin yüksek α parametre değerine sahip maddeleri daha sıklıkla kullanma problemi, diğer madde seçme yöntemlerinde de gözlenmiştir.

Testi durdurma kuralının SH değerinin 0.2'den küçük olarak belirlendiği ve BSD yetenek kestirim yönteminin kullanıldığı BBT koşullarında madde kullanım sıklıklarına ilişkin bulgular Şekil 5'de sunulmuştur.

Şekil 5: Test durdurma kuralının $SH < 0.2$ olarak belirlendiği ve BSD yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları



Şekil 5’da sunulan bulgulara göre, bütün madde seçme yöntemlerinin yüksek α parametre değerine sahip maddeleri kullandıkları görülmektedir. BSD yetenek kestirimlerine göre elde edilen bulgular, EOT yetenek kestiriminden elde edilen bulgularla örtüşmektedir. Madde havuzundaki maddelerden hiç kullanılmayanlar olduğu gibi, çok yüksek düzeyde kullanılanlar da olmuştur. Bu BBT uygulamalarının test güvenliğini etkileyen bir durumdur.

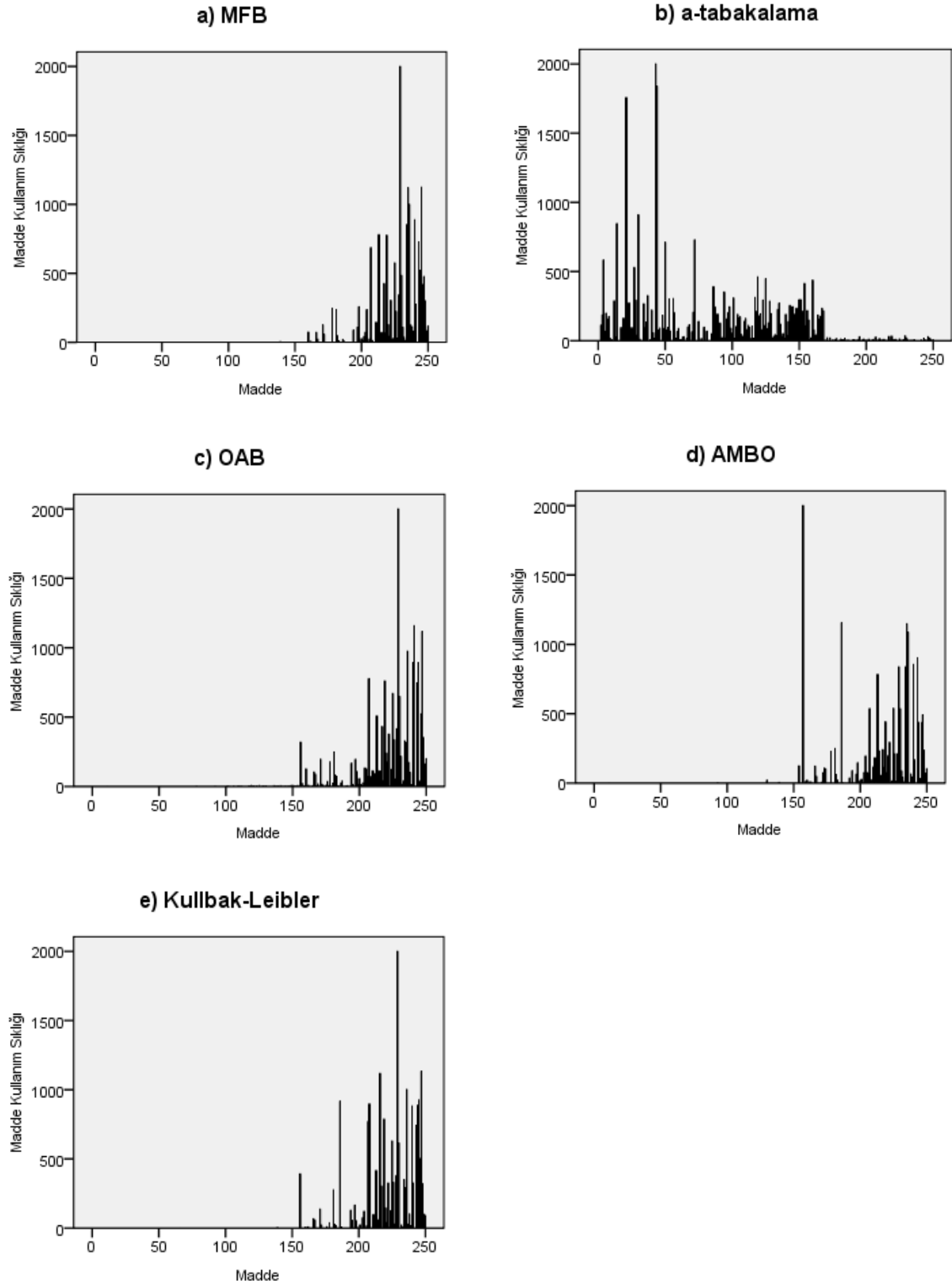
3.4.1. Test Durdurma Kuralı $SH < 0.4$ Olan BBT Koşullarında Madde Kullanım Sıklıklarına İlişkin Bulgular

Testi durdurma kuralının SH değerinin 0.4’den küçük olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında madde kullanım sıklığına ilişkin sonuçlar Şekil 6’de sunulmuştur.

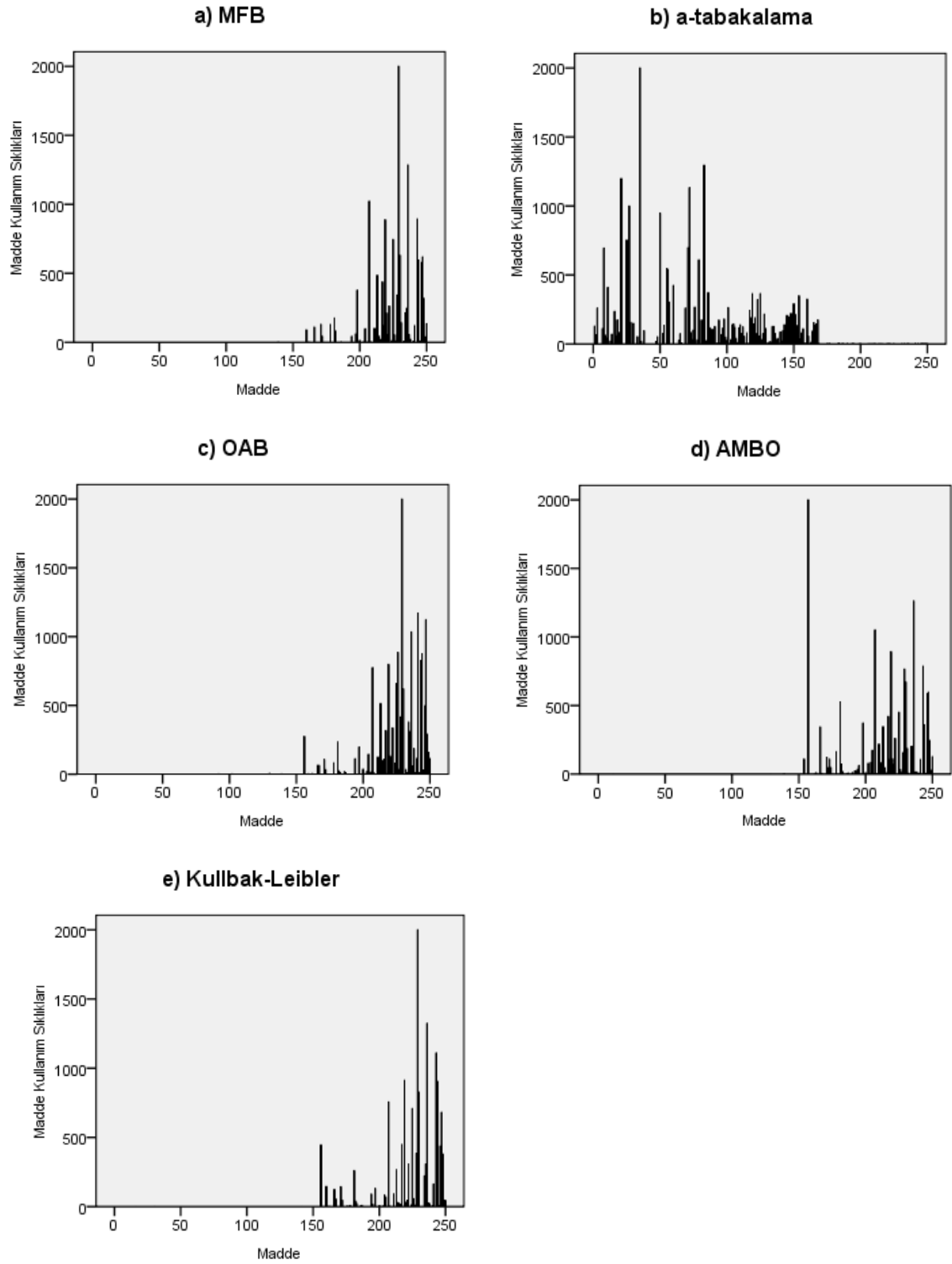
Şekil 6 incelendiğinde; MFB, OAB, AMBO ve Kullbak-Leibler madde seçme yöntemlerinin yüksek α parametre değerine sahip maddeleri kullandıkları düşük α parametresine sahip maddeleri ise hiç kullanmadıkları saptanmıştır. Bunun aksine, α -tabakalama madde seçme yönteminin düşük α parametresine sahip maddeleri daha sıklıkla kullandığı bulunmuştur. Yi, Chang (2003) ve Wen, Chang, Hau (2001) yaptıkları araştırmalarda, BSD yetenek kestiriminin kullanıldığı BBT koşullarında α -tabakalama madde seçme yönteminin madde havuzu kullanımında iyi bir denge sağladığını belirtmişlerdir. Eldeki araştırmada ise benzer bulgular elde edilmemiştir. α -tabakalama madde seçme yöntemi, yüksek α parametre değerine sahip maddeleri kullanma probleminin önüne geçmiştir. Ancak, madde havuzundaki maddelerin kullanımında iyi bir denge göstermemiştir.

Testi durdurma kuralının SH değerinin 0.4’den küçük olarak belirlendiği ve BSD yetenek kestiriminin kullanıldığı BBT koşullarında madde kullanım sıklığına ilişkin sonuçlar Şekil 7’de sunulmuştur.

Şekil 6: Test durdurma kuralının $SH < 0.4$ olarak belirlendiği ve EOT yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları



Şekil 7: Test durdurma kuralının $SH < 0.4$ olarak belirlendiği ve BSD yetenek kestiriminin kullanıldığı BBT koşullarında madde seçme yöntemlerine göre madde kullanım sıklıkları



Şekil 7 incelendiğinde; BSD yetenek kestiriminden elde edilen bulguların, EOT yetenek kestiriminden elde edilen bulgularla örtüştüğü görülmüştür. Araştırmaya alınan tüm madde seçme yöntemlerinde madde havuzu kullanımına ilişkin iyi bir denge olmadığı sonucuna varılmıştır. Nitekim bazı maddelerin kullanılma sıklıkları 0 iken bazı maddeler tüm cevaplayıcılar için kullanılarak kullanım sıklığı 2000 değerine ulaşmıştır. Daha önce de belirtildiği gibi bu test güvenliğini etkileyen bir durumdur. Araştırmaya alınan koşullar altında madde seçme yöntemlerinin birbirine üstünlüğü olmadığı sonucuna varılmıştır.

4. BÖLÜM

SONUÇ VE ÖNERİLER

4.1 Birinci Alt Probleme İlişkin Sonuçlar

Test uzunluğu 5, 10, 20, 30 ve 40 madde olarak belirlendiği BBT koşullarında elde edilen sonuçlar şöyledir:

EOT yetenek kestirimi kullanıldığında;

- Test uzunluğu 5, 10 ve 20 madde iken en düşük SH değeri AMBO madde seçme yönteminden elde edilmiştir.
- Test uzunluğu $n < 30$ iken, a-tabakalama; $n > 30$ iken Kullbak-Leibler madde seçme yöntemi en yüksek SH değerini göstermiştir.
- BBT koşullarının başında ($n < 10$), araştırmaya alınan madde seçme yöntemlerinin SH değerleri arasında büyük farklar olduğu, ancak test uzunluğu arttıkça bu farkın azaldığı görülmüştür.

BSD yetenek kestirimi kullanıldığında,

- Araştırmaya alınan bütün farklı test uzunluklarında en yüksek SH değeri a-tabakalama madde seçme yönteminden elde edilmiştir.
- BBT koşullarının başında ($n < 10$), araştırmaya alınan madde seçme yöntemlerinin SH değerleri arasında büyük farklar olduğu, ancak test uzunluğu arttıkça bu farkın azaldığı görülmüştür. Test uzunluğu 40 madde olduğunda bütün madde seçme yöntemlerinin SH değerleri birbirine eşit sonuçlar vermiştir.

- EOT yetenek kestirimi kullanıldığında elde edilen SH değerleri, BSD yetenek kestirimi kullanıldığında elde edilen SH değerlerinden daha yüksek bulunmuştur.

4.2. İkinci Alt Probleme İlişkin Sonuçlar

- Madde havuzu kullanım sıklığına göre yapılan incelemede, MFB, OAB ve Kullbak-Leibler madde seçme yöntemlerinin yüksek α parametre değerine sahip maddeleri daha çok kullandıkları saptanmıştır. Madde havuzu kullanımında en iyi dengeyi α -tabakalama madde seçme yöntemi göstermiştir.
- Madde havuzu kullanım sıklığından elde edilen sonuçlar, yetenek kestirimlerine göre AMBO madde seçme yöntemi hariç diğer madde seçme yöntemlerinde bir farklılık göstermemiştir.

4.3. Üçüncü Alt Probleme İlişkin Sonuçlar

SH<0.2 olduğu BBT koşullarında

- EOT yetenek kestirimi kullanıldığında en düşük madde sayısı ortalaması AMBO madde seçme yönteminden, en yüksek madde sayısı ortalaması MFB madde seçme yönteminden elde edilmiştir.
- BSD yetenek kestirimi kullanıldığında en düşük madde sayısı ortalaması Kullbak-Leibler madde seçme yönteminden, en yüksek madde sayısı ortalaması OAB madde seçme yönteminden elde edilmiştir.
- EOT ve BSD yetenek kestirimlerinin kullanıldığı her iki durumda da α -tabakalama madde seçme yöntemi sonuç vermemiştir. Bu durumun madde havuzu büyüklüğünün yetersiz kalmasından ve araştırmaya alınan α parametre değeri ranjının düşük olmasından kaynaklandığı sonucuna varılmıştır.

- Madde sayısı ortalamaları, yetenek kestirim yöntemleri bakımından incelendiğinde; EOT yetenek kestiriminin kullanıldığı koşulların, BSD yetenek kestiriminin kullanıldığı koşullardan daha yüksek olduğu sonucuna varılmıştır. $SH < 0.4$ olduğu BBT koşullarında
- EOT yetenek kestirimi kullanıldığında en düşük madde sayısı ortalaması MFB madde seçme yönteminden, en yüksek madde sayısı ortalaması Kullbak-Leibler madde seçme yönteminden elde edilmiştir.
- BSD yetenek kestirimi kullanıldığında en düşük madde sayısı ortalaması MFB ve Kullbak-Leibler madde seçme yöntemlerinden, en yüksek madde sayısı ortalaması a-tabakalama madde seçme yönteminden elde edilmiştir.
- Araştırmaya alınan bütün madde seçme yöntemleri için, EOT yetenek kestiriminden elde edilen ortalama test uzunluğu, BSD yetenek kestiriminden elde edilen ortalama test uzunluğundan yüksek bulunmuştur. BSD yetenek kestiriminin kullanıldığı BBT uygulamalarında daha kısa testler elde edileceği sonucuna varılmıştır.

4.4. Dördüncü Alt Probleme İlişkin Sonuçlar

- Madde havuzu kullanım sıklığı bakımından yapılan incelemede araştırmaya alınan bütün madde seçme yöntemlerinin; $SH < 0.2$ olduğu BBT koşullarında yüksek a parametresine sahip maddeleri kullandıkları sonucuna varılmıştır. Bu durum, yetenek kestirim yöntemine göre bir farklılık göstermemiştir.
- $SH < 0.4$ olduğu BBT koşullarında madde kullanım sıklığı incelendiğinde; MFB, OAB, AMBO ve Kullbak-Leibler madde seçme yöntemlerinin yüksek a parametresine sahip maddeleri kullandıkları saptanmıştır. Bunun aksine, a-tabakalama madde seçme yönteminin düşük a parametresine sahip maddeleri kullandığı görülmüştür. Elde edilen bu bulgu, yetenek kestirim yöntemine göre farklılık göstermemiştir.

- Araştırmaya alınan bütün madde seçme yöntemlerinin; madde havuzu kullanımına ilişkin iyi bir denge göstermediği sonucuna varılmıştır. Madde seçme yöntemlerinin bu bakımdan birbirine üstünlüğü olmadığı sonucuna varılmıştır.

4.5. Öneriler

- Madde seçme yöntemlerine ait SH değerleri, BSD yetenek kestirimi kullanıldığında daha düşük sonuç vermiştir. Operasyonel BBT uygulamalarında BSD yetenek kestiriminin kullanılması önerilmektedir
- BBT uygulamalarının önemli savlarından biri de kağıt kalem testlerine kıyasla daha kısa uzunlukta testler oluşturmalarıdır. Elde edilen sonuçlar değerlendirildiğinde; BBT uygulamalarında BSD yetenek kestirim yönteminin tercih edilmesi önerilmektedir.
- a-tabakalama madde seçme yöntemi, test durdurma kuralı $SH < 0.2$ olduğu koşulda sonuç vermemiştir. Bu üzerinde çalışılması gereken bir bulgudur. Farklı madde havuzu büyüklükleri ve a-parametre değerleri belirlenerek araştırmalar yapılması önerilmektedir. Ayrıca, a-tabakalama madde seçme yönteminde kullanılan tabakalama sayısının da bu değişkenlerle ilişkisi irdelenebilir.

İleride yapılacak çalışmalarla ilgili olarak

- Bu araştırmada, yetenek kestirim yöntemleri EOT ve BSD yetenek kestirim yöntemleri ile sınırlandırılmıştır. Farklı yetenek kestirim yöntemlerine yer veren araştırmalar yapılabilir.
- BBT uygulamalarının bileşenlerinden testi başlatma kuralları ve madde havuzu büyüklüğünün madde seçme yöntemleri üzerindeki etkisi araştırılabilir.
- Bu araştırmada, madde kullanım sıklığı, madde havuzu kullanımı düzeyinde incelenmiştir; madde kullanım sıklığını kontrol eden yöntemlere değinilmemiştir. İleride yapılacak araştırmalarda madde

kullanım sıklığı yöntemlerinin, madde seçme yöntemlerini nasıl etkilediği konusu ele alınabilir.

- BBT uygulamalarını ilgilendiren bir diğer konu içerik balansı(content balance)dır. İçeriğe göre ağırlıklandırılmış madde havuzunda madde seçme yöntemlerinin nasıl işlediği araştırma konusu yapılabilir.
- Bu araştırmada tek boyutlu madde tepki kuramı ele alınmıştır, çok boyutlu madde tepki kuramına dayalı araştırmalara yer verilebilir.
- Eldeki araştırma simülatif olarak yürütülmüştür, operasyonel BBT uygulamalarında elde edilen bulguların nasıl işlediği araştırılabilir.

KAYNAKÇA

- Bock, R. D., Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters. Application of an EM algorithm. *Psychometrika*, 46(4), pp 433-459.
- Costa, D., Karino, C., Moura, F., Andrade, D. (2009). A Comparision of Three Methods of Item Selection for Computerized Adaptive Testing. 2009 GMAC Conference on Computerized Adaptive Testing, June,
- Deng, H., Ansley, T., Chang, H. (2010). Stratified and Maximum Information Item Selection Procedures in Computer Adaptive Testing. *Journal of Educational Measurement*, Vol.47, No.2, pp 202-226.
- Eggen, T. H. J. M. (1999). Item Selection in Adaptive Testing with the Squential Probability Ratio Test. *Applied Psychological Measurement*, Vol.23, No.3., pp 249-261.
- Eggen, T.H.J.M. (2004). Contributions to the Theory and Practice of Computerized Adaptive Testing. Print Partners Ipskamp B.V., Citogroup Arnhem, NL, ISBN: 90-5834-056-2.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer-Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.

- Han, K. (2009). Gradual Maximum Information Ratio Approach to Item Selection in computerized Adaptive Testing. Graduate Management Admission Council Research Reports, RR-09-07, June 25, USA.
- Han, K. (2010). Comparision of Non-Fisher Information Item Selection Criteria in Fixed Length Computerized Adaptive Testing. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Denver.
- Han, K. (2012). SimulCAT: Windows Application That Simulates Computerized Adaptive Test Administration. Applied Psychological Measurement, 36.
- Ho, T. (2010). A Comparison of Item Selection Procedures Using Different Ability Estimation Methods in Computerized Adaptive Testing Based on Generalized Partial Credit Model. Doctoral Dissertation, Graduate School of the University of Texas at Austin.
- Iseri, A. I. (2002). Assessment of Students' Mathematics Achievement Through Computer Adaptive Testing Procedures. Unpublished doctoral dissertation. Middle East Technical University, Turkey.
- Kaptan, F. (1993). Yetenek Kestiriminde Adaptive (bireyselleştirilmiş) Test Uygulaması ile Geleneksel Kağıt-kalem Testi Uygulamasının Karşılaştırılması. Yayımlanmamış doktora tezi, Hacettepe Üniversitesi
- Kalender, İ. (2011). Effects of Different Computerized Adaptive Testing Strategies on Recovery of Ability. Unpublished Doctoral Dissertation. Middle East Technical University, Ankara.

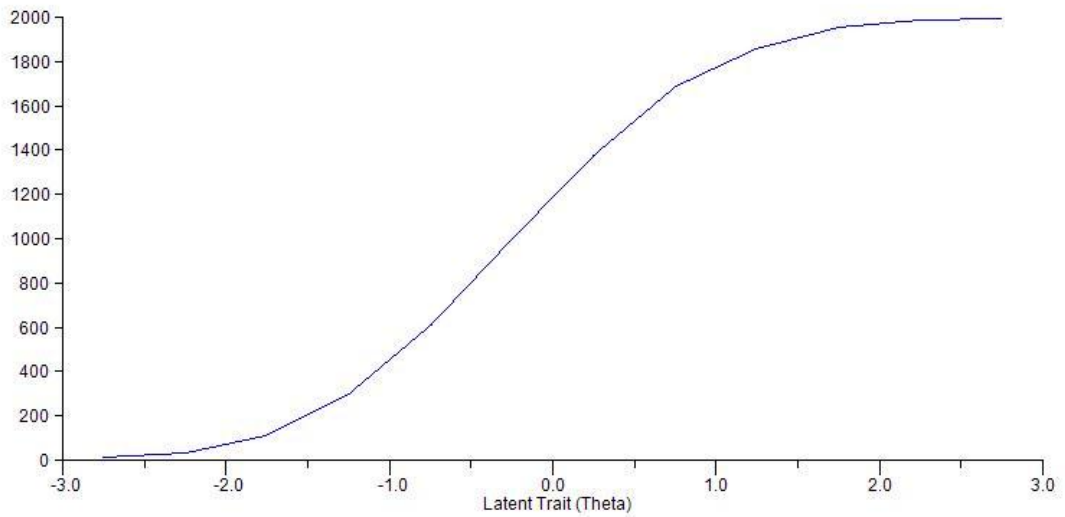
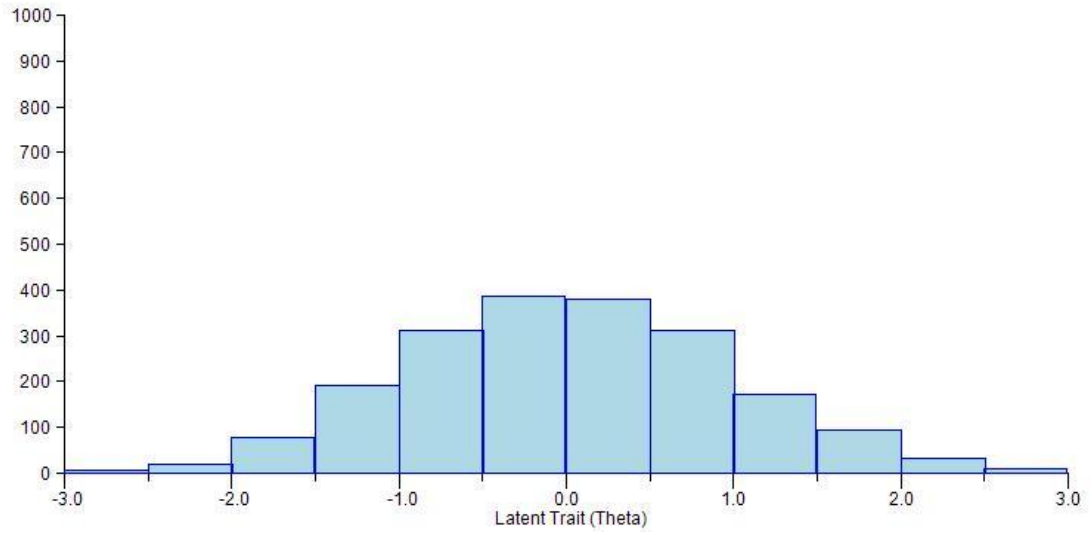
- Kingsbury, G. G., Zara, A. R. (1989). Procedures for Selecting Items for Computerized Adaptive Tests. *Applied Measurement in Education*, 2(4), pp 359-375).
- Köklü, N. (1990). Klasik Test Teorisi'ne Göre Geliştirilen Tailored Test ile Grup Testi Arasında Bir Karşılaştırma. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Türkiye
- Linacre, J. M. (2000). Computer-Adaptive Testing: A Methodology Whose Time Has Come. MESA Memorandum.
- Linda, T. (1996). A comparison of the Traditional Maximum Information Method and the Global Information Method in CAT Item Selection. Annual Meeting of the National Council on Measurement in Education, New York, April.
- McBride, J.R. (1985). Computerized Adaptive Testing. *Educational Leadership*, October.
- Orcutt, V. L. (2002). Computerized Adaptive Testing: Some Issues in Development. Annual Meeting of the Educational Research Exchange, University of North Texas, February, Denton, Texas.
- Sireci, S. (2003). Computerized Adaptive Testing: An Introduction. *Measuring Up: Assessment Issues for Teachers, Counselors and Administrators*, 12p.,
- Slater, S. C. (2001). Pretest Item Calibration Within The Computerized Adaptive Testing Environment. Unpublished Doctoral Dissertation, Graduate School of the University Massachusetts, Amherst.

- Stocking, M. L. (1992). Controlling Item Exposure Rates in a Realistic Adaptive Testing Paradigm. (Research Report 93-2). Princeton, NJ: Educational Testing Service.
- Thissen, D. & Mislevy, R. J. (2000). Testing algorithms. In H. Wainer, (Eds.). Computerized Adaptive Testing: A primer, Mahwah, NH: Lawrence Erlbaum Associates, Inc, pp. 101-133.
- Tian J., Miao, D; Zhu, X; Gong, J. (2007). An Introduction to the Adaptive Testing, US-China Education Review, Volume 4, No.1, ISBN:1548-6613, USA.
- Urry, V. W. (1977). Tailored Testing: A Successful Application of Latent Trait Theory. *Journal of Educational Measurement*, Vol.14, No.2, pp 181-196.
- Van Der Linden, W.J., Glas, C.A.W. (2010). Elements of Adaptive Testing, Statistics for Social and Behavioral Sciences, Springer New York Dordrecht Heidelberg London, ISBN: 978-0-387-85459-5.
- Veerkamp, W.J.J., Berger, M.P.F. (1997). Some New Item Selection Criteria for Adaptive Testing. *Journal of Educational and Behavioral Statistics*, Vol.22, No.2, pp 203-226.
- Veldkamp, B.P. (2012). Ensuring The Future of Computerized Adaptive Testing. In Theo, J.H.M; Veldkamp, B.P. (ed). *Psychometrics in Practice at RCEC*. University of Twente, Netherlands, 978-90-365-3374-4.
- Wainer, H., Dorans, N., Flaugher, R., Green, B., Mislevy, R., Steinberg, L., Thissen, D. (1990) Computerized adaptive testing: A primer. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Wang, T., Vispoel, W. (1998). Properties of Ability Estimation Methods in Computerized Adaptive Testing. *Journal of Educational Measurement*, Vol.35, No.2, pp 109-135.
- Weiss, David J. (1983). Latent Trait Theory and Adaptive Testing. In David J. Weiss (ed.). *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing*. (pp. 5-7). New York: Academic Press.
- Weiss, D.J., Kang, G.K. (2007). Comparison of Computerized Adaptive Testing and Classical Methods for Measuring Individual Change. Graduate Management Admission Council, Item Calibration and Special Applications Paper Session, June 7.
- Weiss, D. J., Kingsbury, G. G. (1984). Application of Computerized Adaptive Testing to Educational Problems. *Journal of Educational Measurement*, 21, 361-375.
- Weiss, D. J. (2010). CAT Central: A Global Resource for Computerized Adaptive Testing Research and Applications [Online]. <http://www.psych.umn.edu/psylabs/CATCentral>. Last visited on 25/11/2012.
- Weissman, A. (2003). Assessing the Efficiency of Item Selection in Computerized Adaptive Testing. Paper presented at the Annual Meeting of the American Educational Research Association, April, Chicago.
- Wen, H., Chang, H., Hau, K. (2001). Adaption of a-stratified Method in Variable Length Computerized Adaptive Testing. American Educational Research Association Annual Meeting, Seattle.

Yi, Q., Chang, H. (2003). α -Stratified CAT Design With Content Blocking. *British Journal of Mathematical and Statistical Psychology*, vol. 56, pp 359–378.

EKLER

Ek:1 Araştırmanın Yürütüldüğü Grup Dağılım Grafiği

Ek:2 SimulCAT Bilgisayar Programı

SimulCAT

File About

STEP1: Examinee / Item Data STEP2: Item Selection STEP3: Test Administration

Examinee Characteristics

Number of Examinees

Distribution

☒ Normal ☐ Uniform ☐ Beta

Mean SD

☐ Regenerate for each replication

Generate True Scores (Theta)

Item Characteristics

Number of Items

Distribution

Par.a Mean SD

Par.b Mean SD

Par.c Mean SD

Content ID 1

☒ Scale to normal metric (scaling factor D=1.702)

☐ Add to the previous item set

Generate True Item Parameters

Examinee(s)

Item Parameters(a,b,c)

N = 0
Mean = 0
SD = 0

Histogram

a: Mean=0 / SD=0
b: Mean=0 / SD=0
c: Mean=0 / SD=0

Plot Item(s)

Ready Elapsed Time 00:00:00

SimulCAT

File About

STEP1: Examinee / Item Data STEP2: Item Selection STEP3: Test Administration

Item Selection Criterion

☒ Maximum Fisher Information (MFI)

☐ α -Stratification (Number of Strata: 3) ☐ b-blocking

☐ Matching b-Value

☐ Random Selection

☐ Interval Information Criterion (IIC)

☐ Likelihood Weighted Information (LWI)

☐ Kullback-Leibler Information (KL), Global Information Constant: 3

☐ Gradual Maximum Information Ratio (GMIR)

☐ Efficiency Balanced Information (EBI)

Item Exposure Control

☒ No Exposure Control

☐ Fade Away Method (FAM, Target Exposure Rate: 0.2)

☐ Randomesque (Randomly select an item among 5 best items)

☐ Probabilistic Approach with Simulations

☐ Simpson and Hotter Method (SHM)

☐ Unconditional Multinomial Method (UMM)

☐ Conditional Multinomial Method (CMM)

☒ Load Item Exposure Parameter Data

☐ Derive Item Exposure Parameters after 20 iterations

Target Exposure Rate: 0.2

(only for CMM, number of intervals: 6 between -3 and 3)

☐ Cumulative Item Usage Criterion for Retirement 0

Item Usage Update (only for FAM and/or Retirement)

☒ Real Time ☐ After Each Test Time Slot ☐ Daily Basis

Test Length

☒ Fixed Length

☐ Variable Length

☐ Terminate when SEE becomes smaller than

☐ Terminates when the change in interim estimates becomes smaller than for the last items

☐ Minimum items

☐ Maximum items

Expected Length: items (only for 'b-plot' and 'GMIR')

Content Balancing

☒ None

☐ By Script

☐ By Weight

Open Content File

Ready Elapsed Time 00:00:00

[Vme] SimulCAT

File About

STEP1: Examinees / Item Data STEP2: Item Selection STEP3: Test Administration

Score Estimation

☒ Maximum Likelihood Estimation (MLE)
☐ Bayesian Maximum a Posteriori (MAP) Post. Mean: 0
☐ Bayes Expected a Posteriori (EAP) Post. SD: 1

Initial Score Value
☐ Fixed Value
☒ Randomly Chosen Value Between -0.5 and 0.5
☐ From a Data File (*.wge) Browse

Options
☒ Limit (Truncate) Range of Estimates
Lower Bound: -3 Upper Bound: 3
☐ Limit the Estimate Jump by for the first items
☐ Use MLE for the Final Estimate

Test Administration

Number of Examinees for Each Test Time Slot
Number of Test Time Slots per Day

Pretest Item Administration

☒ None
☐ Administer Items per Examinee Open Pretest Item File

Extras

☐ Generate Replication Data Sets Number of Replications
☐ Item Pool with DIF/Drift Browse
☐ Previous Item Usage Data Browse
☐ Fixed Seed Value

Output

☐ Save Response Strings and Item IDs
☐ Save Interim Theta Estimates
☒ Save Item Use (Exposure) (in a *.SCU file)
☐ Save Full Response Matrix
☐ Save Interim SEEs and TIFs ☐ Save True Interim SEEs and TIFs

Output File Browse

Run Simulation

Log/Message

Ready Elapsed Time 00:00:00

Ek:3 Madde Havuzu Parametre Değerleri

Madde No	a	b	c
1	1.138	-0.873	0.138
2	1.487	1.254	0.071
3	1.345	2.458	0.079
4	1.251	0.325	0.116
5	1.183	0.917	0.126
6	1.430	-1.694	0.096
7	1.154	-0.332	0.142
8	1.127	1.838	0.134
9	0.939	2.422	0.146
10	1.422	2.252	0.081
11	1.411	1.217	0.139
12	1.432	2.564	0.099
13	1.363	-0.583	0.092
14	1.237	-0.422	0.093
15	1.386	-0.900	0.088
16	0.925	1.555	0.055
17	0.929	-0.458	0.061
18	1.291	1.362	0.054
19	1.330	0.969	0.080
20	1.399	0.563	0.102
21	1.196	-2.730	0.054
22	0.992	1.366	0.069

23	1.477	2.860	0.056
24	1.214	-1.115	0.130
25	1.351	-1.623	0.113
26	1.464	1.633	0.073
27	1.227	-0.593	0.092
28	0.912	1.707	0.087
29	1.326	-2.088	0.078
30	1.443	-2.422	0.116
31	1.298	1.564	0.130
32	1.299	-0.053	0.063
33	0.888	1.662	0.069
34	1.052	0.584	0.054
35	1.155	1.820	0.117
36	1.010	-0.399	0.110
37	1.008	1.124	0.096
38	0.852	2.093	0.086
39	0.817	0.922	0.074
40	1.241	-1.418	0.076
41	1.205	-2.028	0.088
42	1.060	-0.798	0.096
43	1.437	0.091	0.098
44	0.863	1.068	0.074
45	1.259	0.029	0.054
46	0.951	-2.243	0.137
47	1.153	0.739	0.074

48	0.806	-2.654	0.064
49	1.006	0.729	0.065
50	1.233	-1.056	0.129
51	0.921	-0.307	0.130
52	0.920	-0.849	0.075
53	1.006	2.095	0.061
54	0.989	-0.615	0.085
55	1.420	-2.090	0.097
56	1.212	-1.485	0.139
57	1.280	-2.802	0.086
58	1.109	-1.267	0.067
59	0.829	-1.303	0.111
60	1.158	0.109	0.056
61	0.936	2.134	0.142
62	0.881	2.762	0.056
63	1.183	1.428	0.131
64	1.285	1.086	0.085
65	1.452	-0.513	0.075
66	1.070	-2.036	0.058
67	0.905	2.766	0.088
68	1.436	2.423	0.056
69	1.049	-2.699	0.083
70	0.860	-0.359	0.130
71	0.805	-0.622	0.147
72	1.304	-2.229	0.108

73	1.131	-0.452	0.146
74	1.287	-0.501	0.131
75	1.167	-1.843	0.111
76	1.142	-0.209	0.057
77	1.004	2.095	0.087
78	1.126	-1.257	0.098
79	1.097	-2.062	0.138
80	1.252	2.003	0.081
81	1.165	2.176	0.051
82	1.375	0.669	0.149
83	0.857	2.719	0.083
84	0.943	2.129	0.138
85	1.172	-2.641	0.119
86	1.496	-1.241	0.108
87	0.868	0.380	0.074
88	1.169	2.934	0.101
89	1.186	-1.987	0.133
90	1.442	2.274	0.057
91	0.890	2.899	0.105
92	1.037	2.574	0.123
93	0.886	-0.721	0.060
94	1.054	-2.652	0.115
95	1.477	-1.928	0.054
96	1.497	-1.566	0.131
97	1.363	2.540	0.071

98	1.036	-0.418	0.134
99	1.404	-1.267	0.132
100	0.800	-1.521	0.131
101	0.888	2.386	0.053
102	1.302	0.138	0.132
103	0.879	-2.528	0.052
104	1.335	-1.908	0.057
105	1.351	-0.915	0.136
106	0.899	-2.669	0.140
107	1.412	0.853	0.121
108	1.240	0.419	0.127
109	0.881	0.553	0.057
110	0.908	-0.275	0.080
111	0.979	-0.866	0.068
112	0.886	2.215	0.136
113	1.339	1.820	0.137
114	1.214	0.727	0.069
115	1.248	-2.514	0.079
116	1.141	-0.815	0.070
117	1.281	1.999	0.093
118	0.959	2.676	0.139
119	0.981	0.707	0.079
120	1.106	-1.171	0.144
121	1.004	-1.242	0.070
122	1.258	0.444	0.097

123	0.820	-0.551	0.116
124	0.921	-1.212	0.066
125	1.462	2.285	0.098
126	1.107	2.523	0.096
127	1.237	-0.069	0.089
128	1.258	2.681	0.127
129	1.258	2.758	0.074
130	1.148	-0.374	0.135
131	1.363	1.158	0.091
132	0.875	1.457	0.115
133	0.811	-0.684	0.072
134	1.323	0.771	0.094
135	1.496	-2.359	0.079
136	1.358	-1.767	0.052
137	1.099	0.523	0.074
138	1.494	-1.115	0.075
139	0.880	1.617	0.068
140	0.828	-1.359	0.062
141	1.176	2.341	0.129
142	0.844	-2.109	0.086
143	1.257	0.766	0.055
144	1.128	2.668	0.057
145	1.462	-2.402	0.136
146	0.948	-0.697	0.140
147	0.801	1.429	0.061

148	1.185	-1.994	0.150
149	0.915	-1.046	0.149
150	1.325	2.322	0.083
151	0.818	1.329	0.101
152	1.045	2.888	0.098
153	1.367	-2.628	0.113
154	1.456	-0.571	0.141
155	1.429	-0.213	0.084
156	1.182	2.439	0.063
157	0.847	0.089	0.122
158	0.887	-2.525	0.102
159	0.923	-2.556	0.127
160	1.288	0.513	0.103
161	1.069	-1.920	0.099
162	0.903	-1.718	0.116
163	0.889	-2.521	0.105
164	1.332	1.043	0.085
165	1.102	1.029	0.106
166	1.470	-1.272	0.119
167	1.017	2.324	0.119
168	1.394	2.881	0.121
169	1.010	1.798	0.076
170	1.168	-1.651	0.089
171	1.081	-0.113	0.093
172	0.979	2.143	0.141

173	1.472	2.396	0.071
174	1.480	2.426	0.119
175	0.992	-0.730	0.091
176	1.302	2.051	0.071
177	1.469	-0.553	0.065
178	1.473	2.586	0.118
179	1.028	1.987	0.067
180	1.448	0.279	0.054
181	1.059	1.159	0.051
182	1.041	0.886	0.141
183	0.838	-1.318	0.132
184	1.008	1.798	0.089
185	0.810	-2.207	0.077
186	0.831	-0.582	0.140
187	1.425	-2.652	0.050
188	1.492	-2.786	0.092
189	1.425	-0.370	0.097
190	0.942	-2.608	0.137
191	0.996	-0.951	0.059
192	0.911	2.618	0.087
193	1.431	-1.558	0.061
194	1.197	1.738	0.128
195	1.493	-0.866	0.115
196	1.143	-1.074	0.118
197	1.345	2.462	0.119

198	0.891	-2.804	0.112
199	1.301	1.957	0.064
200	1.465	-1.276	0.093
201	1.019	0.848	0.056
202	0.818	1.909	0.137
203	1.425	1.100	0.136
204	1.365	0.613	0.140
205	1.354	1.656	0.146
206	1.265	-1.480	0.081
207	1.342	-2.915	0.066
208	1.429	-0.950	0.106
209	1.295	-0.694	0.139
210	0.886	-1.237	0.054
211	1.337	-1.402	0.140
212	0.901	1.693	0.062
213	1.297	-2.446	0.067
214	1.354	1.555	0.118
215	0.801	1.204	0.062
216	1.017	0.246	0.147
217	1.367	-2.043	0.077
218	0.981	0.597	0.149
219	0.918	-2.429	0.055
220	0.822	2.802	0.132
221	1.382	1.843	0.089
222	1.118	-0.540	0.103

223	1.213	1.182	0.074
224	0.848	1.683	0.106
225	0.879	1.782	0.076
226	0.827	-1.818	0.126
227	1.184	-0.441	0.060
228	1.380	2.793	0.094
229	1.125	-1.447	0.079
230	0.833	-0.637	0.074
231	1.408	-1.021	0.102
232	0.893	2.450	0.094
233	1.059	1.611	0.131
234	1.342	-2.245	0.070
235	1.320	-1.307	0.137
236	0.958	1.740	0.139
237	1.106	1.058	0.121
238	1.444	1.360	0.077
239	1.367	2.647	0.070
240	0.831	-1.646	0.091
241	0.873	1.112	0.074
242	1.483	0.021	0.101
243	1.137	2.698	0.119
244	1.315	-2.065	0.051
245	1.215	1.075	0.062
246	1.410	1.517	0.117
247	1.167	-0.666	0.082

248	1.094	2.921	0.099
249	0.873	1.802	0.130
250	1.402	-2.845	0.120