

**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
ÖLÇME VE DEĞERLENDİRME PROGRAMI**

**BİREYE UYARLANMIŞ TESTLERDE MADDE KULLANIM SIKLIĞI KONTROL
YÖNTEMLERİNİN FARKLI KOŞULLARDA ÖLÇME DUYARLILIĞINA ve TEST
GÜVENLİĞİNE ETKİSİ**

DOKTORA TEZİ

RECEP GÜR

ANKARA, KASIM, 2017

**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
ÖLÇME VE DEĞERLENDİRME PROGRAMI**

**BİREYE UYARLANMIŞ TESTLERDE MADDE KULLANIM SIKLIĞI KONTROL
YÖNTEMLERİNİN FARKLI KOŞULLARDA ÖLÇME DUYARLILIĞINA ve TEST
GÜVENLİĞİNE ETKİSİ**

DOKTORA TEZİ

RECEP GÜR

Danışman: Yrd. Doç. Dr. Hamide Deniz GÜLLEROĞLU

ANKARA, KASIM, 2017

ONAY

Eğitim Bilimleri Enstitüsü Müdürlüğü'ne

Recep GÜR'ün hazırladığı “Bireye Uyarlanmış Testlerde Madde Kullanım Sıklığı Kontrol Yöntemlerinin Farklı Koşullarda Ölçme Duyarlılığına ve Test Güvenliğine Etkisi” başlıklı bu çalışma, jürimiz tarafından Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Ölçme ve Değerlendirme Programı'nda Doktora Tezi Olarak Kabul Edilmiştir.

İmza

Başkan: Prof. Dr. Nizamettin Koç



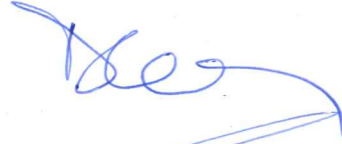
Üye: Prof. Dr. Duygu Anıl



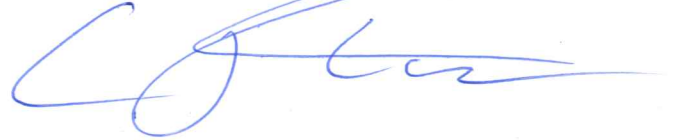
Üye: Doç. Dr. Şeref Tan



Üye: Yrd. Doç. Dr. H. Deniz Gülleroğlu



Üye: Yrd. Doç. Dr. C. Deha Doğan



ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eğitim – Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından/...../20..... tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca/...../20..... tarihinde kabul edilmiştir.

Prof. Dr. İsmail GÜVEN

Eğitim Bilimleri Enstitüsü Müdürü

ETİK BİLDİRİM

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.



Recep GÜR

ÖZET

BİREYE UYARLANMIŞ TESTLERDE MADDE KULLANIM SIKLIĞI KONTROL YÖNTEMLERİNİN FARKLI KOŞULLARDA ÖLÇME DUYARLILIĞINA ve TEST GÜVENLİĞİNE ETKİSİ

Gür, Recep

Doktora, Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanı: Yrd. Doç. Dr. Hamide Deniz Gülleroğlu

Kasım 2017, xiv + 98 Sayfa

Bu çalışmada, bilgisayar ortamında bireye uyarlanmış test (BOBUT) uygulamalarında iki kategorili (1-0) puanlanan maddelerde madde kullanım sıklığı kontrol yöntemlerinin farklı ölçme koşullarında (farklı yetenek dağılımlarında, test uzunluklarında ve örneklem büyüklüklerinde) ölçme duyarlılığına ve test güvenliğine etkisinin incelenmesi amaçlanmıştır. Monte Carlo simülasyon çalışması olan bu araştırma, kuramsal bir araştırma olup çalışma kapsamında kısa test 25 madde, uzun test 50 madde; küçük örneklem 250 birey, geniş örneklem 1000 birey olarak belirlenmiştir. Testi alan bireylerin yetenek parametreleri, her bir örneklem büyüklüğü koşulu için (θ) değeri -3 ve +3 aralığında sırasıyla sağa çarpık, normal, tek biçimli ve sola çarpık dağılım gösterecek şekilde BOBUT örneklemeleri oluşturulmuştur. Araştırmada kullanım sıklığının kontrol edilmediği durum referans olarak alınmış ve madde kullanım sıklığı kontrol yöntemlerinden Sympson Hetter stratejisi ve Aşamalı Düşürme yöntemi tercih edilmiştir. Araştırma kapsamında yer alan değişkenler ve koşullar çaprazlandığında 48 simülasyon koşulu belirlenmiş ve her bir koşulda 100 tekrar yapılarak 4800 veri dosyası oluşturulmuştur. Her bir koşul için ölçme duyarlılığı ve test güvenliği indeksleri hesaplanıp birbirleriyle karşılaştırılmıştır. Yapılan analizler sonucunda, farklı madde kullanım sıklığı kontrolü koşullarında ölçme duyarlılıkları arasında genel olarak büyük farklılık bulunmamıştır. İlgili koşullarda madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme yöntemi seçildiğinde, test güvenliği daha iyi sağlanarak madde kullanım sıklığı dengeli dağılım gösterdiğinden madde havuzu kullanımının daha verimli hale geldiği ve madde havuzu ifşâ olmadan birçok uygulama için süreklilik sağlayabildiği saptanmıştır. Bu doğrultuda, BOBUT

uygulamalarında farklı ölçme koşullarında, madde kullanım sıklığı kontrol yöntemlerinden BOBUT uygulamasından önce, tekrarlı simülasyonlarla kullanım sıklığı kontrol parametrelerinin belirlendiği, BOBUT uygulamalarında en yaygın kullanılan Sympson Hetter stratejisi yerine, BOBUT uygulanma sürecinde kullanım sıklığı kontrol parametrelerinin eş zamanlı olarak belirlendiği Aşamalı Düşürme yöntemi seçildiğinde, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin daha iyi sağlandığı sonucuna ulaşılmıştır.

Anahtar Kelimeler: Bilgisayar Ortamında Bireye Uyarlanmış Test, Madde Kullanım Sıklığı Kontrol Yöntemi, Sympson Hetter yöntemi, Aşamalı Düşürme Stratejisi

SUMMARY

THE EFFECT OF ITEM EXPOSURE CONTROL METHODS ON MEASUREMENT PRECISION AND TEST SECURITY WITHIN DIFFERENT CONDITIONS IN ADAPTIVE TESTING

Gür, Recep

Ph.D., Department of Measurement and Evaluation

Advisor: Assist. Prof. Dr. Hamide Deniz Gülleroğlu

November 2017, xiv + 98 Pages

In this research, it was aimed to investigate the effect of item exposure control methods on measurement precision and test security for dichotomous items within different conditions (different ability distributions, test lengths, sample sizes) in computerized adaptive testing applications. Current research which is designed as a Monte Carlo simulation research is a theoretical research. Within the scope of the research, short test was designed to have 25 items while long test was designed to have 50 items. In addition to this, small sample was simulated to have 250 observations whereas large sample was simulated to have 1000 observations. Regarding the test-takers' ability parameters, computerized adaptive testing samples which have normal, right-skewed, left-skewed and uniform distribution for each sample size condition in which ability parameters were within -3 and +3 range were generated. The condition in which item exposure wasn't controlled for was taken as the reference condition and Simpson Hetter and Fade Away methods were compared with the reference condition. When the variables and conditions within the research were crossed, 48 simulation conditions were established and 4800 data files were generated by doing 100 replications for each condition. Measurement precision and test security indices computed for each condition were compared to one another. The analyses conducted indicate that measurement precision indices corresponding to different item exposure control conditions didn't create substantial differences. It was found out that when Fade Away method was used, item exposure was distributed in a balanced way and higher test security was obtained. As a consequence of this, it was concluded that item pool got

more effective and sustainable without revelation (disclosure) of the items. Accordingly, it was discovered that in computerized adaptive testing applications in different conditions, instead of Simpson Hetter method which is frequently used before computerized adaptive testing applications and in which item exposure control parameters are estimated by iterative simulations, the use of Fade Away method which ensures the item exposure control parameters to be estimated simultaneously resulted in higher test security without reducing the measurement precision.

Keywords: Computerized Adaptive Test, Item Exposure Control Method, Simpson Hetter Method, Fade Away Method



ÖNSÖZ

Bilgisayar, yazılım ve psikometri alanındaki gelişmelerle birlikte dünyada Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) uygulamaları gün geçtikçe yaygınlaşmaktadır. BOBUT uygulamalarından elde edilen puanlar bireyler hakkında verilen önemli kararlara yön vermektedir. Dolayısıyla BOBUT uygulamalarının geçerli ve güvenilir olarak gerçekleştirilmesi, alınan kararların doğruluğunu etkilemektedir.

BOBUT uygulamalarında, bireyler yetenek düzeylerine en uygun maddelerle karşılaştıkları için, özellikle geniş ölçekli testlerde, bir bireyin istediği sayıda ve istediği sıklıkta testi almasına izin verilebilmektedir. Bu tür uygulamalarda, bir sınava birden çok kez katılan bireye, aynı maddeler çok sık uygulandığında maddelerin ifşâ olması, maddelerin psikometrik özelliklerinin yapay olarak düşmesine neden olmaktadır. Bu durum BOBUT uygulamalarının geçerliğini ve güvenilirliğini etkileyecek faktörlerden, test güvenliği ve ölçme duyarlılığına ilişkin tartışmalara yol açmıştır. BOBUT uygulamalarında, madde havuzunun dengeli bir şekilde kullanılmayıp sınırlı sayıda bazı maddelerin kullanılması, sık kullanılan maddelerin ifşâ edilmesiyle psikometrik özelliklerini kaybetmesi, madde havuzunun işlevselliğini kaybederek kullanılamaz hale gelmesi gibi etkenlerden dolayı BOBUT'un temel bileşenleri arasında, madde kullanım sıklığı kontrolü yöntemleri de yer almaya başlamıştır.

Bu araştırmada, BOBUT uygulamalarında iki kategorili (1-0) puanlanan çoktan seçmeli maddelerde madde kullanım sıklığı kontrol yöntemlerinin farklı ölçme koşullarında (farklı yetenek dağılımlarında, test uzunluklarında ve örneklem büyüklüklerinde) ölçme duyarlılığına ve test güvenliğine etkisi üzerine çalışma yapılması gerekli bulunmuştur.

Bu araştırma sonuçlarının, farklı örneklem büyüklükleri, test uzunlukları ve yetenek dağılımları birlikte ele alındığında, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanıp test güvenliğini sağlayan işlevsel madde kullanım sıklığı kontrol yöntemi belirlenerek hem daha geçerli ve güvenilir BOBUT uygulamalarının yapılmasına hem de alanyazındaki söz konusu eksikliğin giderilmesi açısından diğer çalışmalara katkıda bulunması beklenmektedir. Ayrıca çalışmanın incelenilen değişkenler ve koşullar bakımından Türkiye'de yapılan ilk araştırma olarak BOBUT uygulamaları sürecinde farkındalık oluşturabileceği düşünülmektedir.

Bu çalışmanın her aşamasında, tecrübesi ve geribildirimleriyle desteğini hiçbir zaman esirgemeyen, güler yüzlü ve yardımsever yaklaşımlarıyla yol gösteren danışmanım Sayın Hocam Yrd. Doç. Dr. Deniz Gülleroğlu'na teşekkürü bir borç bilirim.

Tez sürecinin belirli aşamalarında önerileriyle yol gösteren tez izleme komitesinin değerli üyeleri Sayın Hocalarım Doç. Dr. Şeref Tan'a ve Yrd. Doç. Dr. Deha Doğan'a teşekkür ederim. Ayrıca önerileriyle çalışmama katkı sağlayan değerli jüri üyeleri Sayın Hocalarım Prof. Dr. Nizamettin Koç'a ve Prof. Dr. Duygu Anıl'a teşekkür ediyorum.

Yaşamım boyunca desteklerini esirgemeyen, iyi günde kötü günde her zaman yanımda olan anneme, babama, kardeşlerime ve biricik eşime; ayrıca üzerimde emeği bulunan tüm öğretmenlerime teşekkürlerimi sunarım.

Doktora eğitimim sürecinde 2211-E Doğrudan Yurt İçi Doktora Burs Programı kapsamında verdiği maddi destekten dolayı TÜBİTAK'a teşekkürlerimi iletiyorum.

Recep Gür

İÇİNDEKİLER

ONAY.....	ii
ETİK BİLDİRİM.....	iii
ÖZET.....	iv
SUMMARY	vi
ÖNSÖZ.....	vii
İÇİNDEKİLER.....	ix
ÇİZELGELER DİZİNİ.....	xii
GRAFİKLER DİZİNİ	xiii
BÖLÜM I	1
1. GİRİŞ.....	1
1.1. Problem.....	1
1.1.1 Madde Kullanım Sıklığı Kontrol Yöntemleri	6
1.1.2 BOBUT Uygulamalarına Yönelik Yapılan Çalışmalar.....	9
1.1.3. Madde Kullanım Sıklığı Kontrol Yöntemleriyle İlgili Yapılan Çalışmalar....	9
1.2. Amaç.....	20
1.3. Araştırmanın Önemi	21
1.4. Araştırmanın Sınırlılıkları.....	23
1.5 Kısaltmalar.....	23
BÖLÜM II.....	24
2. YÖNTEM	24
2.1. Araştırmanın Modeli.....	24
2.2. Veriler	24
2.3. Verilerin Üretilmesi	25
2.4. BOBUT Koşulları	27
2.5. Verilerin Analizi	28

2.5.1. Farklı Ölçme Koşullarında Ölçme Duyarlılığına İlişkin Veri Çözümleme Yöntemleri	29
2.5.2. Farklı Farklı Ölçme Koşullarında Test Güvenliğine İlişkin Veri Çözümleme Yöntemleri	31
BÖLÜM III.....	35
3. BULGULAR VE YORUMLAR	35
3.1. BOBUT Uygulamalarında Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Ölçme Duyarlılığına Etkisi	35
3.2. BOBUT Uygulamalarında Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Test Güvenliğine Etkisi.....	44
BÖLÜM IV	63
4. SONUÇLAR VE ÖNERİLER	63
4.1. Sonuçlar	63
4.1.1. Farklı Ölçme Koşullarında Ölçme Duyarlılığına İlişkin Sonuçlar.....	63
4.1.2. Farklı Ölçme Koşullarında Test Güvenliğine İlişkin Sonuçlar.....	65
4.2. Öneriler	69
KAYNAKÇA	71
EKLER	80
EK 1. BOBUT Uygulamasına ve Sympson Hetter Örnekleme İlişkin Kümülatif ve Histogram Grafikleri.....	80
EK 2. Madde Havuzuna İlişkin Yanıt ve Bilgi Fonksiyonu Grafikleri	82
EK 3. Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Oranlarının Dağılımı.....	83
EK 4. Madde Kullanım Sıklığı Kontrol Yöntemi Kullanılmadığında Madde Kullanım Sıklıklarının Dağılımı	85
EK 5. Sympson Hetter Stratejisi Kullanıldığında Madde Kullanım Sıklıklarının Dağılımı	88
EK 6. Aşamalı Düşürme Stratejisi Kullanıldığında Madde Kullanım Sıklıklarının Dağılımı	91
EK 7. Test Örtüşme İndeksleri Arasındaki Farklara İlişkin z Değerleri	93

EK 8. Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Uyum Katsayılarının r' Dönüştürülmüş Değerleri	94
EK 9. Yanlılık Değerlerine İlişkin Mahalanobis Uzaklıklarının Hesaplanmasında Kullanılan Syntax	95
EK 10. RMSE Değerlerine İlişkin Mahalanobis Uzaklıklarının Hesaplanmasında Kullanılan Syntax	96
EK 11. BOBUT Uygulamalarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Farklı Ölçme Koşullarında Ölçme Duyarlılığına Etkisi	97
EK 12. BOBUT Uygulamalarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Farklı Ölçme Koşullarında Test Güvenliğine Etkisi.....	98



ÇİZELGELER DİZİNİ

Çizelge 1 <i>Araştırma Kapsamında Belirlenen BOBUT Bileşenleri ve Koşullar</i>	28
Çizelge 2 <i>Farklı Ölçme Koşullarına İlişkin Uyum Katsayıları</i>	35
Çizelge 3 <i>Farklı Ölçme Koşullarına İlişkin Cohen q Değerleri</i>	36
Çizelge 4 <i>Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Yanlılık Değerleri</i>	39
Çizelge 5 <i>Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin RMSE Değerleri</i>	41
Çizelge 6 <i>Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Madde Kullanım Sıklığı Oranlarının Standart Sapma Değerleri</i>	45
Çizelge 7 <i>Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Maksimum Madde Kullanım Sıklığı Oranı Dağılımları</i>	48
Çizelge 8 <i>Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Madde Havuzundan Yararlanmama İndeksleri</i>	50
Çizelge 9 <i>Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Madde Kullanım Sıklık Oranlarının Çarpıklık Katsayıları (χ^2 Değerleri)</i>	52
Çizelge 10 <i>Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Test Örtüşme İndeksleri</i>	55
Çizelge 11 <i>Farklı Ölçme Koşullarında Madde Havuzunun Kullanımına İlişkin Dağılım</i>	58

GRAFİKLER DİZİNİ

Grafik 1. <i>BOBUT Uygulamasında Yetenekleri Normal Dağılım Gösteren Küçük Örneklemeye İlişkin Histogram Grafiği</i>	82
Grafik 2. <i>BOBUT Uygulamasında Yetenekleri Normal Dağılım Gösteren Küçük Örneklemeye İlişkin Kümülatif Grafik</i>	82
Grafik 3. <i>Sympson Hetter Örneklemeye İlişkin Histogram Grafiği</i>	83
Grafik 4. <i>Sympson Hetter Örneklemeye İlişkin Kümülatif Grafik</i>	83
Grafik 5. <i>Madde Havuzuna İlişkin Yanıt Fonksiyonu Grafiği.</i>	84
Grafik 6. <i>Madde Havuzuna İlişkin Bilgi Fonksiyonu Grafiği</i>	84
Grafik 7. <i>Küçük örneklem, normal yetenek dağılımı, kısa test ve madde kullanım sıklığı kontrolü yapılmadığında madde kullanım sıklığı oranlarının dağılımı</i>	85
Grafik 8. <i>Küçük örneklem, normal yetenek dağılımı, kısa test ve SH Stratejisi kullanıldığında madde kullanım sıklığı oranlarının dağılımı</i>	85
Grafik 9. <i>Küçük Örneklem, Normal Yetenek Dağılımı, Kısa Test ve Aşamalı Düşürme Stratejisi Kullanıldığında Madde Kullanım Sıklığı Oranlarının Dağılımı</i>	86
Grafik 10. <i>Madde Havuzundan Yararlanmamaya İlişkin Dağılım</i>	87
Grafik 11. <i>Madde Havuzunun Kullanılma Derecesine İlişkin Dağılım</i>	87
Grafik 12. <i>Hedeflenen Kullanım Sıklık Oranından ($r_i \leq .20$) Düşük Oranda Kullanılan Madde Sayısının Dağılımı</i>	88
Grafik 13. <i>Hedeflenen Kullanım Sıklık Oranından ($r_i > .20$) Yüksek Oranda Kullanılan Madde Sayısının Dağılımı</i>	88
Grafik 14. <i>Hedeflenen Kullanım Sıklık Oranından ($r_i = .20$) Yüksek Oranda Madde Kullanım Sıklıklarına İlişkin Çarpıklık Katsayıları Dağılımı</i>	89
Grafik 15. <i>Madde Havuzundan Yararlanmamaya İlişkin Dağılım</i>	90
Grafik 16. <i>Madde Havuzunun Kullanılma Derecesine İlişkin Dağılım</i>	90
Grafik 17. <i>Hedeflenen Kullanım Sıklık Oranından ($r_i \leq .20$) Düşük Oranda Kullanılan Madde Sayısının Dağılımı</i>	91
Grafik 18. <i>Hedeflenen Kullanım Sıklık Oranından ($r_i > .20$) Yüksek Oranda Kullanılan Madde Sayısının Dağılımı</i>	91

Grafik 19. Hedeflenen Kullanım Sıklık Oranından ($r_i = .20$) Yüksek Oranda Madde Kullanım Sıklıklarına İlişkin Çarpıklık Katsayıları Dağılımı	92
Grafik 20. Madde Havuzundan Yararlanmamaya İlişkin Dağılım	93
Grafik 21. Madde Havuzunun Kullanılma Derecesine İlişkin Dağılım	93
Grafik 22. Hedeflenen Kullanım Sıklık Oranından ($r_i \leq .20$) Düşük Oranda Kullanılan Madde Sayısının Dağılımı	94



BÖLÜM I

1. GİRİŞ

Bu bölümde; araştırmanın problemi, amacı, önemi, sınırlılıkları ve kısaltmalar yer almaktadır.

1.1. Problem

Bireylerin psikolojik özelliklerini ve davranışlarını ölçmek amacıyla kullanılan veri toplama araçlarından elde edilen sonuçlar, onların yaşamında önemli bir yere sahiptir. Ölçme sonuçlarının, eğitim ve mesleki alanlara ilişkin önemli kararlarda kullanılabilmesi için, ölçme araçlarından elde edilen puanların geçerli ve güvenilir olması gerekmektedir. Bu gereklilik, araştırmacıları çeşitli kuramlar ve kuramlara dayalı yöntemler geliştirmeye yöneltmiştir.

Eğitimde ve psikolojide kullanılan testler, psikolojik yapıların anlaşılması ve yapılar arası ilişkilerin saptanmasını sağlarken; öğrenciler hakkında öğrenme güçlüklerinin teşhisi, seçme ve yerleştirme, öğrenci başarısının belirlenmesi gibi farklı amaçlarla kullanılmaktadır (Baykul, 2010). Bu testlerden elde edilen puanlar, bireyler hakkında verilen kararların doğruluğunu etkilemektedir. Dolayısıyla eğitimde ve psikolojide birçok örtük özellik doğrudan gözlenemediğinden ilgili değişkenleri ölçme işleminin geçerli ve güvenilir olması için araştırmacılar tarafından iki ölçme kuramı geliştirilmiştir. Bunlar klasik test kuramı (KTK) ve madde tepki kuramı (MTK)'dır. Test geliştirme aşamasında, madde ve test istatistiklerini kestirirken bu kuramlardan yararlanılmaktadır (Hambleton ve Swaminathan, 1985; Lord, 1980).

Klasik Test Kuramı ve Madde Tepki Kuramı'nın her ikisinin de ortak amacı bireyin örtük özelliğini belirlemek iken; KTK ve MTK, bireylerin yanıtları ve yetenek

düzeyleri arasındaki bağıntı bakımından farklılık göstermektedir (Wetzel, Hell ve Passler, 2011). KTK’da gerçek puan bazı varsayımlarla gözlenen puandan kestirilir. Dolayısıyla bu kurama “gerçek puan kuramı”da denir. Bu kuramda bireyin gözlenen özelliğinde, hem hata puanının hem de gerçek puanın bulunduğu kabul edilmektedir. KTK’da bireyin ölçekten aldığı ham puan ölçülmek istenen özelliğe ilişkin puanı ifade etmekte iken, ham puanların standartlaştırılmış değerleri ise bireyin yetenek ölçüsüne karşılık gelmektedir (Baykul, 2010).

KTK’da bireylerin yetenekleri her ne kadar sabit kalsa da testin güçlük düzeyi arttıkça, bireyler daha düşük gözlenen puana sahip olabilmektedirler (Hambleton ve Jones, 1993). KTK’nın bu sınırlılığının üstesinden gelebilmek amacıyla Madde Tepki Kuramı (MTK) geliştirilmiştir. KTK’da gerçek puan, gözlenen puandan kestirilirken; MTK, bireyin yetenek düzeyi ile herhangi bir maddeyi yanıtlayabilme olasılığı arasındaki ilişkiye odaklanmaktadır (Embretson ve Reise, 2000). Hambleton ve Swaminathan (1985) bu olasılığın önemine vurgu yaparak, bireyin ilgili örtük özelliğinin (başarı, tutum vb.), maddeyi yanıtlama olasılığını yordayabileceği bunun yanı sıra bireyin ilgili örtük özelliği ile madde performansı arasındaki ilişkinin monotonik artan bir fonksiyonla (madde karakteristik eğrisi) gösterilebileceğini belirtmekte ve MTK’nın bu iki temel dayanak üzerine kurulu olduğunu vurgulamaktadır. Bu temel dayanaklar doğrultusunda, MTK’nın testten ziyade maddeye odaklandığı ifade edilebilir. MTK’da, katılımcı yanıtları ve katılımcıların yetenek düzeyleri arasındaki ilişki, doğrusal olmayan matematiksel bir fonksiyon ile açıklanmaktadır (Lord, 1980).

KTK ve MTK arasında önemli farklılıklar bulunmaktadır. KTK’ya göre, test geliştirirken yaygın olarak ele alınan madde güçlüğü ve ayırt ediciliği, elde edildiği örnekleme göre değişiklik göstermektedir. Bir başka ifadeyle, madde ve test istatistiklerinin değeri, grubun performansına bağlıdır (Embretson ve Reise, 2000). Bu durumda, ortalama yetenek düzeyi ve yetenek puanlarının ranjı, puanların elde edildiği örneklemden etkilenmektedir (Hambleton ve Swaminathan, 1985). Dolayısıyla, test ve madde istatistiklerinin, uygulanan gruba bağımlı olması, bireyin yeteneklerinin maddelere bağımlı olması, maddenin doğru yanıtlanma olasılığıyla ilgili kestirim imkânı sunmaması, KTK’nın sınırlılıkları olarak ön plana çıkmaktadır (De Ayala, 2009).

MTK modellerinin varsayımları (tek boyutluluk, yerel bağımsızlık, testin hız testi olmaması, model veri uyumu ve madde karakteristik eğrisinin (MKE) yapısı ve değişmezliği) sağlandığında, KTK'nın sınırlılıklarına alternatif olarak birçok avantajı bulunmaktadır (Embretson ve Reise, 2000; Hambleton, Swaminathan ve Rogers, 1991; Orlando ve Thissen, 2000). MTK'nın, KTK'ya göre en önemli avantajı, madde parametreleri kestirimlerinin bir cevaplayıcı grubundan diğerine; yetenek parametrelerinin kestirimlerinin de bir grup madde örnekleminde diğerine değişmemesidir (Crocker ve Algina, 1986). Bu değişmezlik özelliği, maddelerin MTK'ya göre bir kez ölçeklendikten sonra, diğer uygulamalarda da kullanılmasına olanak sağlamaktadır (Hambleton ve diğerleri, 1991; Kelecioğlu, 2001).

Madde Tepki Kuramı, eğitim ve psikoloji alanında yetenek ve başarının ölçülmesinde sıklıkla kullanılmaktadır. Çünkü MTK, bireyin örtük özelliğini, test maddelerine verdiği yanıtlardan kestirmekte ve temelde bireylerin ilgili örtük özellikleri ile maddelere verdikleri yanıtlar arasındaki bağıntının matematiksel olarak gösterilebilmesine imkân sunmaktadır (Embretson ve Reise, 2000; Thissen ve Steinberg, 2009). Eğer örtük değişkenlerin (yetenek, tutum vs.) ölçümü ile ilgileniliyorsa MTK modellerinin kullanılması her zaman daha uygun olmaktadır. Ayrıca Hambleton ve Swaminathan (1985) MTK'nın madde havuzu oluşturma, bireye uyarlanmış test geliştirme, madde yanlılığının belirlenmesi ve test eşitleme konularında karşılaşılan sorunlara çözüm getirdiğini vurgulamaktadır.

MTK'nın sağladığı avantajlardan biri de bireye uyarlanmış (adaptive) test geliştirmeye imkân sunmasıdır. Binet ve Simon tarafından 1905 yılında geliştirilen Binet zekâ testi, bireye uyarlanmış testin tipik özelliklerini göstermekle birlikte ilk bireye uyarlanmış test uygulaması olarak kabul edilmektedir (Weiss, 1988). Bireye uyarlanmış test uygulamasının etkin olabilmesi, eşzamanlı hesaplamalar ve bireyin yetenek düzeyine en uygun maddelerle hızlıca karşılaşabilmesini gerektirdiğinden, bilgisayarların ucuzlaması ve teknolojideki gelişmelerle birlikte 1970'li yıllarda, bireye uyarlanmış test uygulamalarının yapılabilmesi mümkün hale gelmiştir. Böylece “Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) Uygulaması” ortaya çıkmıştır (Çıkrıkçı-Demirtaşlı, 1999).

Madde Tepki Kuramı'nın ölçme yaklaşımı bakımından avantajlar sağladığı ölçme uygulamalarından biri olan BOBUT, geleneksel kâğıt kalem uygulaması yerine

bilgisayar uygulamasına dayanan bir ölçme yaklaşımıdır. Klasik testlerden farklı olarak bilgisayar ortamında uygulanabilen bireye uyarlanmış testler, yanıtlayıcıların bilgisayarlarla yönetildiği ölçme araçlarıdır. BOBUT uygulamaları, MTK'nın değişmezlik özelliğini kullanarak, her birey için nitelikli ölçme yapabilen maddelerin madde havuzundan seçilerek bireye sunulmasını sağlayan bir algoritma ile çalışmaktadır (Embertson ve Reise, 2000; Way, 2005).

BOBUT uygulamalarında, maddeler, bireyin yetenek (θ :theta) düzeyi ile eşleşecek şekilde seçilmektedir. Bu süreç çeşitli yöntemler kullanılarak farklı şekilde işleyebilmektedir. Genellikle ilk adımda, madde havuzundaki maddelerden orta derecede zorluğa sahip olan bir madde seçilerek bireyden bu maddeye yanıt vermesi beklenmektedir. Verilen yanıtı göre, bireyin θ düzeyine ilişkin kestirim elde edildikten sonra, madde havuzundaki hangi maddenin, bireyin θ kestirimi için en fazla bilgi sağlayacağına karar verilmektedir. BOBUT uygulamalarında madde seçiminin temel mantığı, bireyin doğru yanıtladığı maddeden sonra daha zor, yanlış yanıtladığı maddeden sonra ise daha kolay maddelerle karşılaşmasını sağlamaktır. Bireylerin her maddeye verdiği yanıtın sonra, yetenek düzeyleri tekrar hesaplanır ve yeni bir θ kestirimi yapılır. Farklı sonlandırma kuralları gerçekleştiğinde test sona ermektedir (Bulut ve Kan, 2012; Lord ve Stocking, 1988).

Bireylere kendi yetenek düzeylerine göre çok kolay bir ölçme aracı uygulamak, zaman kaybının yanı sıra dikkatsizlik nedeniyle hata yapılması gibi istenmeyen birey davranışlarına neden olabilmektedir. Diğer taraftan birey için çok zor olan bir testi uygulamak da bilgi verici olmayan sonuçlar ortaya çıkaracaktır. Çünkü bireyler maddeleri yanıtlamak için istenmeyen davranışlara (kopya, şans başarısı vb.) yönelerek doğru yanıtları bulmaya çalışacaklardır. Bu nedenle bireyin ölçülen özellikteki düzeyine bağlı olarak bireylere test uygulamak, daha güvenilir ve geçerli ölçme sonuçları elde etmeye olanak sağlamaktadır (Kalender, 2011; Öztuna, 2008).

BOBUT uygulamalarında, bireyin maddelere verdiği yanıtlara bağlı olarak bireylerin ölçülen psikolojik özellik düzeyi belirlenebilmektedir. Yeni maddeler bireyin ilgili yetenek düzeyine en uygun olacak şekilde madde havuzundan seçilerek, bireyin bu maddeleri alması sağlanmaktadır (Weiss, 2004). Teorik olarak BOBUT uygulamalarında, iteratif bir test süreci mantığı olmasına rağmen, uygulamada hız

faktörü, maddelerin ifşâ olmaması, kapsam geçerliğinin sağlanması gibi çeşitli gerekliliklerden dolayı kısıtlamalar yapılabilmektedir (Weiss, 1988).

BOBUT uygulamalarından elde edilen puanlar bireyler hakkında verilen önemli kararlara yön vermektedir. Dolayısıyla BOBUT uygulamalarının geçerli ve güvenilir olarak gerçekleştirilmesi, alınan kararların doğruluğunu etkilemektedir. Bilgisayar, yazılım ve psikometri alanındaki gelişmelerle birlikte BOBUT uygulamalarının yaygınlaşmaya başlamasıyla BOBUT uygulamalarının geçerliğini ve güvenilirliğini etkileyecek faktörlerden, test güvenliği (test security) ve ölçme duyarlılığı (measurement precision) gündeme gelmektedir (Boyd, Dodd ve Fitzpatrick, 2013; Weiss, 2004).

BOBUT uygulamalarında, test güvenliğini ve ölçme duyarlılığını olumsuz yönde etkileyecek durumlara ilişkin, madde kullanım sıklığı kontrolü (item exposure control) ile BOBUT algoritmasında kısıtlamalar yapılabilmektedir (Han, 2009). Madde havuzunun kalibrasyonu, başlatma ve sonlandırma kuralları, madde seçme ve yetenek kestirim yöntemleri, içerik dengelenmesinin (content balancing) yanı sıra madde kullanım sıklığı kontrolünün de BOBUT'un temel bileşenlerinden biri olduğu vurgulanmaktadır (Boyd, 2003; Segall, 2004; Magis ve Raîche, 2012).

BOBUT uygulamalarında, bireyler yetenek düzeylerine en uygun maddelerle karşılaştıkları için, özellikle geniş ölçekli testlerde, bir bireyin istediği sayıda ve istediği sıklıkta testi almasına izin verilebilmektedir. Bu tür uygulamalarda, bir sınava birden çok kez katılan bireye, aynı maddeler çok sık uygulandığında (over exposure) maddelerin ifşâ edilmesi maddelerin psikometrik özelliklerinin yapay olarak düşmesine neden olmaktadır (Segall, 2004; Revuelta ve Ponsoda, 1998). Bu durum ölçme sonuçlarının geçerliğini ve test güvenliğini olumsuz yönde etkilemektedir (Georgiadou, Triantafillou ve Economide, 2007; Lee ve Dodd, 2012). Örneğin, bir sınava birden çok kez katılan bireyin aynı maddelerle karşılaşması durumunda, ilgili maddelerin ifşâ olmasının bireyin kestirilen yetenek düzeyi üzerinde etkisinin olup olmadığı belirlenememektedir.

Madde bilgi fonksiyonu ile yetenek kestiriminin standart hatası arasında ters yönde ilişki bulunmaktadır. Madde bilgi fonksiyonu, BOBUT uygulamalarında madde seçimi için ölçüt alındığından yüksek bilgi değerine sahip maddelerin

seçilmesi, yetenek kestiriminin standart hatasını düşürerek ölçme duyarlılığını artırmaktadır (Kalender, 2009). Fakat ölçme duyarlılığını maksimum yapacak maddeler seçildiğinde, kullanım sıklığı oranları düzgün dağılmayacağından madde havuzundaki bazı maddeler çok sık uygulanırken, bazı maddeler ise hiçbir bireye uygulanmamaktadır (Pastor, Dodd ve Chang, 2002). Bu durum, madde havuzunun dengeli bir şekilde kullanılmayıp sadece sınırlı sayıda bazı maddelerin kullanılmasına sebep olmaktadır (Han, 2009). Bu nedenle, madde kullanım sıklığı kontrol yöntemleriyle, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin sağlanması hedeflenmektedir (Boyd, 2003; Boyd, Dodd ve Fitzpatrick, 2013; Davis ve Dodd, 2005; Pastor, Dodd ve Chang, 2002).

Bireysel farklılıkları gözeterek, bireye kendi yetenek düzeyine en uygun maddelerin seçilmesiyle gerçekleştirilen BOBUT uygulamaları için geniş madde havuzuna ihtiyaç duyulmaktadır (Embretson ve Reise, 2000; Magis ve Raîche, 2012). Çok sayıda yüksek derecede ayırt ediciliğe sahip, her bir yetenek düzeyine hitap edebilecek güçlük düzeyindeki maddelerden oluşan madde havuzlarıyla BOBUT uygulamaları daha iyi sonuçlar vermektedir (Veldkamp ve van der Linden, 2010; Weiss, 2004). Dolayısıyla madde havuzunun dengeli bir şekilde dağılım göstermeyip sadece sınırlı sayıda bazı maddelerin kullanılması, madde havuzu geliştirmek için harcanan zaman ve işgücünün iyi değerlendirilmemesine sebep olmaktadır (Aytuğ-Koşan, 2013). Belirtilen bu nedenlerle test güvenliğini sağlayabilmek, madde havuzu kullanımını daha verimli hale getirebilmek ve madde havuzunun işlevsel olarak sürekliliği için madde kullanım sıklığı kontrol yöntemleri geliştirilmiştir (Davis, 2002; Boyd, 2003; Revuelta ve Ponsoda, 1998).

1.1.1. Madde Kullanım Sıklığı Kontrol Yöntemleri

Madde kullanım sıklığı kontrol problemine ilişkin geliştirilen ilk yöntemlerden biri, 5-4-3-2-1 stratejisidir (McBride ve Martin, 1983). Tesadüfi seçme yöntemlerinden 5-4-3-2-1 stratejisi, BOBUT iterasyonundaki ilk adımda beş madde arasından, ikinci adımda dört madde arasından şeklinde, tesadüfi olarak seçilen bir maddenin uygulanma sıklığını kontrol etmeyi amaçlamaktadır. Kingsbury ve Zara (1989) ve Thomasson (1998) ise bütün maddelerin kullanım sıklık oranını azaltmak için farklı tesadüfi seçme yöntemleri geliştirmişlerdir (Akt. Veldkamp, Vershoor ve Eggen, 2007). Önceden belirlenmiş en çok bilgi veren on maddeden birinin rastgele

seçildiği Randlesque yöntemi (Kingsbury ve Zara, 1989); hedeflenen madde güçlük düzeyinin .10 logit aralığı ranjından rastgele maddelerin seçildiği .10 logit aralığı (within .10 logit) yöntemi (Lunz ve Stahl, 1998), test ilerledikçe madde seçme üzerindeki tesadüfi bir bileşenin etkisini azaltarak madde bilgisinin öneminin belirgin bir şekilde artırılmasının amaçlandığı İlerlemeli yöntem [Progressive Method-(İY)] (Revuelta ve Ponsoda, 1996) tesadüfi seçme yöntemlerindendir. Tesadüfi seçme yöntemlerinde ortak amaç, maksimum bilginin en ideal olduğu aralıktaki maddeler arasından rastgele seçilen bir maddenin uygulanma sıklığını kontrol etmektir (Akt. Georgiadou, Triantafillou ve Economide, 2007).

Koşullu seçme yöntemlerinde, BOBUT uygulamasından önce, tekrarlı simülasyonlarla belirlenen kullanım sıklığı kontrol parametresi kullanılarak maddenin kullanım sıklık oranı kontrol edilmektedir. Davey ve Parshall yöntemi, Stocking ve Lewis çok terimli (mutinomial) yöntemi, hedeflenen (targeted) kullanım sıklığı kontrol yöntemi, Sınırlandırılmış Maksimum Bilgi (Restricted Maximum Information-[SMB]) yöntemi gibi koşullu seçme yöntemleri bulunmaktadır. Davey ve Parshall (1995) yönteminde tek tek maddelerin aşırı kullanımını kısıtlamakla kalmayıp aynı zamanda bireylerin aynı madde setleriyle karşılaşması engellenmeye çalışılırken; Stocking ve Lewis çok terimli yönteminde, her bir madde için her bir yetenek seviyesi sayısı (n) kadar madde kullanım sıklığı kontrol parametresi hesaplanmaktadır (Stocking ve Lewis, 1995). Hedeflenen kullanım sıklığı kontrol yöntemi, maddelerin aşırı kullanımını kontrol etmeye odaklanmak yerine kullanılmayan maddelerin uygulanma olasılıklarını artırmaya yöneliktir (Thompson, 2002). Koşullu seçme yöntemlerinden bir diğeri SMB yönteminde ise, hiçbir maddenin önceden belirlenmiş kullanım sıklık oranından daha fazlasının kullanılmasına izin verilmemektedir (Revuelta ve Ponsoda, 1998).

Dönüşümlü (rotating) madde havuzu yönteminde (Ariel, Veldkamp ve van der Linden, 2008; Way, 1998) maddelerin kullanım sıklık oranını azaltmaya yönelik madde havuzunu içerik ve madde parametreleri bakımından benzer dağılım gösterecek şekilde, önsel dağılım kullanılarak maddeler farklı testlerde uygulanmaktadır. Bu yöntemlerin yanı sıra, madde havuzunun tabakalandırılarak kullanılmasının amaçlandığı tabakalı yöntemler (a-tabakalama stratejisi (Chang ve Ying 1999), b parametresi blokeli a tabakalama (Chang, Qian ve Ying, 2001), içerik

blokeli a tabakalamalı BOBUT deseni (Yi ve Chang, 2001), 0-1 tabakalamalı strateji (Chang ve Van der Linden, 2003)) ve farklı yöntemler bir araya getirilerek birleştirilmiş yöntemler (İlrmeli sınırlandırılmış strateji (Revuelta ve Ponsoda, 1998); a tabakalama ile SH stratejisinin birleşimi (Leung, Chang ve Hau, 2003); içerik blokeli a tabakalamalı ile SH stratejisinin birleşimi (Yi, 2002)) geliştirilmiştir (Akt. Georgiadou, Triantafillou ve Economide, 2007). Farklı yöntemler bulunmasına rağmen, BOBUT uygulamalarında en yaygın kullanılan madde kullanım sıklığı kontrol yöntemi, koşullu seçme yöntemlerinden Sympson Hetter (SH) yöntemidir (Veldkamp, Vershoor ve Eggen, 2007).

BOBUT uygulaması öncesinde hedeflenen değer olarak maksimum kullanım sıklık oranı belirlenip madde seçildiğinde, bireyin o maddeyle karşılaşması, maddenin kullanım sıklığı kontrol parametresine bağlı olmaktadır (Davis ve Dodd, 2005). Örneğin, bütün maddeler için bu parametre .25 olarak tanımlanırsa, bir maddenin seçilmesi durumunda, yaklaşık olarak her dört bilgisayar ortamında bireye uyarlanmış testlerden birinde, ilgili madde uygulanabilmektedir (Weiss ve Guyer, 2012).

BOBUT uygulamalarında çok fazla seçilen i. maddenin kullanım sıklık oranının ($P_i(S)$), BOBUT uygulaması öncesinde belirlenen madde kullanım sıklığı kontrol parametresinden ($P_i(A|S)$) büyük olması ($P_i(A|S) < P_i(S)$), i. maddenin seçilmesi durumunda uygulanma olasılığını ($P_i(A)$) kısıtlarken; daha az kullanılan maddeler için madde kullanım sıklık oranı, madde kullanım sıklığı kontrol parametresinden daha küçük ($P_i(S) < P_i(A|S)$) olduğunda ilgili maddelerin uygulanma olasılığını artırmaktadır (Segall, 2004; Stocking ve Lewis, 2002). Böylelikle uygulanacak maddenin koşullu olasılığı $P_i(A|S)$ kontrol parametresi olarak tanımlanıp $P_i(A) = P_i(A|S) * P_i(S)$ eşitliğinde kullanılması, hedeflenen kullanım sıklığı oranına ulaşılabilme, seçilen maddenin uygulanıp uygulanmadığını kontrol etme ya da birden fazla BOBUT uygulamasına katılan bireyler için seçilen maddenin geçici olarak kullanılmamasına imkân vermektedir (Pastor, Dodd ve Chang, 2002; Veldkamp, Vershoor ve Eggen, 2007).

Han (2009) geliştirdiği yöntemle, kullanım sıklığı kontrol parametrelerinin belirlenmesinde, BOBUT uygulamasından önce, tekrarlı simülasyon yapılmasına gerek olmadığını, BOBUT uygulanma sürecinde eş zamanlı olarak kontrol parametrelerinin belirlenebileceğini vurgulamaktadır. Madde seçme sürecinde,

havuzdan seçilmesi uygun olan her bir madde için madde seçme ölçütü, hedeflenen kullanım sıklık oranı (e) ve hali hazırda/gözlenen kullanıma sıklık oranı (r_i) arasındaki oran ile ters bir şekilde ($I_i [\hat{\theta}_{m-1}]^{\frac{e-r_i}{e}}$) ağırlıklandırılmaktadır. Böylelikle $e > r_i$ ise, madde seçme ölçütü Maksimum Fisher Bilgisi (MFB) $I_i [\hat{\theta}_{m-1}]$ pozitif iken; $r_i > e$ ise, $I_i [\hat{\theta}_{m-1}]$ negatif olmaktadır (Han, 2012). Bu durumda, az kullanılan maddeler daha sık kullanılırken, fazla kullanılan maddelerde ise, seçilme oranı aşamalı düşürüleceğinden, Han tarafından bu yöntem Aşamalı Düşürme (AD) Yöntemi (fade-away method) olarak adlandırılmaktadır.

1.1.2. BOBUT Uygulamalarına Yönelik Yapılan Çalışmalar

BOBUT uygulamalarıyla ilgili yapılan çalışmalarda, genellikle geleneksel kâğıt kalem testi ile farklı stratejiler (farklı başlatma, sonlandırma kuralları, madde seçme ve yetenek kestirim yöntemleri) kullanılarak yapılan BOBUT uygulamalarıyla kestirilen yetenek düzeyleri arasında manidar bir ilişki bulunup bulunmadığı incelenmiştir (Bulut ve Kan, 2012; Cömert, 2008; Eroğlu, 2013; Gökçe, 2012; İşeri, 2002; Kalender, 2011; Kaptan, 1993; Kaskatı, 2011; Kezer, 2013; McDonald, 2002; Özbaşı, 2014; Öztuna, 2008; Scullard, 2007; Smits, Cuijper ve Straten, 2011; Sulak, 2013; Wang, Kuo, Tsai ve Laio, 2012; Zitny, Halama, Jelinek ve Kveton, 2012). İlgili alanyazında, BOBUT uygulaması sonucunda, kâğıt kalem testlerine benzer yetenek kestirimleri yapıldığı ve farklı BOBUT stratejileri kullanılarak yetenek düzeyleri kestirildiğinde benzer sonuçlar alınabileceği bulgusuna ulaşılmıştır. Ayrıca BOBUT uygulamalarının, kâğıt kalem testlerine göre ölçme duyarlılığını artırdığı ve yetenek kestirimi için gerekli madde sayıları dolayısıyla zaman bakımından da önemli ölçüde ekonomiklik sağladığı sonucuna ulaşılmıştır.

1.1.3. Madde Kullanım Sıklığı Kontrol Yöntemleriyle İlgili Yapılan Çalışmalar

BOBUT'un temel bileşenlerinden, madde kullanım sıklığı kontrol yöntemleriyle ilgili yurtdışında yapılan çalışmalar incelendiğinde, çoklu puanlanan MTK modellerine (Burt, Davis ve Dodd, 2003; Davis, 2002, 2004; Davis ve Dodd, 2005); farklı madde seçme yöntemlerine (Han, 2009, 2012); farklı madde ayırt edicilik dağılımlarına sahip madde havuzlarına (Revuelta ve Ponsoda, 1998), farklı ortalama güçlük düzeylerine sahip madde havuzlarına (Lee ve Dodd, 2012), farklı madde havuzu büyüklüklerine (Chang ve Twu, 1998; Pastor, Dodd ve Chang, 2002;

Revuelta ve Ponsoda,1998); farklı test sonlandırma kurallarına (French ve Thompson, 2003; Revuelta ve Ponsoda,1998) ve testlet (madde takımları) temelli BOBUT uygulamalarında farklı yetenek kestirimlerine (Boyd, 2003; Davis ve Dodd, 2003) göre madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı görülmüştür. İlgili çalışmalara aşağıda sırasıyla yer verilmiştir.

Çoklu puanlanan MTK modellerine göre madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı çalışmalar incelendiğinde, Burt, Davis ve Dodd (2003) araştırmasında çoklu puanlanan MTK modellerinden genelleştirilmiş kısmî bilgi modeline (Generalized Partial Credit Model) dayalı BOBUT uygulaması için Ulusal Eğitimde İlerleme Değerlendirmesi (National Assessment of Educational Progress, [NAEP]) 1996 fen testinden 210 madde seçilerek madde havuzu oluşturmuştur. Araştırma kapsamında başlatma kuralı $\theta=0$, en yüksek olabilirlik yetenek kestirim yöntemi (EYOY), sonlandırma kuralı olarak sabit test uzunluğu stratejisi (20 madde), içerik dengelenmesi yöntemlerinden Kingsbury ve Zara yöntemleri; madde kullanım sıklığı kontrol yöntemlerinden .10 logit aralığı, Bilgi Koşullu Tesadüfi Seçme yöntemi (3 ya da 6 aday madde) ve SH stratejisi kullanılmıştır. SH stratejisi kullanıldığında diğer madde kullanım sıklığı kontrol yöntemlerine göre maksimum madde kullanım sıklığı oranının daha düşük olduğu; Bilgi Koşullu Tesadüfi Seçme yöntemi seçilmesi durumunda ise, diğer madde kullanım sıklığı kontrol yöntemlerine göre ölçme duyarlılığının azaldığı ve madde havuzunun kullanılma derecesinin arttığı sonucuna ulaşılmıştır.

Davis (2002) çalışmasında, farklı çoklu puanlanan MTK modellerine (Kısmî bilgi modeline (Partial Credit Model [KBM]), genelleştirilmiş kısmî bilgi modeline ve aşamalı tepki modeline (Graded Response Model) dayalı BOBUT uygulamalarında, madde kullanım sıklığı kontrol yöntemlerinden .10 logit aralığı (3 ve 6 madde) yöntemi, Bilgi Koşullu Tesadüfi (3 ve 6 madde) yöntemi, SH stratejisi, koşullu Sympton Hetter stratejisi, a-tabakalama ve geliştirilmiş a- tabakalama yöntemlerini incelemiştir. Araştırmada .10 logit aralığı (6 madde) ve Bilgi Koşullu Tesadüfi (6 madde) yöntemlerinin ölçme duyarlılığı bakımından diğer yöntemlere göre en iyi madde kullanım sıklığı kontrol yöntemi olduğu; a-tabakalama ve geliştirilmiş a- tabakalama yöntemlerinin ise en kullanışsız yöntemler olduğu sonucuna ulaşmıştır. Ayrıca diğer yöntemlere göre uygulaması zor ve zaman alıcı bir yöntem olduğu

belirtilen SH stratejisinde her ne kadar BOBUT uygulamasından önce, hedeflenen kullanım sıklığı oranına ulaşılabilme için tekrarlı simülasyonlarla kullanım sıklığı kontrol parametreleri belirlense de madde örtüşme indeksi ve madde havuzunun kullanılma derecesi bakımından hedeflenen düzeyde madde kullanım sıklığı sonucuna ulaşılmamıştır.

Davis (2004), genelleştirilmiş kısmî bilgi modeline dayalı simule edilmiş bir dizi BOBUT uygulaması için altı farklı madde kullanım sıklığı kontrol yöntemini incelemiştir. Araştırma kapsamında kullanım sıklığının kontrol edilmediği durum referans olarak alınmış ve madde kullanım sıklığı kontrol yöntemlerinden Randomesque yöntemi, değiştirilmiş .10 logit aralığı yöntemi, SH, koşullu SH stratejisi, çok katmanlı a-tabakalama, çok katmanlı geliştirilmiş a-tabakalama yöntemi kullanım sıklığı oranlarını kontrol etmek için uygulanmıştır. Araştırma sonucunda, maksimum kullanım sıklık oranı bakımından koşullu SH stratejisi kullanıldığında, diğer madde kullanım sıklığı kontrol yöntemlerine göre daha düşük sonuçlar elde edilmesine rağmen, madde havuzunun kullanılma derecesi ve örtüşme indeksleri bakımından Randomesque ve değiştirilmiş .10 logit aralığı yöntemlerinin daha kullanışlı madde kullanım sıklığı kontrol yöntemi olduğu bulgusuna ulaşılmıştır.

Çoklu puanlanan MTK modellerinden kısmî bilgi modeline dayalı post-hoc BOBUT simülasyon çalışmasında Davis ve Dodd (2005) çoklu puanlanan geniş ölçekli teste katılım gösteren 157 öğrenciden elde edilen madde parametrelerini kullanmıştır. Davis ve Dodd, başlatma kuralı olarak $\theta=0$, MLE yetenek kestirim yöntemi, sonlandırma kuralı olarak 20 maddelik sabit test uzunluğu, içerik dengelenmesi yöntemlerinden Kingsbury - Zara yöntemi; madde kullanım sıklığı kontrol yöntemlerinden SH stratejisi, koşullu Sympton Hetter stratejisi, değiştirilmiş .10 logit aralığı (3 ve 6 madde) ve Kingsbury - Zara (3 ve 6 madde) yöntemini tercih etmiştir. Araştırma sonucunda, değiştirilmiş .10 logit aralığı ve Kingsbury - Zara yöntemi (6 madde) ile koşullu Sympton Hetter stratejisi arasında, madde kullanım sıklık oranı, örtüşme indeksi ve madde havuzunun kullanılma derecesi bakımından benzer sonuçlar gösterdiği bulgusuna ulaşılmıştır.

Farklı madde seçme yöntemlerine göre madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı çalışmalar incelendiğinde, Han (2009) Aşamalı Maksimum Bilgi Oranı (Gradual Maximum Information Ratio [AEYBO]) madde

seçme yönteminin ölçme duyarlılığı ve test güvenliği bakımından etkililiğini test etmek için GMAT (Graduate Management Admission Test) çoktan seçmeli matematik alt testine ait madde havuzundan 500 madde seçerek madde havuzu oluşturmuştur. Araştırma kapsamında $-.50 < b < .50$ aralığında başlatma kuralı, BSD yetenek kestirim yöntemi, sonlandırma kuralı olarak sabit test uzunluğu stratejisi (40 madde); madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme stratejisi kullanılırken; sonuçların genellenebilirliğini artırmak amacıyla içerik dengelenmesi yöntemleri kullanılmamıştır. Standart normal dağılımdan $N(0,1)$ 10.000 birey simüle edilmiştir. Günde iki zaman diliminde (time slot) olmak üzere her bir zaman dilimi için 250 bireye 40 madde uygulanarak toplam 20 günde simülatif BOBUT uygulaması gerçekleştirilmiştir. Her bir BOBUT uygulaması 100 kez tekrar edilmiştir. Çalışmada, beş farklı madde seçme yöntemi uygulanarak karşılaştırılmıştır. Birinci yöntemde kestirilen yetenek düzeyine en yakın b parametrelerine sahip beş maddeden birinin tesadüfi olarak seçilip uygulanması ve madde kullanım sıklık oranının .20 ile kısıtlanması ($r < .20$); ikinci yöntemde MFB ve $r < .20$; üçüncü yöntemde MFB ve AD Stratejisi; dördüncü yöntemde AEYBO ve $r < .20$; beşinci yöntemde ise AEYBO ve AD Stratejisi uygulanmıştır. Araştırma sonucunda, ölçme duyarlılığı bakımından AEYBO madde seçme yönteminin daha kullanışlı bir yöntem olduğu; AEYBO madde seçme yöntemin ve madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisinin birlikte kullanıldığında (beşinci yöntem) test güvenliğinin daha az zedelendiği bulgusuna ulaşılmıştır.

Han (2012) çalışmasında, madde kullanım sıklığı kontrol yöntemlerinden SH ve AD stratejisi kullanılarak madde seçme yöntemlerine (Etkili Dengelenmiş Bilgi (efficiency balanced information) yöntemi, MFB, b Eşleme (b Matching) yöntemi, b parametresinin bloke edildiği a -tabakalama yöntemi) göre, ölçme duyarlılığı ve test güvenliği indeksleri bakımından karşılaştırmıştır. Araştırma kapsamında $-.50 < b < .50$ aralığında başlatma kuralı, MLE yetenek kestirim yöntemi, madde kullanım sıklığı kontrolü yapılmaması koşulu ile madde kullanım sıklığı kontrol yöntemlerinden AD ve SH stratejisi kullanmanın yanı sıra sonlandırma kuralı olarak sabit test uzunluğu (30 madde) ve değişken test uzunluğu stratejisi ($SE(\theta) < .30$ ya da min.10 - maks. 40 madde) kullanılarak iki farklı simülasyon koşulu oluşturulmuştur. Yetenek düzeyleri normal dağılım gösteren 10.000 birey simüle edilmiştir. GMAT çoktan seçmeli matematik alt testi madde havuzundan 500 madde seçilerek madde havuzu

oluşturulmuş ve yükseköğretim için BOBUT uygulamasına benzer şekilde simülatif BOBUT uygulaması gerçekleştirilmiştir. Her bir BOBUT uygulaması 25 kez tekrar edilmiş ve elde edilen sonuçların ortalaması alınmıştır. Araştırmada Han tarafından geliştirilen Etkili Dengelenmiş Bilgi (efficiency balanced information) madde seçme yönteminin madde kullanım sıklığı kontrol yöntemlerinden AD stratejisiyle birlikte kullanıldığında, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin diğer yöntemlere göre daha iyi sağlandığı sonucuna ulaşılmıştır.

Farklı madde ayırt edicilik dağılımlarına sahip madde havuzlarına göre madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı Revuelta ve Ponsoda (1998)'nın çalışması incelendiğinde, Revuelta ve Ponsoda tarafından geliştirilen Sınırlandırılmış Maksimum Bilgi (SMB) yöntemi ve İlerlemeli yöntem (İY) ile diğer altı yöntem (En Yüksek Bilgi [EYB], Bir parametre [1P], McBride ve Martin [MM], Randomesque [RA] SH, Tesadüfi Bilgi yöntemi [TBY]) ölçme duyarlılığı ve madde kullanım sıklığı değişkenleri bakımından karşılaştırılmıştır. Çalışmada öncelikle Ponsoda, Olea ve Revuelta (1994) tarafından bireylerin İngilizce kelime becerisini ölçmeye yönelik geliştirilen 221 maddelik gerçek madde havuzunda, sabit test uzunluğu (35 madde) ve değişken test uzunluğu ($SE(\theta) < .22$ ya da maks. 50 madde) koşullarında ilgili yöntemler karşılaştırılmıştır. İkinci aşamada ise EYB, İY, SMB, İlermeli sınırlandırılmış strateji (İSMB) ve SH yöntemleri farklı test uzunluklarında (20, 40 ve 60 madde) ve farklı madde ayırt edicilik dağılımlarına (madde havuzundaki maddelerin a parametrelerinin her biri Lognormal (-.25, .50) dağılımdan olmak üzere ortalamaları sırasıyla .93; 1.07; 1.30) sahip madde havuzlarına göre karşılaştırılmıştır. Madde havuzu büyüklükleri 500 ve 1000 maddeden oluşmaktadır. Her iki aşamada da başlatma kuralı olarak $\theta=0$ seçilmiş olup, yetenek düzeyleri normal dağılım $N(0,1)$ gösteren 2.000 birey simüle edilmiştir. Birinci aşamada elde edilen sonuçların farklı ayırt edicilik düzeylerine sahip madde havuzlarında da elde edilip edilemeyeceği incelenmiştir. Araştırmada maksimum kullanım sıklık oranını azaltmada Sınırlandırılmış yöntemin, kullanılmayan madde sayısını azaltmada ise İlerlemeli yönteminin etkili bir madde kullanım sıklığı kontrol yöntemi olduğu, her iki yöntemin de ölçme duyarlılığını düşürmediği sonucuna ulaşılmıştır. Bu sonuç doğrultusunda, ölçme duyarlılığını düşürmeden test güvenliğini sağlayabilecek birleştirilmiş bir

madde kullanım sıklığı kontrol yöntemi “İlerlemeli-Sınırlandırılmış yöntem” önerilmiştir.

Farklı ortalama güçlük düzeylerine sahip madde havuzlarına göre madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı Lee ve Dodd (2012)’un çalışması incelendiğinde, kısmî bilgi modeline dayalı simülatif BOBUT uygulamasında farklı yetenek dağılımlarına (sola çarpık ve normal dağılımda) ve farklı ortalama güçlük düzeylerine sahip madde havuzlarına (her biri 120’şer maddeden oluşan zor, orta ve kolay güçlükte madde havuzlarına) göre madde kullanım sıklığı kontrol yöntemlerinden Randomesque, Maksimum Bilgi ve İlerlemeli-Sınırlandırılmış yöntemi karşılaştırıldığında ölçme duyarlılığı ve test güvenliği indekslerinin nasıl değiştiğinin incelenmesi amaçlanmıştır. Araştırma bulguları, sola çarpık yetenek dağılımlarındaki bireylerin yetenek düzeylerine hitap edebilecek maddeler seçmek yerine, kolay madde havuzundaki maddeler seçildiği durumda ortaya çıkan uyumsuzluk sorunu ölçme duyarlılığı ve test güvenliğinin en çok zedelenmesine neden olurken, orta güçlükte madde havuzunun kullanıma derecesinin diğerlerine göre daha iyi olduğunu; Randomesque ve Maksimum Bilgi yöntemi yerine İlerlemeli-Sınırlandırılmış yöntem seçildiğinde test güvenliğinin arttığını göstermektedir.

Farklı madde havuzu büyüklüklerine göre madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı çalışmalar incelendiğinde, Chang ve Twu (1998)’nin çalışmasında çoktan seçmeli ACT (American College Testing) Matematik alt testinden (ACT-1997) çekilen madde havuzu büyüklükleri 360 ve 720 maddeden oluşan iki madde havuzuna göre, madde kullanım sıklığı kontrol yöntemlerinden tesadüfî seçme yöntemlerinden 5-4-3-2-1 stratejisi, SH yöntemi, Davey ve Parshall yöntemi, Stocking Lewis koşullu ve koşulsuz çok terimli yöntemleri karşılaştırılmıştır. Küçük madde havuzu için madde parametrelerinin ortalama ve standart sapma değerleri sırasıyla a (1.02, .33), b (.19, 1.04) ve c (.18, .08) şeklinde dağılım göstermekte iken; küçük madde havuzu parametrelerinin dağılımına benzer şekilde geniş madde havuzu için a (1.02, .33), b (.16, 1.06) ve c (.17, .08) şeklinde dağılım göstermektedir. Madde kullanım sıklığı kontrol yöntemi uygulanmadığında hem geniş hem de küçük madde havuzları için maksimum madde kullanım sıklık oranı 1.00 bulunmuştur. Dolayısıyla bu sonuç doğrultusunda her ne kadar geniş

madde havuzları kullanılsa da test güvenliğinin zedelenmeyeceği anlamına gelmeyeceği belirtilmiştir. Davey ve Parshall yöntemi, Stocking Lewis koşullu çok terimli yöntemi için geniş madde havuzlarında gözlemlenen kullanım sıklık oranı yarı yarıya (.08'den .04'e) azalsa da SH (.09'den .07'ye) ve Stocking Lewis koşullu çok terimli yöntemi (.11'den.10'a) ihmal edilebilecek düzeyde azalmıştır. Bu nedenle Davey ve Parshall, Stocking Lewis koşullu çok terimli yöntemi için geniş madde havuzu kullanılması diğer yöntemlere göre test güvenliğini daha olumlu yönde etkilemektedir.

Revuelta ve Ponsoda (1998)'nın çalışmasında geniş madde havuzu (1000 madde) kullanıldığında küçük madde havuzuna (500 madde) göre ölçme duyarlılığının arttığı ve test uzunluğunu arttırmaktan ziyade madde havuzu genişliğini artırmanın minimum kullanım sıklık oranının sıfır olmasına sebep olduğu sonucuna ulaşmıştır. Pastor, Dodd ve Chang (2002) hangi madde kullanım sıklık yöntemi kullanılması gerektiğine karar verilirken dikkat edilmesi gereken hususlardan birinin de madde havuzu genişliği olduğu, küçük madde havuzu (60 madde) kullanıldığında, test güvenliği bakımından önemsenmeyecek farklar bulunmakta iken; geniş madde havuzu (100 madde) kullanıldığında koşullu Sympson Hetter ve koşullu genişletilmiş tabakalama deseninin test güvenliği bakımından daha iyi sonuçlar verdiği saptanmıştır.

Farklı test sonlandırma kurallarına göre madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı çalışmalar incelendiğinde, French ve Thompson, (2003) minimum 7 madde ve maksimum 12 madde değişken test uzunluklarında sonlandırma kurallarının kullanıldığı BOBUT uygulamasında, madde kullanım sıklığı kontrol yöntemlerinden hedeflenen kullanım sıklığı kontrol yöntemi, SH yöntemi, b parametresi blokeli a tabakalama yöntemlerinin ölçme duyarlılığına ve test güvenliğine etkisini incelemiştir. Yetenek düzeyleri normal dağılım $N(0,1)$ gösteren 10.000 birey simüle edilmiştir. Madde havuzu için işlevsel (operational) matematik kursuna yerleştirmeye yönelik gerçekleştirilen BOBUT uygulamasındaki 269 maddenin 3PL'ye dayalı marjinal en yüksek olabilirlik yöntemine göre kestirilen madde parametreleri kullanılmıştır. Araştırmada madde kullanım sıklığı kontrol yöntemleri arasında ölçme duyarlılığı bakımından benzer sonuçlar elde edilirken; test güvenliği açısından en iyi sonuçlar hedeflenen kullanım sıklığı kontrol yönteminde

elde edilmiştir. Ayrıca ortalama test uzunluğu bakımından en kısa SH yöntemi (7.72); en uzun ise b parametresi blokeli a tabakalama yönteminde (9.45) elde edilmiştir.

Revuelta ve Ponsoda (1998)'nin çalışmasında, sabit test uzunluğu (35 madde) ve değişken test uzunluğu ($SE(\theta) < .22$ ya da max. 50 madde) test sonlandırma koşullarında madde kullanım sıklığı kontrol yöntemlerinden (SMB, İY, EYB, 1P, MM, RA, SH ve TBY) ölçme duyarlılığı ve madde kullanım sıklığı değişkenleri bakımından karşılaştırılmıştır. Sabit test uzunluğu (35 madde) sonlandırma kuralı koşulunda en yüksek ölçme duyarlılığı EYB, MM ve RA yöntemlerinde elde edilirken; TBY ve 1P yöntemleri kullanıldığında en düşük ölçme duyarlılığı sonucuna ulaşılmıştır. Değişken test uzunluğu ($SE(\theta) < .22$ ya da max. 50 madde) sonlandırma kuralı koşulunda bireylere en az madde (en kısa test) EYB, MM ve RA yöntemlerinde uygulanırken; en fazla madde (en uzun test) TBY ve 1P yöntemlerinde uygulanmıştır. TBY yönteminde $SE(\theta) < .22$ sonlandırma kuralı koşuluna ulaşamadığından bireylere 50 madde uygulanmıştır.

Testlet (madde takımı) temelli BOBUT uygulamalarında farklı yetenek kestirimlerine göre madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı çalışmalar incelendiğinde Boyd (2003) çalışmasında, üç parametrelili lojistik madde takımı tepki modeline ve kısmî bilgi modeline dayalı madde takımları içeren BOBUT uygulamalarında en ideal madde kullanım sıklığı kontrol yöntemini belirlemeyi amaçlamıştır. Araştırma kapsamında kullanım sıklığının kontrol edilmediği durum referans olarak alınmış ve kullanım sıklığı kontrol yöntemlerinden Randomesque yöntemi, değiştirilmiş .10 logit aralığı yöntemi, ilerlemeli sınırlandırılmış yöntem ve SH yöntemi, maksimum bilgi, ölçme duyarlılığı ve kullanım sıklığı oranları bakımından karşılaştırılmıştır. Araştırma sonucunda kullanım sıklığı kontrol yöntemlerine ilişkin ölçme duyarlılığı bakımından benzer sonuçlar elde edilirken, madde havuzunun verimliliği bakımından en düşük performansı maksimum madde kullanım sıklık oranları önceden belirlenmesine rağmen koşullu seçme yöntemlerinden SH yöntemi gösterirken; tesadüfi seçme yöntemlerinden Randomesque yöntemi ve değiştirilmiş .10 logit aralığı yönteminde ise her ne kadar madde takımı havuzunun %70'i kullanılsa da maksimum madde kullanım sıklık oranı düşük bulunmuştur. Koşullu ve tesadüfi seçme yöntemlerinin birleşimi olan ilerlemeli sınırlandırılmış yöntem seçildiğinde, bütün madde takımı havuzu kullanılmakla

birlikte, maksimum madde kullanım sıklık oranını kontrol etme bakımından en iyi performansı göstermiştir. İlerlemeli sınırlandırılmış yöntem, hem üç parametrelili lojistik madde takımı tepki modeli hem de kısmî bilgi modeline dayalı madde takımları içeren BOBUT uygulamalarında en ideal kullanım sıklığı kontrol yöntemi olduğu saptanmıştır.

Tıp Fakültesi'ne Giriş Testi (Medical College Admission Test [MCAT]) Sözel Muhakeme bölümündeki okuma parçalarının madde takımı puanlanırken kısmî bilgi modeline dayalı dört BOBUT uygulaması gerçekleştirilerek madde kullanım sıklığı sonuçları incelenmiştir (Davis ve Dodd, 2003). BOBUT uygulamalarının birincisinde maksimum bilgi madde seçme yöntemi, ikincisinde tesadüfi madde seçme yöntemi, üçüncüsünde .10 logit aralığı yöntemi, sonuncusunda ise Luecht ve Nungester bilgisayar ortamında bireye uyarlanmış sıralı (sequential) test (CAST) kullanılmıştır. Araştırma sonucunda kullanım sıklığı kontrolü ve ölçme duyarlılığı bakımından CAST uygulamasının en iyi performansı gösterdiği belirtilmiştir.

BOBUT'un temel bileşenlerinden, madde kullanım sıklığı kontrol yöntemleriyle ilgili yurtdışında yapılan çalışmalar incelendiğinde, genelde kullanım sıklığının kontrol edilmediği durum referans olarak alınmış ve madde havuzunun ölçme duyarlılığı, maksimum madde kullanım sıklığı oranı ve test örtüşme (overlap) oranı bakımından karşılaştırmalar yapılmıştır. Yapılan çalışmalar sonucunda genel olarak madde kullanım sıklığı kontrol yöntemlerinin farklı madde havuzu büyüklüklerinden etkilendiği, orta güçlükteki maddelerden oluşan madde havuzlarında kullanılmayan madde sayısının daha az ve ölçme duyarlılığının daha iyi olduğu, normal dağılıma sahip örneklemelerde madde havuzunun güçlük düzeyi arttıkça kullanılmayan madde sayısının arttığı ortaya çıkmıştır.

Türkiye'de madde kullanım sıklığı kontrol yöntemlerinin karşılaştırılmasına konu olan sadece Boztunç-Öztürk (2014)'ün çalışmasına rastlanılmıştır. İlgili çalışmada, madde seçme yöntemlerine ve farklı güçlük düzeyindeki madde havuzlarına göre, madde kullanım sıklığı kontrol yöntemlerinin karşılaştırılması amaçlanmıştır. Araştırmada, ortalama güçlüğü orta düzeyde olan madde havuzunun daha verimli/işlevsel kullanıldığı; yüksek güçlük düzeyine sahip madde havuzlarında ise, genellikle zor maddelerin kullanılmadığı sonucuna ulaşılmıştır. Ayrıca madde seçme yöntemlerinden a-tabakalama madde seçme yöntemi ve Aşamalı Düşürme

madde kullanım sıklığı kontrol yöntemlerinin birlikte kullanıldığı durumlarda test güvenliğinin arttığı görülmüştür.

Ölçme sonuçlarına göre alınan önemli kararların, bireysel, politik ve sosyal yönden kritik sonuçları olabilmektedir (Clauser ve Mazor, 1998). Dolayısıyla bireylerin yeteneklerini mümkün olduğunca objektif ve doğru bir şekilde ölçerek ölçme duyarlılığını artırmanın ve testlerin standart koşullarda gerçekleştirilmesi bakımından test güvenliğini sağlamanın ne kadar önemli olduğu tartışılmaz bir gerçektir (Cook ve Eignor, 1991).

Rudman (1987) BOBUT uygulamalarının 21. yüzyılın ölçme yöntemi olarak nitelendirilebileceğini savunarak, BOBUT stratejilerine yönelik yapılacak çalışmaların önemine dikkat çekmektedir. Psikometrik yönden, BOBUT uygulamalarının iki önemli avantajı bulunmaktadır. Bunlardan biri ölçme duyarlılığını arttırması, diğeri ise güvenli bir test ortamının sağlanmasıdır (Weiss, 2004).

BOBUT uygulamaları, yurtdışında Graduate Management Admission Test (GMAT), Graduate Record Examination (GRE), North American Pharmacist Licensure Examination (NAPLEX) gibi yüksek risk içeren (high stakes) sınavlarda kullanılmaktadır. BOBUT uygulamasının sağlayacağı avantajlar göz önüne alındığında, Türkiye’de özellikle ulusal düzeyde yapılan (YKS, YDS, KPSS gibi) yüksek risk içeren test sonuçlarına dayanarak bireyler hakkında kritik kararların alındığı sınavlarda da BOBUT uygulamasına geçilmesinin gerekli olduğu düşünülmektedir.

Dünyada BOBUT uygulamalarının gün geçtikçe yaygınlaşmasıyla birlikte, testlerin güvenilirliği ve geçerliğini olumsuz yönde etkileyebilecek bazı etkenler de tartışılmaya başlanmıştır. BOBUT uygulamalarında, madde havuzunun dengeli bir şekilde kullanılmayıp sınırlı sayıda bazı maddelerin kullanılması, sık kullanılan maddelerin ifşâ edilmesiyle psikometrik özelliklerini kaybetmesi, madde havuzunun işlevselliğini kaybederek kullanılamaz hale gelmesi gibi etkenlerden dolayı, BOBUT’un temel bileşenleri arasında, madde kullanım sıklığı kontrolü yöntemleri de yer almaya başlamıştır.

Madde kullanım sıklığı kontrolü yöntemleriyle, ölçme duyarlılığını düşürmeden test güvenliğini sağlayabilmek için madde havuzunun daha dengeli

kullanılması, testler arasındaki madde örtüşme oranının azaltılması ve madde havuzunun yıllarca kullanılabilir olması hedeflenmektedir. Uluslararası Bilgisayar Ortamında Bireye Uyarlanmış Test Derneği (International Association for Computerized Adaptive Testing [IACAT], 2017) madde kullanım sıklığı kontrol yöntemlerinin, BOBUT algoritmasında yapılandırılması gereken kısıtlamalara yönelik ele alınan tartışmalı ve çözüm önerilerine ilişkin güncel konulardan biri olduğuna dikkat çekmektedir.

BOBUT uygulamalarında bir sınava birden çok kez katılan bireyin aynı maddelerle karşılaşması durumunda, önceden pratik yapacak olması, Yen (1993)'e göre, madde yerel bağımlılığına neden olan faktörler arasında gösterilmektedir. MTK'ya dayalı ölçekleme yapılan BOBUT uygulamalarında, MTK'nın yerel bağımsızlık varsayımının ihlal edilmesi, güvenilirliği şişirmekte, yetenek ve madde parametrelerinin kestirimine ilişkin standart hataları etkilemekte ve buna bağlı olarak, yetenek ve madde parametrelerinin hatalı kestirilmesine sebep olmaktadır (Demars, 2006). Madde kullanım sıklığı kontrol yöntemlerine yönelik yapılacak çalışmaların bu tür sorunlara da çözüm getireceği düşünülmektedir.

Bu tartışmalar doğrultusunda yapılan araştırmalarda, geleneksel kâğıt kalem testi ile farklı stratejiler kullanılarak yapılan BOBUT uygulamalarıyla kestirilen yetenek düzeyleri arasında manidar bir ilişki bulunup bulunmadığı; farklı madde seçme yöntemlerine, yetenek kestirim yöntemlerine, sonlandırma kurallarına, farklı ortalama güçlüğüne sahip madde havuzlarına, madde havuzu büyüklüklerine göre, madde kullanım sıklığı kontrol yöntemlerinin karşılaştırılması üzerine odaklanılmıştır. Ancak madde kullanım sıklığı kontrol yöntemlerinin incelenmesinde, farklı ölçme koşullarının (test uzunlukları, örneklem büyüklükleri ve yetenek dağılımlarının) birlikte ele alınıp değerlendirildiği bir araştırmaya alanyazında rastlanamamıştır.

MTK modellerinin uygulanmasında ve yorumlanmasında test uzunluğu, örneklem büyüklüğü ve yetenek dağılımı önemli bir etkiye sahiptir. Dolayısıyla farklı ölçme koşullarında (örneğin küçük örneklem - kısa test ve sola çarpık yetenek dağılımlarında) hangi madde kullanım sıklığı kontrol yönteminin daha işlevsel olduğunun belirlenmesi, hem test güvenliğinin artmasına hem de bireylerin yeteneklerin daha az hata ile kestirilmesine katkı sağlamaktadır. Bu bağlamda, farklı örneklem büyüklüklerine, test uzunluklarına ve yetenek dağılımlarına göre madde

kullanım sıklığı kontrol yöntemleri karşılaştırıldığında ölçme duyarlılığı ve test güvenliği indekslerinin nasıl değiştiğinin incelenmesi bu araştırmanın problemini oluşturmaktadır.

1.2. Amaç

Bu çalışmada, simülatif bilgisayar ortamında bireye uyarlanmış test uygulamalarında iki kategorili (1-0) puanlanan çoktan seçmeli maddelerde madde kullanım sıklığı kontrol yöntemlerinin farklı ölçme koşullarında, ölçme duyarlılığına ve test güvenliğine etkisinin incelenmesi amaçlanmaktadır. Bu genel amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

- 1) *BOBUT uygulamalarında, madde kullanım sıklığı kontrolü yapılmadığında*
 - Test uzunluklarına kısa test (25 madde), uzun test (50 madde),
 - Örneklem büyüklüklerine (küçük örneklem ($n=250$), geniş örneklem ($N=1000$))
 - Yetenek dağılımlarına (sağa çarpık, normal, tek biçimli dağılım, sola çarpık) göre ölçme duyarlılığı ve test güvenliği nasıl değişmektedir?
- 2) *BOBUT uygulamalarında, madde kullanım sıklığı kontrolü yöntemlerinden*
 - 2.1.) *Sympson Hetter Stratejisi kullanıldığında;*
 - Test uzunluklarına (kısa test (25 madde), uzun test (50 madde)),
 - Örneklem büyüklüklerine (küçük örneklem ($n=250$), geniş örneklem ($N=1000$))
 - Yetenek dağılımlarına (sağa çarpık, normal, tek biçimli dağılım, sola çarpık) göre ölçme duyarlılığı ve test güvenliği nasıl değişmektedir?
 - 2.2.) *Aşamalı Düşürme Stratejisi kullanıldığında;*
 - Test uzunluklarına (kısa test (25 madde), uzun test (50 madde)),
 - Örneklem büyüklüklerine (küçük örneklem ($n=250$), geniş örneklem ($N=1000$))
 - Yetenek dağılımlarına (sağa çarpık, normal, tek biçimli dağılım, sola çarpık) göre ölçme duyarlılığı ve test güvenliği nasıl değişmektedir?

1.3. Araştırmanın Önemi

Küreselleşen dünyada, bilgisayar ve teknolojiadaki gelişmeler hayatımızın her alanına yön verdiği gibi, eğitim alanındaki yeniliklere ve değişimlere de yön vermeye başlamıştır. Bilgisayar ve yazılımlardaki gelişmelerin yanı sıra psikometri alanındaki gelişmeler ile birlikte testlerin bilgisayar ortamında bireye uyarlanmış test olarak uygulanması giderek yaygınlaşmaya başlamıştır.

21. yüzyılın ölçme yöntemi olarak nitelendirilen (Rudman, 1987) BOBUT uygulamalarındaki gelişmelerle birlikte, GRE, GMAT, TOEFL, NAPLEX gibi yüksek risk içeren sınavlar, BOBUT olarak uygulanmaya başlanmıştır. Bu süreçte birçok kurum, kuruluş ve test şirketi de (Educational Testing Services, Cambridge Üniversitesi ve British Council vb.) testlerin BOBUT olarak uygulanmasına yönelik imkânlar sunmaya başlamıştır.

BOBUT uygulamalarında, bireyler yetenek düzeylerine en uygun maddelerle karşılaştıkları için, özellikle geniş ölçekli testlerde, bir bireyin istediği sayıda ve istediği sıklıkta testi almasına izin verilebilmektedir. Bu şekilde BOBUT uygulamalarının sürekliliğin olması, her ne kadar sınav takvimini planlama, test uygulamaları için test merkezlerinde, okullarda vs. yeterli mekân ve bilgisayar sağlanması bakımından avantajlar sağlasa da dezavantajları da bulunmaktadır. Bu tür uygulamalarla, bir sınava birden çok kez katılan bireyin önceden karşılaştığı maddeleri hatırlaması, internet aracılığıyla sosyal medyada ya da arkadaşlarıyla çok sık uygulanan maddeleri paylaşması riskini ortaya çıkarmaktadır. Ayrıca BOBUT uygulamalarında, bireylerin ilgili yetenek düzeyleri ile yüksek bilgi değerine sahip maddelerin bilgisayar ortamında eşleştirilmesi, madde havuzunun dengeli bir şekilde kullanılmayıp sınırlı sayıda bazı maddelerin kullanılması, sık kullanılan maddelerin ifşâ edilmesiyle psikometrik özelliklerinin düşmesi ve madde havuzunun işlevselliğini kaybetmesi gibi sorunları da beraberinde getirmektedir. Bu tür etkenler, BOBUT uygulamalarının geçerliğini ve güvenilirliğini etkileyecek faktörlerden, test güvenliği ve ölçme duyarlılığına ilişkin tartışmalara yol açmıştır.

Test güvenliği, Uluslararası Bilgisayar Ortamında Bireye Uyarlanmış Test Derneği (2017) tarafından tartışmalı ve güncel konulardan olduğu vurgulanmakla birlikte, son zamanlarda sınav soruların çalınması, sonuçların yanlış hesaplanması,

sınav sistemlerinin değiştirilmesi gibi konularla, test güvenliği ulusal medyada da sık sık gündeme gelmektedir. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Derneği [EPODDER] (2017) tarafından hazırlanan Temel Eğitimden Ortaöğretime Geçiş Konusundaki Basın Bildirisi'nde de vurgulandığı gibi yaşanan bu tür olumsuzluklar, sınav hizmeti veren kurumların kamuoyundaki saygınlığını ve güvenilirliğini zedelemektedir. Kamuoyundaki kurumsal saygınlık ve güvenirliliğin tekrar kazanılabilmesi için atılacak adımların neler olması gerektiği hususunda, test güvenliğini artırmaya yönelik yapılacak çalışmaların, sınav hizmeti veren kurum ve kuruluşlara ışık tutması beklenmektedir.

BOBUT uygulamalarıyla elde edilen ölçme sonuçları öğrenciler, öğretmenler, aileler ve politikacılar üzerinde önemli etkilere sahip olabilmektedir. Bu nedenle ölçme sonucunda elde edilen bilginin doğruluğu alınan kararları da etkilemektedir. Dolayısıyla BOBUT uygulamaları sonucunda geçerli ve güvenilir yetenek kestirimlerinin yapılabilmesi için, farklı koşullarda ölçme duyarlılığını ve test güvenliğini artırmaya yönelik madde seçme algoritmasında yapılması gereken kısıtlamaların neler olması gerektiği hususunda yapılacak çalışmaların araştırmaya değer bir konu olduğu düşünülmektedir.

Alanyazında yapılan incelemeler sonucunda, farklı test uzunlukları, örneklem büyüklükleri ve yetenek dağılımları birlikte ele alınarak, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğini sağlayan en işlevsel madde kullanım sıklığı kontrol yönteminin belirlenmesine yönelik herhangi bir çalışmayla karşılaşılmamıştır. Farklı ölçme koşullarına göre madde kullanım sıklığı yöntemlerinin incelenmesiyle elde edilecek sonuçların, hem daha geçerli ve güvenilir BOBUT uygulamalarının yapılmasına hem de alanyazındaki söz konusu eksikliğin giderilmesine önemli katkıları olacağı düşünülmektedir. Ayrıca çalışmanın incelenilen değişkenler ve koşullar bakımından Türkiye'de yapılan ilk simülatif araştırma olarak, BOBUT uygulamaları sürecinde yol gösterici olması da ayrı bir önem taşımaktadır.

1.4. Araştırmanın Sınırlılıkları

Bu çalışma, BOBUT uygulaması için madde parametreleri 3 parametrelili lojistik (3PL) modele göre kestirilen 1000 adet 1-0 puanlanan iki kategorili maddelerden oluşan simülatif madde havuzu, başlatma kuralı $-.50 < b < .50$, yetenek kestirim yöntemi BSD, sonlandırma kuralı olarak sabit test uzunluğu stratejisinin kullanılmasının yanı sıra araştırmanın amacı doğrultusunda ele alınan madde kullanım sıklığı kontrol yöntemleriyle sınırlıdır.

1.5 Kısaltmalar

KTK: Klasik Test Kuramı

MTK: Madde Tepki Kuramı

MFB: Maksimum Fisher Bilgisi

AEYBO: Aşamalı En Yüksek Bilgi Oranı

BSD: Beklenen Sonsal Dağılım Yöntemi

EYOY: En Yüksek Olabilirlik Yöntemi

YOK: Madde kullanım sıklığı kontrolü yapılmaması

SH: Sympson Hetter stratejisi

AD: Aşamalı Düşürme yöntemi

SMB : Sınırlandırılmış Maksimum Bilgi yöntemi

İY: İlerlemeli yöntem

EYB: En Yüksek Bilgi yöntemi

MM: McBride ve Martin yöntemi

RA: Randomesque yöntemi

TBY: Tesadüfi Bilgi yöntemi

KBM: Kısmî Bilgi Modeli

BÖLÜM II

2. YÖNTEM

Bu bölümde, araştırmanın modeli, veriler, verilerin üretilmesi, BOBUT koşulları ve verilerin analizi ile ilgili bilgiler yer almaktadır.

2.1. Araştırmanın Modeli

Ampirik veri bulmanın zor olduğu karmaşık çalışmalarda, yöntemler arası bağıntıların anlaşılmasına yönelik yapılan simülasyon çalışmalarının kuramın gelişmesinde önemli katkılar sağladığı belirtilmektedir (Davis, Eisenhardt ve Bingham, 2007). Bu çalışma ile monte-carlo simülatif BOBUT uygulamalarında, 1-0 puanlanan iki kategorili maddelerde madde kullanım sıklığı kontrol yöntemleri, farklı ölçme koşullarına göre karşılaştırıldığında ölçme duyarlılığı ve test güvenliği indekslerinin nasıl değiştiğinin incelenmesi amaçlandığı için kuramsal araştırma olarak değerlendirilebilir. Kuramsal araştırmalar, var olan kuram veya yöntemleri test ederek mevcut kuramsal bilgi alanına yenilerini katmak amacıyla yapılan araştırmalar olarak tanımlanmaktadır (Bailey, 1994).

2.2. Veriler

Naylor, Balintfy, Burdick ve Chu (1968) simülasyon çalışmalarında problem tanımlandıktan sonra, bağımsız ve bağımlı değişkenlerin tanımlanması gerektiğini vurgulamaktadır (Akt. Harwell, Stone, Hsu ve Kirişçi, 1996). Bu bilgi doğrultusunda ilgili araştırmadaki bağımsız değişkenler; örneklem büyüklükleri, test uzunlukları ve yetenek dağılımları ve madde kullanım sıklığı kontrol yöntemleridir. Bağımlı değişkenler ise farklı koşullar doğrultusunda gerçekleştirilen BOBUT uygulamaları sonucunda elde edilen ölçme duyarlılığı ve test güvenliği indeksleridir.

2.3. Verilerin Üretilmesi

BOBUT uygulamasında MTK'ya göre ölçeklenmiş maddelerden oluşan geniş madde havuzunun yanı sıra BOBUT uygulamalarında en az hata ile kestirimler elde edilebilmesi için çok fazla sayıda katılımcının yanıt örüntülerine ihtiyaç duyulmaktadır. Bunun yanı sıra, araştırmanın amacı doğrultusunda farklı ölçme koşullarını sağlayabilecek veri setlerine ulaşmak pratik anlamda neredeyse imkânsız olduğundan simülatif veriler kullanılmıştır. Revuelta ve Ponsoda (1998) simülasyon çalışmalarının farklı ölçme koşullarında (test uzunluğu, madde havuzu genişliği vb.) karşılaştırmalar yapılabilmesini mümkün hale getirdiğini vurgulamaktadır. BOBUT uygulamalarına yönelik simülasyon çalışmalarında, genellikle post-hoc (gerçek veri) simülasyonları ya da monte-carlo simülasyonları uygulanmaktadır. Post-hoc simülasyon çalışmalarında gerçek bireylerin madde yanıt örüntüsüne ihtiyaç duyulmakta iken; monte-carlo simülasyon çalışmalarında ise, hiçbir veri seti mevcut olmadığında, araştırmacı tarafından belirtilen koşullar doğrultusunda madde yanıt örüntüsü oluşturulmaktadır. Monte-carlo simülasyonu, varsayımsal BOBUT uygulamalarını kontrol eden değişken parametrelerle bağlantılı olarak verilerin farklı yönlerinin hızlı ve verimli bir şekilde değiştirilmesine olanak tanımaktadır (Weiss ve Guyer, 2012). Bir başka ifadeyle monte-carlo simülasyon çalışmaları, farklı stratejilerde veri setlerini çeşitlendirerek BOBUT uygulamalarına ilişkin etkin ve hızlı bir şekilde karşılaştırma imkânı sunmaktadır.

Simülatif verilerin üretilmesi için SimulCAT yazılımından faydalanılmıştır. SimulCAT, farklı stratejilere (başlatma ve sonlandırma kuralları, madde seçme ve yetenek kestirim yöntemleri, içerik dengelenmesi ve madde kullanım sıklığı kontrolü yöntemlerine) göre karşılaştırma imkânları sunan kullanıcı dostu BOBUT simülasyon yazılım aracıdır. SimulCAT uygulamalarına ilişkin (.sca) uzantılı çıktı dosyalarında replikasyon numarası, birey numarası, gerçek yetenek (true theta) değeri, uygulanan madde numarası, nihâi yetenek kestirimi, nihâi yetenek kestiriminin standart hatası gibi bilgiler mevcuttur (Han, 2011).

MTK modellerinin uygulanmasında ve yorumlanmasında test uzunluğu ya da örneklem büyüklüğü önemli bir etkiye sahiptir. Şahin ve Anıl'a (2017) göre, tek boyutlu iki kategorili (1-0) puanlanan MTK modellerinde, test uzunluğu ve örneklem büyüklüğü birlikte ele alınıp değerlendirildiğinde, madde parametreleri kestirimi

açısından daha da önemli bir etkiye sahip olduğu vurgulanmaktadır. Alanyazında test uzunluğunun ve örneklem büyüklüğünün dağılımına ilişkin farklılıklar bulunmakla birlikte genellikle kısa test için 25 madde (Demars, 2006; Guyer ve Thompson, 2011; Harwell, Stone, Hsu ve Kirişçi, 1996; Weiss ve Von Minden, 2012; Yoes, 1995) uzun test için 50 madde (Demars, 2006; Glas, 2002; Weiss ve Von Minden, 2012); küçük örneklem 250 birey (Çetin, 2009; Goldman ve Raju, 1986; Harwell ve Janosky, 1991; Speron, 2009; Vaughn ve Wang, 2010; Yoes, 1995), geniş örneklem ise 1000 birey (Çetin, 2009; Glas, 2002; Goldman ve Raju, 1986; Guyer ve Thompson, 2012; Hulin, Lissak ve Drasgow, 1982; Köse, 2010; Lord, 1968; Patsula ve Gessaroli, 1995; Speron, 2009; Tang, Way ve Carey, 1993; Thissen ve Wainer, 1982; Vaughn ve Wang, 2010; Yen, 1987; Yoes, 1995; Weiss ve Von Minden, 2012) olduğu vurgulanmaktadır. İlgili alanyazın bilgisi doğrultusunda, bu çalışmada, kısa test 25 madde, uzun test 50 madde; küçük örneklem 250 birey, geniş örneklem 1000 birey olarak belirlenmiştir.

Testi alan bireylerin yetenek parametreleri her bir örneklem büyüklüğü koşulu için (θ) değeri -3 ve +3 aralığında i) normal dağılım $\theta \sim N(0, 1)$; Weiss ve Guyer (2012)'e göre ii) sağa çarpık dağılım $\beta(1, 4)$; iii) sola çarpık dağılım $\beta(4, 1)$; iv) tek biçimli (uniform) dağılım $\beta(1, 1)$ gösterecek şekilde üretilerek BOBUT örneklemeleri oluşturulmuştur Bunun yanı sıra, tek biçimli dağılımda ortalama ve standart sapma değerleri için Agresti ve Coull (1998)'a göre, $U(-3,3)$ için $\bar{x} \cong .00$, $SS = \frac{6}{\sqrt{12}} \cong 1.732$ olacak şekilde BOBUT örneklemeleri elde edilmiştir.

Madde kullanım sıklığı kontrol yöntemlerinden, SH stratejisinde BOBUT uygulamasından önce, tekrarlı simülasyonlarla kullanım sıklığı kontrol parametrelerinin belirlenmesi gerekmektedir (Veldkamp, Vershoor ve Eggen, 2007). Bunun için her bir BOBUT örneğinde yer alan bireylerin yetenek dağılımları ile benzer dağılıma sahip olacak şekilde (örnek olarak yetenek dağılımı normal dağılım gösteren koşulda $\bar{X}_{BOBUT-\theta} = .007$, $SS_{BOBUT-\theta} = .998$, $Ky_{BOBUT-\theta} = -.15$, $Bs_{BOBUT-\theta} = -.79$; $\bar{X}_{SH-\theta} = -.006$, $SS_{SH-\theta} = .998$, $Ky_{SH-\theta} = .01$, $Bs_{SH-\theta} = -.02$) Sympson Hetter örnekleme için örneklem büyüklüğü 6000 ve replikasyon sayısı 5 alınarak madde kullanım sıklığı kontrol parametreleri hesaplanmıştır (Gu ve Reckase, 2007). Eğer her bir maddenin beş replikasyon sonrasında ortalama seçilme oranı ($P_i(S)$), maksimum hedeflenen kullanım sıklık oranından ($r = .20$) büyükse, madde kullanım sıklığı kontrol

parametresi $k_i=r/P_i(S)$ olarak hesaplanmakta; $(P_i(S))<r$ ise $k_i=1$ alınarak ilgili maddelerin uygulanma olasılığı artırılmaktadır (Eggen, 2001; Veldkamp ve van der Linden, 2008). BOBUT örnekleme ve Sympson Hetter örnekleme ilişkili histogram ve kümülatif grafikleri EK-1’de yer almaktadır.

BOBUT uygulamalarında maddeler her bir yetenek düzeyine hitap edecek şekilde yeterli sayıda ve eşit olarak dağılım gösteren ve ayırt ediciliği yüksek maddelerden oluşan havuzla çalışıldığı takdirde daha iyi sonuçlar vereceğinden, Weiss ve Von Minden (2012)’e göre a (.25 ile 1.75), b (-3 ile +3), c ise (.20 ile .30) ranj aralığında, ayırt edicilik ve şans parametreleri tek biçimli dağılım, güçlük parametresi ise normal dağılım gösteren 1000 maddeden oluşan simülatif madde havuzu elde edilmiştir. Ayrıca tek boyutlu iki kategorili (1-0) puanlanan çoktan seçmeli maddelerde, şans başarısı parametresini de dikkate alan bir model olduğu için 3PL modele göre kestirimler yapılmıştır (Crocker ve Algina, 1986; Hambleton ve diğerleri, 1991). Madde havuzuna ilişkin yanıt fonksiyonu ve bilgi fonksiyonu grafikleri EK-2’de yer almaktadır.

2.4. BOBUT Koşulları

Araştırma kapsamında simülatif BOBUT uygulaması için, madde seçme yöntemlerinden en yaygın kullanılan Maksimum Fisher Bilgisi yöntemi ve başlatma kuralı stratejilerinden $-.50 < b < .50$ aralığı stratejisi kullanılmıştır (Weiss, 1988). Yetenek kestirim yöntemi olarak maddelere tümüyle doğru ya da tümüyle yanlış yanıt veren bireyler bulunabileceği için en yüksek olabilirlik kestirim yöntemi yerine önsel dağılım kullanarak kestirim yapan beklenen sonsal dağılım (BSD) yöntemi [Expected A Posteriori] tercih edilmiştir (Embretson ve Reise, 2000).

Madde kullanım sıklığı kontrol yöntemlerinden, BOBUT uygulamasından önce, tekrarlı simülasyonlarla kullanım sıklığı kontrol parametrelerinin belirlendiği BOBUT uygulamalarında en yaygın kullanılan Sympson Hetter stratejisi (Veldkamp, Vershoor ve Eggen, 2007) ve BOBUT uygulamasından önce, tekrarlı simülasyonlar gerektirmeyen, BOBUT uygulanma sürecinde kullanım sıklığı kontrol parametrelerinin eş zamanlı olarak belirlendiği Aşamalı Düşürme Stratejisi (Han, 2009) seçilmiştir. Sonlandırma kuralı olarak sabit test uzunluğu (kısa test:25 madde;

uzun test:50 madde) tanımlanmıştır. Araştırma kapsamında BOBUT uygulaması için belirlenen BOBUT bileşenleri ve farklı koşullar Çizelge 1’de özetlenmiştir.

Çizelge 1

Araştırma Kapsamında Belirlenen BOBUT Bileşenleri ve Koşullar

<i>BOBUT Bileşenleri</i>	<i>Koşullar</i>	<i>Koşul Sayısı</i>
Madde Havuzu Genişliği	1000 madde	1
Örneklem Büyüklüğü	Küçük Örneklem:250 birey	2
	Geniş Örneklem:1000 birey	
Yetenek Dağılımı	Sağa çarpık dağılım	4
	Normal dağılım	
	Tek biçimli dağılım	
	Sola çarpık dağılım	
Yetenek Kestirim Yöntemi	BSD	1
Başlatma Kuralı	$-.50 < b < .50$	1
Madde Seçme Yöntemi	MFB	1
Madde kullanım sıklığı kontrolü	Madde kullanım sıklığı kontrolünün yapılmaması (YOK)	3
	Sympson Hetter (SH) Stratejisi	
Sonlandırma Kuralı	Aşamalı Düşürme (AD) Stratejisi	2
	Kısa Test:25 madde	
	Uzun Test:50 madde	

Çizelge 1’de yer alan değişkenler ve koşullar çaprazlandığında $2*4*3*2=48$ simülasyon koşulu belirlenmiştir. Ayrıca örneklem yanlılığı oluşmaması için (Harwell 1996; akt. Evans, 2010) her bir koşulda 100 tekrar kullanılmıştır (Han, 2009). Böylece 4800 veri dosyası oluşturulmuştur.

2.5. Verilerin Analizi

Araştırma kapsamında her bir koşulda ölçme duyarlılığı ve test güvenliği indeksleri hesaplanıp birbirleriyle karşılaştırılmıştır.

2.5.1. Farklı Ölçme Koşullarında Ölçme Duyarlılığına İlişkin Veri Çözümleme Yöntemleri

Ölçme duyarlılığı arttığında, hata değerleri azalacağı için, araştırmada ölçme duyarlılığının belirlenmesi amacıyla, her bir koşulda hata göstergeleri olan uyum (fidelity), yanlılık (bias) ve Hataların Ortalama Karekökü (Root Mean Squared Error [RMSE]) katsayıları hesaplanmıştır (French ve Thompson, 2003).

Uyum katsayısı, BOBUT uygulaması ile kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasındaki korelasyon katsayısıdır. Uyum katsayısını hesaplamak için Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılmıştır. BOBUT uygulamasında kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki ortalama anlamlı farklılığa ilişkin yanlılık ve mutlak farklılık için RMSE değerleri aşağıda sırasıyla Eşitlik 1, Eşitlik 2 yardımıyla hesaplanmıştır (Gu ve Reckase, 2007; Leroux, Lopez, Hembry ve Dodd, 2013; Wang ve Vispoel, 1998; Zheng ve Chang, 2014):

$$\text{Yanlılık} = \frac{\sum_{k=1}^n (\widehat{\theta}_k - \theta_k)}{n} \dots\dots(\text{Eşitlik 1})$$

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^n (\widehat{\theta}_k - \theta_k)^2}{n}} \dots\dots(\text{Eşitlik 2})$$

Formüllerde n=toplam birey sayısını; $\widehat{\theta}_k$: k. bireyin kestirilen yetenek düzeyini, θ_k : k. bireyin gerçek yetenek düzeyini göstermektedir. Ayrıca RMSE ile yanlılık değeri arasında $(\text{RMSE})^2 = (\text{Yanlılık})^2 + (\text{SE})^2$ bir ilişki bulunmaktadır. Dolayısıyla RMSE değeri, yanlılık ve tahminin standart hatasını içermektedir. Ölçme duyarlılığının belirlenmesinde, bu indeksler birlikte değerlendirilerek yorumlanmıştır.

Farklı ölçme koşullarında ölçme duyarlılığına ilişkin uyum değerleri arasında manidar bir farklılık bulunup bulunmadığı, iki korelasyon katsayısı arasındaki farkın

karşılaştırılmasına imkân veren Fisher'in z testi ile incelenmiştir (Howell, 2010; Şencan, 2005). Fisher (1921) iki farklı korelasyon katsayısı arasında manidar bir fark bulunup bulunmadığını incelemek için, korelasyon (r) katsayılarının sırasıyla $r' = .50$ ve $r' = .50$ dönüştürülerek $z = \frac{r'_1 - r'_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$ değerinin hesaplanması gerektiğini vurgulamaktadır. Hesaplanan z değeri .05 manidarlık düzeyinde ∓ 1.96 kritik değerleri sınırları dışında kaldığında $H_0 = \text{İki farklı uyum katsayısı arasında manidar bir fark bulunmamaktadır. } (r_1 = r_2)$ hipotezi red edilmektedir (Akt. Howell, 2010). Dolayısıyla iki dönüştürülmüş korelasyon katsayısı arasındaki uzaklık $|r'_1 - r'_2| > z_{.05} \cdot \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$ ise H_0 hipotezi red edilmektedir. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin korelasyon katsayılarının r' dönüştürülmüş değerlerine EK-8'de yer verilmiştir.

Farklı ölçme koşullarında *küçük örneklem*lere ilişkin karşılaştırmalarda serbestlik derecesi $\sqrt{\frac{1}{250 - 3} + \frac{1}{250 - 3}} = .09$ ve kritik değer $z_{.05} = 1.96$ olmak üzere, $|r'_1 - r'_2| > 1.96 \cdot .09 = .176$ olduğunda H_0 hipotezi red edilirken, *geniş örneklem*lere ilişkin karşılaştırmalarda ise serbestlik derecesi $\sqrt{\frac{1}{1000 - 3} + \frac{1}{1000 - 3}} = .045$ olmak üzere $|r'_1 - r'_2| > 1.96 \cdot .045 = .088$ olduğunda H_0 hipotezi red edilmiştir.

Farklı değişkenlere (farklı ölçme koşullarında yanlılık / RMSE değerlerine) ait değişimlerin gruplandırılmasına/kümeler arasındaki uzaklığın hesaplanmasına imkân sunduğundan, örüntüleri bilinen sonuçların karşılaştırmasında ve aşırı gözlemleri test etmede Mahalanobis uzaklıkları kullanılabilir (Hair, Anderson, Tatham ve Black, 1998; Pallant, 2010). Dolayısıyla farklı ölçme koşullarında ölçme duyarlılığına ilişkin yanlılık / RMSE değerleri arasında manidar bir fark bulunup bulunmadığını incelemek için Mahalanobis uzaklıkları hesaplanmıştır.

Mahalanobis uzaklığı (D^2), n değişken ve kovaryansları tarafından tanımlanan, n boyutlu bir uzaydaki her bir grup için, herhangi bir gözlemin grup merkezine (ortalama vektörüne) olan uzaklığı dikkate alınarak hesaplanmaktadır. Dolayısıyla D^2 değeri arttıkça ilgili gözlemin veri merkezine olan uzaklığı artmaktadır (McLachlan, 1999; Khatree ve Naik, 2000).

Mahalanobis uzaklığı, ilgili özelliklerin (yanlılık / RMSE değerlerinin) kovaryans matrisini dikkate almaktadır. Böylelikle tesadüfi bir değişkenin iki bileşeni arasındaki çapraz kovaryansların etkisini doğru bir şekilde değerlendirme imkânı sunmaktadır. Örneğin, madde kullanım sıklığı yöntemlerinden Sympson Hetter Stratejisi ile Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan yanlılık / RMSE değerleri arasındaki çapraz kovaryansların etkisi değerlendirilebilmektedir.

Mahalanobis uzaklık değeri, değişken sayısı p olmak üzere, p serbestlik dereceli χ^2 dağılımında .001 manidarlık düzeyinde Mahalanobis uzaklığı kritik değerinden küçük olduğunda manidar bir farklılık bulunmadığı belirtilmiştir. Dolayısıyla çok değişkenli herhangi bir gözlemin, merkezi (centroid) değerden sapan bir değer olmadığı (Kalaycı, 2010) bir başka ifadeyle ilgili gözlemin aşırı bir gözlem ya da başka bir evrene ait olmadığı, verilerin çok değişkenli homojenlik gösterdiği ifade edilmiştir (Alpar, 2011; Khatree ve Naik, 2000).

2.5.2. Farklı Ölçme Koşullarında Test Güvenliğine İlişkin Veri Çözümleme Yöntemleri

Her bir koşulda, kullanım sıklığı oranı, madde havuzundan yararlanmama (pool utilization) ve test örtüşme (test overlap) indeksi birlikte değerlendirilerek test güvenliği hakkında yorum yapılabilmektedir (Davis ve Dodd, 2005). Madde havuzundan yararlanmama (pool utilization) ile BOBUT sürecinde uygulanmayan maddelerin yüzdesi kastedilmektedir (Leroux ve diğerleri 2013). Madde kullanım sıklığı oranlarının (r_i : maddelerin kullanılma sayısının örneklem büyüklüğüne oranlarının) dağılımı, madde kullanım sıklığı oranlarının standart sapması ve maksimum madde kullanım sıklığı oranı hesaplanarak her bir koşulda *kullanım sıklığı oranları* belirlenmektedir. Ayrıca uygulanmayan madde sayısı ve oranı hesaplanarak elde edilen *madde havuzundan yararlanmama indeksi*, madde havuzunun kullanılma derecesinin belirlenmesine imkân sunmaktadır.

Test örtüşme (test overlap) indeksinde her bir koşulda, tesadüfi olarak seçilen iki birey için örtüşen maddelerin sayısı/karşılaştıkları aynı maddelerin sayısı dikkate alınmaktadır. Test örtüşme indeksi aşağıdaki eşitlikler yardımıyla hesaplanmıştır (Chen, Ankenmann ve Spray, 1999):

$$\sum_{i=1}^n \binom{m_i}{2}$$

BOBUT sürecinde m_i : uygulanan i. maddenin kullanılma sayısı n : madde havuzu genişliği olmak üzere, eğer $m_i < 2$ ise $\binom{m_i}{2} = 0$ olmaktadır. Başka bir ifadeyle, BOBUT uygulamalarında, herhangi bir madde, yalnızca bir kez uygulandıysa, test örtüşme indeksini etkilememektedir. Genel olarak testler arası örtüşme indeksinin ortalaması aşağıdaki formülle hesaplanmıştır.

$$\bar{T} = \frac{\sum_{i=1}^n \binom{m_i}{2}}{k \binom{N}{2}}$$

N :BOBUT uygulamasına katılan bireylerin sayısı ve k :BOBUT uygulamasındaki sabit test uzunluğunu göstermektedir. Formül kavramsal olarak tanımlanacak olursa, BOBUT uygulamasına katılan bireyler arasından tesadüfi olarak seçilen ikili bireyler için örtüşen maddelerin sayısının toplamı, tesadüfi olarak seçilen her bir ikili bireyin katıldıkları BOBUT uygulamasındaki test uzunluklarına (bireylerin ikili kombinasyonları ile test uzunluklarının çarpımına) oranı, testler arası örtüşme indeksinin ortalamasını vermektedir. Test örtüşme indeksi ortalamasının büyük olması, test güvenliğinin zedelendiği anlamına gelmektedir (Huang, Chen ve Wang, 2012).

Araştırmanın sınırlılıkları doğrultusunda en ideal madde kullanım sıklık yöntemine karar verilmesi için F istatistiği sonucu kullanılmıştır (Chang, Qian ve Ying, 2001; Grubbs, 1973). Chang ve Ying (1999, s.215)’a göre, “Madde kullanım sıklık oranları karşılaştırılırken, χ^2 istatistikleri hesaplanarak, $\chi^2_{yöntem1}$ ve $\chi^2_{yöntem2}$ karşılaştırılması” gerektiğini belirtmektedir.

Karşılaştırma ölçüsü olarak ($F_{\chi^2_{yöntem1}, \chi^2_{yöntem2}} = \frac{\chi^2_{yöntem1}}{\chi^2_{yöntem2}}$) istatistiği tanımlanmıştır. Eğer $F_{\chi^2_{yöntem1}, \chi^2_{yöntem2}} < 1$ ise, madde kullanım sıklığı oranlarının genel olarak dengelenmesi bakımından birinci yöntemin, ikinci yönteme göre daha iyi bir yöntem olduğu şeklinde yorumlanmaktadır.” χ^2 istatistiği aşağıdaki Eşitlik 3 yardımıyla hesaplanmıştır (Tay, 2015):

$$\chi^2 = \frac{\sum_{i=1}^n (r_i - uni(r_i))^2}{uni(r_i)} \dots \dots (Eşitlik 3)$$

Madde kullanım sıklık oranlarının (r_i), bütün maddeler için istenilen tek biçimli madde kullanım sıklık oranına (test uzunluğunun madde havuzu büyüklüğüne oranına ($uni(r_i) = k/n$)) farkının kareleri toplamı, istenilen tek biçimli madde kullanım sıklık oranına bölünerek hesaplanmıştır.

Farklı ölçme koşullarında test güvenliğine ilişkin madde kullanım sıklığı oranlarının standart sapma değerleri arasında manidar bir fark bulunup bulunmadığı “İki varyansa ilişkin hipotez testi” ile incelenmiştir. Dolayısıyla öncelikle ilgili koşullardaki madde kullanım sıklığı oranlarının standart sapma değerlerinin kareleri alınarak varyansları hesaplanmıştır. İki örneklem varyansının birbirine oranı F dağılımını ($F_h \sim \frac{S_1^2}{S_2^2}$) vermektedir. Hesaplanan F_h değeri, çift yönlü hipotezi test ettiğimizden, serbestlik dereceleri n_1-1 ve n_2-1 olmak üzere, normal dağılım eğrisinin altındaki sol uçtaki $.025^{F_{n_1-1, n_2-1}} = \frac{1}{F_{n_2-1, n_1-1}, .025}$ kritik değerinden küçük veya sağ uçtaki F_{n_1-1, n_2-1} kritik değerinden büyük ise $H_0 = \sigma_1 = \sigma_2$ hipotezi red edilmiştir (Ünver, Gamgam ve Altunkaynak, 2016). Dolayısıyla H_0 hipotezi red edildiğinde, ilgili ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin madde kullanım sıklığı oranlarının standart sapma değerleri arasında manidar bir fark bulunduğu şeklinde yorumlanmıştır.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin maksimum madde kullanım sıklığı oranları; madde havuzundan yararlanmama oranları; test örtüşme indeksleri arasında manidar bir fark bulunup bulunmadığı “İki oran arasındaki farkın testi” ile incelenmiştir. İki oran farkı için hipotez testlerinde, örneklem büyüklükleri ($n \geq 250$) yeterince büyük olduğundan merkezi limit teoremi gereğince iki oran farkının ($P_1 - P_2$ istatistiğinin) örnekleme dağılımı normale yaklaşmaktadır. Dolayısıyla iki oran arasındaki farka ilişkin hipotez testlerinde z dağılımı test istatistiği kullanılmıştır (Arıcıgil Çılan, 2009). Bu bilgi doğrultusunda, z değeri aşağıdaki gibi (Eşitlik 4) hesaplanmıştır (Green ve Salkind, 2007; Howell, 2010; Mazmanoğlu, 2016):

$$z = \frac{P_1 - P_2}{\sqrt{\frac{P_1(1-P_1)}{N_1} + \frac{P_2(1-P_2)}{N_2}}} \dots\dots(Eşitlik 4)$$

İki oran arasındaki farka ilişkin hesaplanan z değeri .05 manidarlık düzeyinde $z_{.05} = \pm 1.96$ kritik değerleri sınırları dışında kalırsa $H_0 = \text{İki farklı madde kullanım}$

sıklığı kontrol yöntemlerine ilişkin hesaplanan oranlar (maksimum madde kullanım sıklığı oranları vb.) arasında manidar bir fark bulunmamaktadır. ($P_1 = P_2$) hipotezi red edilmiştir. Dolayısıyla ilgili iki oran arasındaki farklılıkların manidar olduğu belirtilmiştir.

Her bir koşulda 100 tekrar sonucunda elde edilen değerlerin ortalamaları (Leroux ve diğerleri 2013; Ross, 2013; Zheng ve Chang, 2014) dikkate alınarak ölçme duyarlılığı için uyum, yanlılık ve RMSE katsayıları; test güvenliği için, maksimum madde kullanım sıklığı oranı, madde kullanım sıklığının standart sapması, madde havuzundan yararlanmama ve test örtüşme indeksi; madde kullanım sıklık oranlarını karşılaştırmak için χ^2 istatistikleri; en ideal madde kullanım sıklık yöntemine karar verilmesi için ise F istatistiğine ilişkin değerler Excel’de ayrı ayrı hesaplanmıştır. Yanlılık ve RMSE değerlerinin Mahalanobis uzaklıklarının hesaplanması için R yazılım programında yer alan {stats} paketindeki “mahalanobis” komutu kullanılmıştır (R Core Team, 2017).

BÖLÜM III

3. BULGULAR VE YORUMLAR

Bu bölümde, araştırmanın amacı çerçevesinde cevap aranan araştırma sorularına ilişkin bulgular ve yorumlar yer almaktadır.

3.1. BOBUT Uygulamalarında Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Ölçme Duyarlılığına Etkisi

BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin ölçme duyarlılığına etkisini incelemek amacıyla aşağıda sırasıyla ölçme duyarlılığı indekslerinden uyum katsayılarına, yanlışlık değerlerine ve RMSE değerlerine ilişkin bulgular ve yorumlara yer verilmiştir. Bu doğrultuda, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin uyum katsayıları Çizelge 2’de sunulmuştur.

Çizelge 2

Farklı Ölçme Koşullarına İlişkin Uyum Katsayıları

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	YOK	SH	AD
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.9990	.9988	.9954
		Normal	.9997	.9997	.9997
		Tek Biçimli	.9997	.9997	.9997
		Sola Çarpık	.9992	.9991	.9977
	Geniş (N=1000)	Sağa Çarpık	.9990	.9989	.9962
		Normal	.9997	.9997	.9997
		Tek Biçimli	.9998	.9998	.9998
		Sola Çarpık	.9993	.9993	.9985
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.9994	.9994	.9964
		Normal	.9999	.9999	.9999
		Tek Biçimli	.9999	.9999	.9999
		Sola Çarpık	.9996	.9996	.9985
	Geniş (N=1000)	Sağa Çarpık	.9995	.9996	.9973
		Normal	.9998	.9998	.9998
		Tek Biçimli	.9998	.9998	.9998
		Sola Çarpık	.9997	.9997	.9989

Çizelge 2 incelendiğinde, BOBUT uygulamalarında, farklı ölçme koşullarındaki uyum katsayılarının .9954 ile .9999 arasında değiştiği görülmektedir. Bu durumda, farklı ölçme koşullarında kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasında yüksek düzeyde uyum bulunduğu belirtilebilir. Uyum katsayılarını hesaplamak için Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılmıştır. Dolayısıyla farklı ölçme koşullarında uyum katsayıları arasında manidar fark olup olmadığı, iki korelasyon katsayısı arasındaki farkın karşılaştırılmasında kullanılan Fisher'in z testi ile incelenmiştir (Şencan, 2005). İki korelasyon katsayısı arasındaki farka ilişkin etki büyüklüğünü hesaplarken ise Cohen q ($Cohen\ q = r'_1 - r'_2$) katsayısı kullanılmıştır. Cohen q katsayısının .10'dan küçük olması etkinin olmadığı şeklinde yorumlanırken; .10 ile .30 arası küçük, .30 ile .50 arası orta düzeyde ve .50'den büyük olması ise geniş etki büyüklüğü şeklinde yorumlanmaktadır (Cohen, 1988). Bu bilgiler doğrultusunda, farklı ölçme koşullarında uyum katsayıları arasında manidar farklılık bulunan madde kullanım sıklığı kontrol yöntemlerine ait etki büyüklüklerine Çizelge 3'te yer verilmiştir.

Çizelge 3

Farklı Ölçme Koşullarına İlişkin Cohen q Değerleri

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	YOK-SH	SH-AD	YOK-AD
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.09	.67*	.76*
		Normal	.00	.00	.00
		Tek Biçimli	.03	.10	.07
		Sola Çarpık	.06	.47*	.53*
	Geniş (N=1000)	Sağa Çarpık	.05	.62*	.67*
		Normal	.00	.00	.00
		Tek Biçimli	.00	.15	.15
		Sola Çarpık	.00	.38*	.38*
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.00	.90*	.90*
		Normal	.00	.13	.13
		Tek Biçimli	.05	.05	.10
		Sola Çarpık	.00	.66*	.66*
	Geniş (N=1000)	Sağa Çarpık	.03	.90*	.87*
		Normal	.04	.02	.06
		Tek Biçimli	.00	.00	.00
		Sola Çarpık	.00	.65*	.65*

*: $p < .05$

Çizelge 3 incelendiğinde, BOBUT uygulamalarında, farklı ölçme koşullarındaki uyum katsayılarına ilişkin Cohen q değerlerinin .00 ile .90 arasında değiştiği görülmektedir. Farklı ölçme koşullarında, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığı (YOK-SH) durumlarda uyum katsayıları arasında manidar bir farklılık bulunmamaktadır ($p>.05$). Ayrıca normal ve tek biçimli yetenek dağılımlarında, kullanılan madde kullanım sıklık yöntemlerine ilişkin hesaplanan uyum katsayıları arasında manidar farklılık bulunmadığı sonucuna ulaşılmıştır ($p>.05$).

Etki büyüklüğü bakımından, sağa çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayılarına göre, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları lehine geniş düzeyde manidar bir farklılık bulunmuştur ($p<.05$). Sola çarpık dağılımlarda ise, Aşamalı Düşürme Stratejisi ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları arasında test uzunluğu kısa olduğunda orta düzeyde manidar farklılık bulunmakta iken, uzun testlerde geniş düzeyde manidar bir farklılık bulunmuştur ($p<.05$). Bunun yanı sıra, sola çarpık dağılımlarda kısa test- geniş örneklem koşulu haricinde, Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayıları ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında hesaplanan uyum katsayıları arasında geniş düzeyde manidar bir farklılık bulunmakta iken, kısa test- geniş örneklem koşulunda (*Cohen $q=.38$*) orta düzeyde manidar farklılık bulunduğu sonucuna ulaşılmıştır.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin uyum katsayıları genel olarak değerlendirildiğinde, her bir koşulda kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasında yüksek düzeyde uyum bulunduğu sonucuna ulaşılmıştır. Bunun yanı sıra diğer ölçme koşulları sabit tutulduğunda, yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça madde kullanım sıklığı kontrol yöntemlerine ilişkin uyum katsayıları her ne kadar artsa da aralarında manidar bir farklılık bulunmamıştır. Sağa ve sola çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayılarına göre, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları lehine manidar bir farklılık bulunduğu sonucuna ulaşılmıştır. Diğer ölçme koşullarında ise madde kullanım sıklığı

kontrol yöntemlerine ilişkin hesaplanan uyum katsayıları arasında manidar farklılık bulunmamıştır.

Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayıları ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları arasında kısa test - sola çarpık dağılımlarda orta düzeyde ya da orta düzeye yakın (*Cohen q*=.53) etki büyüklüğü bulunmakta iken; kısa test – sağa çarpık dağılımlarda ise, geniş düzeyde etki büyüklüğü bulunmuştur. Ayrıca uzun test- hem sağa çarpık dağılımlarda hem de sola çarpık dağılımlarda, Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayıları ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları arasında geniş düzeyde etki büyüklüğü elde edilmiştir. Dolayısıyla çarpık dağılımlarda madde kullanım sıklık yöntemlerinden Aşamalı Düşürme Stratejisi seçilmesi durumunda diğer madde kullanım sıklık yöntemlerine göre kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki uyum katsayısının önemli derecede azaldığı belirtilebilir. Bir başka ifadeyle, ilgili koşullarda kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki uyum katsayısı büyük oranda azalacağı için, hatanın artacağı söylenebilir. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin yanlılık değerleri Çizelge 4'te yer almaktadır.

Çizelge 4

Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Yanlılık Değerleri

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrolü Yapılmadığında	Sympson Hetter Stratejisi	Aşamalı Düşürme Stratejisi
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.1236	.1249	.2111
		Normal	-.0025	-.0035	-.0044
		Tek Biçimli	.0111	.0094	.1277
		Sola Çarpık	-.0922	-.0963	-.1420
	Geniş (N=1000)	Sağa Çarpık	.1113	.1135	.1844
		Normal	.0009	-.0013	-.0016
		Tek Biçimli	.0095	.0097	.0125
		Sola Çarpık	-.0881	-.0895	-.1263
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.0668	.0689	.1526
		Normal	-.0004	-.0003	-.0025
		Tek Biçimli	.0056	.0044	.0076
		Sola Çarpık	-.0522	-.0524	-.0985
	Geniş (N=1000)	Sağa Çarpık	.0587	.0597	.1275
		Normal	-.0005	-.0001	-.0002
		Tek Biçimli	.0059	.0051	.0074
		Sola Çarpık	-.0465	-.0467	-.0816

Çizelge 4 incelendiğinde, yanlılık değerlerinin $-.1420$ ile $.2111$ aralığında değiştiği görülmektedir. Uzun test, geniş örneklem, normal yetenek dağılımı ve Sympson Hetter Stratejisi koşulunda yanlılık değeri ($-.0001$) sıfıra en yakın değer olduğu için kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki ortalama farklılık en düşüktür.

Farklı değişkenlere (farklı ölçme koşullarında madde kullanım sıklığı yöntemlerinin yanlılık değerlerine) ait değişimlemelerin gruplandırılmasına/kümeler arasındaki uzaklığın hesaplanmasına imkân sunduğundan, örüntüleri bilinen sonuçların karşılaştırmasında Mahalanobis uzaklıkları kullanılabilir (Hair, Anderson, Tatham ve Black, 1998). Dolayısıyla farklı ölçme koşullarında madde

kullanım sıklığı kontrol yöntemlerine ilişkin yanlılık değerleri arasında manidar bir fark bulunup bulunmadığı Mahalanobis uzaklıkları hesaplanarak incelenmiştir.

Madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan yanlılık katsayılarının Mahalanobis uzaklık (EK-9) değerleri .19 ile 14.04 arasında değişmektedir ($p>.001$). Bu doğrultuda, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin kestirilen yetenek puanları ile gerçek yetenek puanları arasındaki ortalama farklılıklar bakımından manidar bir değişim bulunmadığı yorumu yapılabilir.

Uzun test-geniş örneklem ve normal yetenek dağılımında, madde kullanım sıklığı kontrolü yapılmadığında hesaplanan yanlılık değeri, Sympson Hetter Stratejisi ile Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan yanlılık değerlerine göre sifıra daha uzaktır. Bu koşul haricindeki tüm koşullarda, madde kullanım sıklığı kontrol yöntemleri bakımından Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan yanlılık değerlerinin, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan yanlılık değerlerine göre, sifıra daha uzak olduğu görülmektedir. Dolayısıyla diğer ölçme koşullarının sabit tutulması şartıyla, madde kullanım sıklığı kontrolü yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında kestirilen yetenek düzeyi ile gerçek yetenek düzeyi arasındaki ortalama farklılık diğer yöntemlere göre daha yüksektir.

Yanlılık değerleri her koşulda sifıra yakın değerler olduğu için, her bir koşulda kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasındaki ortalama farklılığın düşük olduğu belirtilebilir. Bunun yanı sıra, her bir ölçme koşulunda yanlılık değerlerine ilişkin elde edilen bulgularda manidar farklılıklar bulunmamakla birlikte, diğer ölçme koşulları sabit tutulduğunda, genel olarak, örneklem büyüklüğü ve test uzunluğu arttıkça, yetenek dağılımlarının çarpıklık katsayısı sifıra yaklaştıkça, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanılmadığında kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki ortalama farklılığın düştüğü bulgusuna ulaşılmıştır. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin RMSE değerlerine Çizelge 5'te yer verilmiştir.

Çizelge 5

Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin RMSE Değerleri

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrolü Yapılmadığında	Sympson Hetter Stratejisi	Aşamalı Düşürme Stratejisi
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.1523	.1584	.2715
		Normal	.0428	.0434	.0497
		Tek Biçimli	.1014	.0987	.1112
		Sola Çarpık	.1155	.1201	.1822
	Geniş (N=1000)	Sağa Çarpık	.0045	.0046	.0076
		Normal	.0015	.0015	.0017
		Tek Biçimli	.0032	.0032	.0036
		Sola Çarpık	.0036	.0036	.0051
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.0863	.089	.2043
		Normal	.0241	.0242	.0285
		Tek Biçimli	.0546	.0544	.0638
		Sola Çarpık	.0681	.0677	.1321
	Geniş (N=1000)	Sağa Çarpık	.0025	.0025	.0055
		Normal	.0008	.0009	.0009
		Tek Biçimli	.0018	.0018	.0020
		Sola Çarpık	.0020	.0020	.0035

Çizelge 5 incelendiğinde, uzun test-geniş örneklem-normal yetenek dağılımı ve madde kullanım sıklığı kontrol yöntemi uygulanmadığı koşulda en düşük RMSE değeri (.0008); kısa test-küçük örneklem-sağa çarpık dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında en yüksek RMSE değeri (.2715) bulunmuştur.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin RMSE değerleri arasında manidar bir fark bulunup bulunmadığı Mahalanobis uzaklıkları hesaplanarak incelenmiştir. Madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan RMSE katsayılarının Mahalanobis uzaklık (EK-10) değerleri .61 ile 7.32 arasında değişmektedir ($p > .001$). Bu doğrultuda, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin kestirilen yetenek düzeyleri ile

gerçek yetenek düzeyleri arasındaki mutlak farklılıklar bakımından manidar bir değişim bulunmadığı sonucuna ulaşılmıştır.

Kestirimin hatasının ortalama karekökü (Root Mean Square Error, [RMSE]), gerçek yetenek düzeyleri arasındaki kovaryans ile kestirilen yetenek düzeylerine ilişkin kovaryans matrisi arasındaki farkın bir başka ifadeyle hatanın derecesi doğrultusunda geliştirilmiş bir mutlak uyum indeksidir (Jackson, Deze, Douglas ve Shimeall, 2005; Sümer, 2000). RMSE, modelden kestirilen yetenek parametrelerinin, evren kovaryansları (gerçek yetenek parametreleri) ile ne derece uyumlu olduğunu göstermekte olup .00 ile 1.00 arasında değer almaktadır (Byrne, 1998). Modelden kestirilen yetenek parametrelerinin, gerçek yetenek parametrelerine ilişkin kovaryanslar arasındaki farklılık sıfıra yaklaştıkça, modelin uyumlu olduğu söylenebilmektedir (Byrne, 1998; Kline, 2005).

RMSE değerinin .05'den küçük olması mükemmel uyum (Brown, 2006; Jöreskog ve Sörbom, 1993); .05 ile .08 arasında olması kabul edilebilir uyum (Hooper, Coughlan ve Mullen, 2008; Howell, 2010; Tabachnick ve Fidell, 2007) ve .10'dan küçük olması zayıf uyum (Kelloway, 1998) olarak değerlendirilmektedir. Eğer örneklem küçükse ve modelde hesaplanan parametre sayısı fazla ise RMSE değeri .10'dan büyük çıkabilmektedir (Şimşek, 2007). Bu bilgiler doğrultusunda uzun test-geniş örneklemelerde mükemmel düzeyde uyum bulunduğu görülmüştür. Ayrıca farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin RMSE değerlerine göre, en yüksekten en düşüğe doğru Aşamalı Düşürme Stratejisi, Sympson Hetter Stratejisi, madde kullanım sıklığı kontrol yöntemi uygulanmadığında hesaplanan RMSE değerleri olarak sıralanmaktadır. Buna göre, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemi uygulanmadığında, madde kullanım sıklığı kontrolünün yapıldığı koşullara göre, kestirilen yetenek parametrelerine ilişkin kovaryans ile gerçek yetenek parametrelerine ilişkin kovaryans arasındaki farklılığın sıfıra daha çok yaklaştığı ortaya çıkmıştır. Dolayısıyla, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemi uygulanmadığında hesaplanan RMSE değerleri, Sympson Hetter Stratejisi ve Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan RMSE değerlerine göre daha küçük olduğundan, madde kullanım sıklığı kontrol yöntemi uygulanmadığında kestirilen yetenek düzeyleri ile

gerçek yetenek düzeyleri arasındaki mutlak farklılıkların azalabileceği çıkarımında bulunulabilir.

Farklı ölçme koşullarında RMSE değerlerine ilişkin elde edilen bulgularda manidar farklılıklar bulunmamakla birlikte, diğer ölçme koşulları sabit tutulduğunda, genel olarak, örneklem büyüklüğü ve test uzunluğu arttıkça, yetenek dağılımlarının çarpıklık katsayısı sifıra yaklaştıkça ve madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanılmadığında kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki mutlak farklılık azalmıştır.

BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin ölçme duyarlılığına etkisini inceleyebilmek için, ölçme duyarlılığı arttığında, hata değerleri azalacağı için, her bir koşulda hata göstergeleri olan uyum, yanlılık ve RMSE katsayıları birlikte ele alınarak değerlendirilmelidir (Gu ve Reckase, 2007; Wang ve Vispoel, 1998; Zheng ve Chang, 2014). Bu bilgi doğrultusunda, ölçme duyarlılığına ilişkin elde edilen bulgular genel olarak değerlendirildiğinde, örneklem büyüklüğü ve test uzunluğu arttıkça yetenek dağılımlarının çarpıklık katsayısı sifıra yaklaştıkça genelde ölçme duyarlılığının arttığı sonucuna ulaşılmıştır. Ayrıca madde kullanım sıklığı kontrolü yapılmadığında sifıra en yakın yanlılık ve en düşük RMSE değeri bulgusuna ulaşılsa da yanlılık ve RMSE değerleri bakımından madde kullanım sıklığı kontrolü yöntemleri arasında manidar farklılıklar bulunmadığı sonucuna ulaşılmıştır.

Uyum katsayıları bakımından ise, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığı durumlara ilişkin uyum katsayıları arasında manidar bir farklılık bulunmamakta iken; sağa ve sola çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayılarına göre, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları lehine manidar bir farklılık bulunmuştur.

Madde kullanım sıklığı kontrolü yapılmadığında, madde kullanım sıklığı kontrolünün yapıldığı koşullara göre ölçme duyarlılığının genelde daha yüksek olduğu bulgusuna ulaşılmıştır. Elde edilen bu bulgu, kullanım sıklığının kontrol edilmediği durumun referans olarak alındığı Boyd (2003)'ün ve Davis (2004)'in

çalışmasındaki bulgularla paralellik göstermektedir. BOBUT uygulamalarında madde seçiminde madde bilgi fonksiyonu ölçüt alındığından, madde kullanım sıklığı kontrolü yapılmadığında madde seçimi kısıtlanmayacağı için, yüksek bilgi değerine sahip maddeler seçilerek yetenek kestiriminin standart hatasının düşmesi bu bulguya gerekçe gösterilebilir. Bir başka ifadeyle, yüksek bilgi değerine sahip maddeler seçildiğinde yetenek kestiriminin standart hatasının düşmesi, madde kullanım sıklığı kontrolü yapılmadığında, madde kullanım sıklığı kontrolünün yapıldığı koşullara göre ölçme duyarlılığının genelde daha yüksek bulunmasına neden olduğu söylenebilir.

Ölçme duyarlılığını maksimum yapacak maddeler seçildiğinde, kullanım sıklığı oranları düzgün dağılmayacağından (Pastor, Dodd ve Chang, 2002) madde havuzunun dengeli bir şekilde kullanılmayıp sadece sınırlı sayıda bazı maddelerin kullanılmasına sebep olmamak için (Hulin, Drasgow ve Parson 1983; akt. Revuelta ve Ponsoda, 1998) madde kullanım sıklığı kontrol yöntemleriyle, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin sağlanması amaçlanmaktadır (Boyd, 2003; Boyd, Dodd ve Fitzpatrick, 2013; Davis ve Dodd, 2005; Pastor, Dodd ve Chang, 2002). Bu durumda, ilgili değişkenler birlikte ele alındığında, farklı madde kullanım sıklığı kontrolü koşullarında ölçme duyarlılıkları arasında genel olarak büyük farklılıklar bulunmamıştır. Bu bulgu sonucunda, ilgili madde kullanım sıklığı kontrol yöntemleri kullanıldığında, ölçme duyarlılığının düşürülmeyeceği ifade edilebilir.

3.2. BOBUT Uygulamalarında Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Test Güvenliğine Etkisi

BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin test güvenliğine etkisini incelemek amacıyla aşağıda sırasıyla madde kullanım sıklığı oranlarının standart sapma değerleri, maksimum kullanım sıklık oranı dağılımları, madde havuzundan yararlanmama indeksleri, madde kullanım sıklık oranlarının çarpıklık katsayıları (χ^2 değerleri) ve test örtüşme indekslerine ilişkin bulgular ve yorumlara yer verilmektedir. Bu doğrultuda, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin madde kullanım sıklığı oranlarının standart sapma değerleri Çizelge 6'da sunulmuştur.

Çizelge 6

Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Madde Kullanım Sıklığı Oranlarının Standart Sapma Değerleri

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrolü Yapılmadığında	Sympson Hetter Stratejisi	Aşamalı Düşürme Stratejisi
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.097	.091	.048
		Normal	.074	.071	.037
		Tek Biçimli	.068	.064	.038
		Sola Çarpık	.101	.092	.044
	Geniş (N=1000)	Sağa Çarpık	.095	.090	.047
		Normal	.075	.072	.037
		Tek Biçimli	.068	.064	.038
		Sola Çarpık	.099	.091	.044
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.142	.138	.063
		Normal	.098	.096	.054
		Tek Biçimli	.089	.087	.055
		Sola Çarpık	.143	.137	.060
	Geniş (N=1000)	Sağa Çarpık	.139	.136	.063
		Normal	.099	.098	.054
		Tek Biçimli	.089	.086	.055
		Sola Çarpık	.139	.134	.060

Çizelge 6 incelendiğinde, madde kullanım sıklığı oranlarının standart sapma değerlerine ilişkin, en yüksek değer uzun test-küçük örneklem-sola çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında (.143); en düşük değer kısa test-küçük örneklem-normal dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında (.037) bulunmuştur. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin madde kullanım sıklığı oranlarının standart sapma değerleri arasında manidar bir fark bulunup bulunmadığı “İki varyansa ilişkin hipotez testi” ile incelenmiştir.

Farklı ölçme koşullarında *küçük örneklem*lere ilişkin karşılaştırmalarda pay ve paydanın serbestlik dereceleri 249 ve $.025^{F_{249,249,.05}} = \frac{1}{F_{249,249,.025}} = .78 < F_h < F_{249,249,.025} = 1.28$; *geniş örneklem*lere ilişkin karşılaştırmalarda pay ve paydanın serbestlik dereceleri 999 ve $.025^{F_{999,999,.05}} = \frac{1}{F_{999,999,.025}} = .88 < F_h < F_{999,999,.025} = 1.13$ olmak üzere, F_h ilgili güven aralığında (H_0 kabul bölgesinde) ise H_0 hipotezi red edilmemiştir.

İlgili ölçme koşullarındaki, ikili örneklem varyanslarının birbirlerine oranları doğrultusunda hesaplanan F_{YOK-AD} ve F_{SH-AD} değerleri, tüm ölçme koşullarında red bölgesinde yer almakta iken; sadece F_{YOK-SH} kısa test-geniş örneklem-sola çarpık yetenek dağılımı koşulu haricindeki tüm ölçme koşullarında kabul bölgesinde yer aldığı görülmüştür. Dolayısıyla madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında (kısa test-geniş örneklem-sola çarpık yetenek dağılımı koşulu haricinde) madde kullanım sıklığı oranlarının standart sapma değerleri arasında manidar bir farklılık bulunmamıştır ($p > .05$). Ayrıca Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan madde kullanım sıklığı oranlarının standart sapma değerleri ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında (F_{YOK-AD}) ve Sympson Hetter Stratejisi kullanıldığında hesaplanan (F_{SH-AD}) madde kullanım sıklığı oranlarının standart sapma değerleri arasında manidar bir farklılık bulunduğu sonucuna ulaşılmıştır ($p < .05$).

Kısa test-geniş örneklem-sola çarpık yetenek dağılımı koşulunda madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında madde kullanım sıklığı oranlarının standart sapma değerleri arasında her ne kadar manidar farklılık bulunsun da F değerinin kritik değere çok yakın olduğu saptanmıştır ($F_{YOK-SH} = 1.18 > F_{999,999,.025} = 1.13$). Bu bulguya gerekçe olarak, sola çarpık yetenek dağılımındaki (üstün yetenekli) bireylerin (geniş örneklem) yetenek düzeylerine hitap edebilecek madde havuzundaki maddeleri seçmede yaşanan güçlüklerin yanı sıra test sonlandırma kuralının kısa test olması, madde seçme yönteminde de önsel dağılım kullanılarak bir kısıtlamanın yapılıp yapılmamasına göre madde kullanım sıklığı oranları arasındaki çarpıklığın artmasına neden olabileceği söylenebilir.

Farklı ölçme koşullarında madde kullanım sıklığı oranlarının standart sapma değerlerine ilişkin elde edilen bulgular genel olarak değerlendirildiğinde, SH ile YOK arasında genelde manidar farklılık bulunmamakta iken; SH ile AD arasında ve YOK ile AD arasında manidar farklılık bulunmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan madde kullanım sıklığı oranlarının standart sapma değerleri, diğer iki yöntemle göre manidar bir şekilde daha küçüktür. Dolayısıyla Aşamalı Düşürme Stratejisi kullanıldığında madde kullanım sıklığı oranlarının diğer stratejilere göre manidar bir şekilde daha homojen dağılım gösterdiği söylenebilir.

Diğer ölçme koşulları sabit tutulduğunda, genel olarak, test uzunluğu azaldıkça, yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça, örneklem büyüklüğü genelde arttıkça ve madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan madde kullanım sıklığı oranlarının standart sapma değerleri azaldığından madde kullanım sıklığı oranlarının daha homojen dağılım gösterdiği ortaya çıkmaktadır. Bir başka ifadeyle, ilgili koşullarda madde kullanım sıklığı oranları daha homojen dağılım gösterdiğinden, madde kullanım sıklığı dengeli dağılım göstermekte ve madde havuzu daha verimli kullanılmaktadır. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin maksimum madde kullanım sıklığı oranı dağılımları Çizelge 7'de yer almaktadır.

Çizelge 7

Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Maksimum Madde Kullanım Sıklığı Oranı Dağılımları

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrolü Yapılmadığında	Sympson Hetter Stratejisi	Aşamalı Düşürme Stratejisi
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.834	.622	.193
		Normal	.750	.486	.117
		Tek Biçimli	.677	.448	.152
		Sola Çarpık	.922	.642	.184
	Geniş (N=1000)	Sağa Çarpık	.830	.594	.189
		Normal	.739	.492	.117
		Tek Biçimli	.683	.454	.151
		Sola Çarpık	.915	.623	.181
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.840	.727	.200
		Normal	.761	.508	.151
		Tek Biçimli	.686	.463	.183
		Sola Çarpık	.929	.733	.196
	Geniş (N=1000)	Sağa Çarpık	.836	.721	.197
		Normal	.749	.523	.151
		Tek Biçimli	.689	.461	.181
		Sola Çarpık	.921	.714	.192

Çizelge 7 incelendiğinde, maksimum madde kullanım sıklığı oranlarına ilişkin, en yüksek değer uzun test-küçük örneklem-sola çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında (.929); en düşük değer kısa test-hem küçük hem de geniş örneklem-normal dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında (.117) bulunmuştur.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin *maksimum madde kullanım sıklığı oranları* arasında manidar bir farklılık bulunup bulunmadığı “iki oran arasındaki farkın testi” ile incelenmiştir (Green ve Salkind, 2007). Madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan *maksimum madde kullanım sıklığı oranları arasındaki farklara ilişkin z* değerleri 5.19 ile 48.82

arasında deęiřtięinden ilgili iki oran arasındaki farklılıkların manidar olduęu sonucuna ulařılmıřtır ($p<.05$).

Farklı ölçme kořullarında madde kullanım sıklıęı kontrol yöntemlerine iliřkin hesaplanan maksimum madde kullanım sıklıęı oranlarına göre, en düşükten en yükseęe doęru Ařamalı Düşürme Stratejisi, Sympson Hetter Stratejisi, madde kullanım sıklıęı kontrolünün yapılmadıęı kořul olarak sıralanmaktadır. Maddelerin kullanım sıklıęı oranı arttıkça maddelerin ifřâ olma ihtimalleri arttıęı için ilgili kořullarda Ařamalı Düşürme Stratejisi kullanıldıęında, maddelerin ifřâ olarak psikometrik özelliklerini kaybetme riskinin dięer kořullara göre manidar bir řekilde azaldıęı görölmektedir.

Dięer ölçme kořulları sabit tutulduęunda, genel olarak test uzunluęu azaldıkça, yetenek daęılımlarının çarpıklık katsayısı sıfıra yaklařtıkça, örneklem büyüklüęü genelde arttıkça ve madde kullanım sıklıęı kontrol yöntemlerinden Ařamalı Düşürme Stratejisi kullanıldıęında hesaplanan maksimum madde kullanım sıklıęı oranları azaldıęından maddelerin ifřâ olarak psikometrik özelliklerini kaybetme riskinin azaldıęı ifade edilebilir. Farklı ölçme kořullarında madde kullanım sıklıęı kontrol yöntemlerine iliřkin BOBUT sürecinde uygulanmayan maddelerin yüzdesi bir bařka ifadeyle madde havuzundan yararlanmama indeksleri Çizelge 8'de yer almaktadır.

Çizelge 8

Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Madde Havuzundan Yararlanmama İndeksleri

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrolü Yapılmadığında	Sympson Hetter Stratejisi	Aşamalı Düşürme Stratejisi
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.1259	.1256	.0980
		Normal	.1254	.1253	.0939
		Tek Biçimli	.1242	.1242	.0952
		Sola Çarpık	.1272	.1264	.0928
	Geniş (N=1000)	Sağa Çarpık	.0308	.0308	.0243
		Normal	.0309	.0309	.0232
		Tek Biçimli	.0310	.0309	.0238
		Sola Çarpık	.0312	.0308	.0239
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.0534	.0528	.0312
		Normal	.0530	.0536	.0322
		Tek Biçimli	.0510	.0504	.0330
		Sola Çarpık	.0533	.0530	.0288
	Geniş (N=1000)	Sağa Çarpık	.0127	.0126	.0075
		Normal	.0127	.0126	.0078
		Tek Biçimli	.0126	.0126	.0082
		Sola Çarpık	.0128	.0127	.0069

Her bir koşulda madde havuzunun kullanılma derecesini belirleyebilmek için, madde havuzundan yararlanmama oranları incelendiğinde, en yüksek değer kısa test-küçük örneklem-sağa çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında (.1259); en düşük değer uzun test-geniş örneklem-sola çarpık dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında (.0069) bulunmuştur. Dolayısıyla farklı ölçme koşullarında madde havuzunun kullanılma derecesi $(1-.1259=\%87.41)$ ile $(1-.0069=\%99.31)$ arasında değişim göstermektedir.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin madde havuzundan yararlanmama oranları arasında manidar bir fark bulunup

bulunmadığı “iki oran arasındaki farkın testi” ile incelenmiştir. Madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan *madde havuzundan yararlanmama oranları arasındaki farklara ilişkin z* değerleri -.03 ile 1.38 arasında değiştiğinden ilgili iki oran arasındaki farklılıkların manidar olmadıkları sonucuna ulaşılmıştır ($p>.05$).

Farklı ölçme koşullarında madde havuzundan yararlanmama oranları arasında manidar farklılıklar bulunmamakla birlikte, diğer ölçme koşulları sabit tutulduğunda, örneklem büyüklüğü ve test uzunluğu arttıkça, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde havuzundan yararlanmama oranları azalmaktadır. Dolayısıyla örneklem büyüklüğü ve test uzunluğu arttıkça, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde havuzunun kullanılma derecesinin arttığı söylenebilir.

Madde kullanım sıklık yöntemlerine ilişkin madde kullanım sıklık oranlarının istenilen kullanım sıklık oranına uzaklığının belirlenmesi bir başka ifadeyle madde havuzu kullanımının ne kadar dengeli dağılım gösterdiğinin saptanabilmesi için farklı ölçme koşullarında madde kullanım sıklık oranlarının çarpıklık katsayısı (χ^2 değerleri) Çizelge 9'da sunulmuştur.

Çizelge 9

Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Madde Kullanım Sıklık Oranlarının Çarpıklık Katsayıları (χ^2 Değerleri)

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrolü Yapılmadığında	Sympson Hetter Stratejisi	Aşamalı Düşürme Stratejisi
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	375.06	330.99	90.78
		Normal	218.66	201.84	54.50
		Tek Biçimli	183.84	165.06	58.49
		Sola Çarpık	406.71	341.64	77.97
	Geniş (N=1000)	Sağa Çarpık	362.35	321.22	89.39
		Normal	224.52	208.19	54.17
		Tek Biçimli	182.74	164.07	58.23
		Sola Çarpık	387.70	328.23	76.78
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	403.50	381.86	79.54
		Normal	193.19	185.85	58.51
		Tek Biçimli	158.61	150.22	60.25
		Sola Çarpık	406.23	375.25	72.14
	Geniş (N=1000)	Sağa Çarpık	389.23	369.19	78.38
		Normal	197.60	191.56	57.01
		Tek Biçimli	157.27	148.79	59.82
		Sola Çarpık	387.87	359.85	71.50

Çizelge 9 incelendiğinde, en yüksek χ^2 değeri kısa test-küçük örneklem-sola çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında ($\chi^2 = 406.71$) elde edilirken; en düşük χ^2 değeri kısa test-geniş örneklem-normal dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında en düşük ($\chi^2 = 54.17$) elde edilmiştir.

Madde kullanım sıklığı kontrolü yapılmadığında ya da madde kullanım sıklığı kontrol yöntemlerinden Sympson Hetter stratejisi ve Aşamalı Düşürme stratejisi kullanıldığında hangisinin en ideal madde kullanım sıklık yöntemi olduğuna karar verilebilmesi için, karşılaştırma ölçüsü olarak F istatistiği

$(F_{\chi^2_{yöntem1}, \chi^2_{yöntem2}} = \frac{\chi^2_{yöntem1}}{\chi^2_{yöntem2}})$ tanımlanmıştır. $F_{\chi^2_{yöntem1}, \chi^2_{yöntem2}} < 1$ olduğunda, birinci yöntemin madde kullanım sıklığı oranlarının genel olarak dengelenmesi bakımından, ikinci yönteme göre daha iyi bir yöntem olduğu şeklinde yorumlanmıştır (Chang ve Ying, 1999; Grubbs, 1973).

Farklı ölçme koşullarında hesaplanan χ^2 değerlerine göre, en düşükten en yükseğe doğru Aşamalı Düşürme Stratejisi, Sympson Hetter Stratejisi, madde kullanım sıklığı kontrol yöntemi uygulanmadığında hesaplanan χ^2 değerleri olarak sıralanmaktadır. Dolayısıyla F istatistiklerine ilişkin her bir koşulda $F_{\chi^2_{AD}, \chi^2_{SH}} < 1$ ve $F_{\chi^2_{YOK}, \chi^2_{SH}} > 1$ sonuçlarına ulaşılmıştır. Bu durumda her bir koşulda, madde kullanım sıklığı oranlarının genel olarak dengelenmesi bakımından, madde kullanım sıklığı kontrolü yapılmamasının, Sympson Hetter Stratejisine göre daha kullanışsız ($F_{\chi^2_{YOK}, \chi^2_{SH}} > 1$) bir yöntem olduğu; Aşamalı Düşürme Stratejisinin ise Sympson Hetter Stratejisine göre daha iyi bir yöntem ($F_{\chi^2_{AD}, \chi^2_{SH}} < 1$) olduğu görülmektedir.

Elde edilen bulgular ve ilgili koşullar doğrultusunda madde kullanım sıklığı kontrolü yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde havuzunun daha dengeli dağılım gösterdiği sonucuna ulaşılmıştır. Bu sonuca benzer olarak, EK-3'te yer alan BOBUT uygulamalarında küçük örneklem, normal yetenek dağılımı, kısa test ve madde kullanım sıklığı kontrolü yapılmadığında (Grafik 7); Sympson Hetter Stratejisi kullanıldığında (Grafik 8) ve Aşamalı Düşürme Stratejisi kullanıldığında (Grafik 9) madde kullanım sıklığı oranlarının dağılımına ilişkin grafikler karşılaştırıldığında Aşamalı Düşürme Stratejisi için madde havuzunun daha dengeli dağılım gösterdiği görülmüştür.

Diğer ölçme koşulları sabit tutulduğunda, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında *test uzunluğu arttığında*, χ^2 değeri sağa çarpık dağılımda ve sola çarpık dağılımda (küçük örneklem-kontrol yöntemi uygulanmadığı koşul dışında) artmakta iken normal ve tek biçimli dağılımlarda χ^2 değerleri azalmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında ise diğer iki strateji koşullarında elde edilen bulguların aksine, *test uzunluğu arttığında*, χ^2 değeri sağa çarpık dağılımda ve sola çarpık dağılımda azalmakta iken normal ve tek biçimli dağılımlarda χ^2 değerleri artmaktadır.

Örneklem büyüklüğü arttığında sağa çarpık, sola çarpık ve tek biçimli dağılımlarda, χ^2 değerleri azalmaktadır. Normal yetenek dağılımlarında ise, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında χ^2 değeri artmakta iken; Aşamalı Düşürme Stratejisi kullanıldığında χ^2 değeri azalmaktadır.

Yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça, χ^2 değeri azalmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında sırasıyla *sağa çarpık, sola çarpık, tek biçimli ve normal dağılıma doğru* χ^2 değeri azalmaktadır. Ayrıca, Aşamalı Düşürme Stratejisi kullanıldığında, *normal dağılımlarda* χ^2 değeri en küçük iken; madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında *tek biçimli dağılımlarda* χ^2 değeri en küçüktür.

Farklı ölçme koşullarında madde kullanım sıklık oranlarının çarpıklık katsayılarına (χ^2 değerlerine) ilişkin elde edilen bulgular genel olarak incelendiğinde, Aşamalı Düşürme Stratejisi kullanıldığında en düşük χ^2 değeri, kısa test-geniş örneklem-normal dağılım koşulunda elde edilmiştir. Madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında, en düşük χ^2 değeri, uzun test-geniş örneklem- tek biçimli dağılım koşulunda elde edilmiştir. Fakat tüm koşullarda da Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan χ^2 değeri diğer stratejilere göre daha küçük bulunmuştur. Dolayısıyla madde kullanım sıklık yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde kullanım sıklık oranının istenilen kullanım sıklık oranına (kısa testlerde .025; uzun testlerde .05) uzaklığının azaldığı söylenebilir. Başka bir deyişle Aşamalı Düşürme Stratejisi kullanıldığında, madde kullanım sıklık oranının çarpıklık katsayısı sıfıra yaklaştığı için madde havuzu kullanımı daha dengeli dağılım göstermektedir. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin test örtüşme indekslerine Çizelge 10'da yer verilmektedir.

Çizelge 10

Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine İlişkin Test Örtüşme İndeksleri

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrolü Yapılmadığında	Sympson Hetter Stratejisi	Aşamalı Düşürme Stratejisi
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.396	.353	.112
		Normal	.239	.223	.076
		Tek Biçimli	.205	.187	.079
		Sola Çarpık	.429	.363	.099
	Geniş (N=1000)	Sağa Çarpık	.386	.346	.114
		Normal	.249	.232	.078
		Tek Biçimli	.207	.188	.082
		Sola Çarpık	.412	.352	.101
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.452	.429	.126
		Normal	.239	.232	.105
		Tek Biçimli	.205	.197	.107
		Sola Çarpık	.454	.422	.119
	Geniş (N=1000)	Sağa Çarpık	.439	.419	.128
		Normal	.247	.241	.106
		Tek Biçimli	.206	.198	.109
		Sola Çarpık	.437	.409	.121

Çizelge 10 incelendiğinde, farklı ölçme koşullarında test örtüşme indekslerine ilişkin, en yüksek değer uzun test-küçük örneklem-sola çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında (.454); en düşük değer kısa test-küçük örneklem-normal dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında (.076) bulunmuştur.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin test örtüşme indeksleri arasında manidar bir fark bulunup bulunmadığı “İki oran arasındaki farkın testi” ile incelenmiştir. Madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan test örtüşme indeksleri arasındaki farklara ilişkin (EK-7) z_{YOK-AD} değerleri 3.05 ile 17.06 arasında değişmekte iken; z_{SH-AD} değerleri 2.82 ile 15.47

arasında değişmektedir. Dolayısıyla İlgili ölçme koşullarında, Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan test örtüşme indeksleri ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında (Z_{YOK-AD}) ve Simpson Hetter Stratejisi kullanıldığında hesaplanan (Z_{SH-AD}) test örtüşme indeksleri arasında manidar farklılık olduğu sonucuna ulaşılmıştır ($p<.05$).

Madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Simpson Hetter Stratejisi kullanıldığında (kısa test-geniş örneklem-sola çarpık yetenek dağılımı koşulu $Z_{YOK-SH} = 2.78$ haricinde) hesaplanan test örtüşme indeksleri arasında manidar bir farklılık bulunmamıştır ($p>.05$).

Farklı ölçme koşullarında hesaplanan test örtüşme indekslerine ilişkin elde edilen bulgular genel olarak değerlendirildiğinde, SH ile YOK arasında genelde manidar farklılık bulunmamakta iken; SH ile AD arasında ve YOK ile AD arasında manidar farklılık bulunmuştur. Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan test örtüşme indekslerinin diğer iki yöntemle göre manidar bir şekilde daha küçük olduğu ortaya çıkmıştır. Dolayısıyla test örtüşme indeksi büyüdükçe, test güvenliği zedeleneyeceğinden (Huang, Chen ve Wang, 2012). Aşamalı Düşürme stratejisi kullanıldığında, madde kullanım sıklığı kontrolü yapılmadığı ve Simpson Hetter stratejisi kullanıldığı koşullara göre test güvenliğinin daha az zedelendiği ifade edilebilir.

Test uzunluğu ve yetenek dağılımı sabit tutulup, madde kullanım sıklığı kontrol yöntemi uygulanmadığı ve Simpson Hetter Stratejisi kullanıldığı koşullarda *örneklem büyüklüğü arttığında* sağa çarpık ve sola çarpık dağılımlarda, test örtüşme indeksleri azalmakta iken; normal ve tek biçimli yetenek dağılımlarında ise, test örtüşme indeksleri artmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında *örneklem büyüklüğü arttıkça* test örtüşme indeksleri de artmaktadır.

Test uzunluğu ve yetenek dağılımı sabit tutulup örneklem büyüklüğü arttığında; madde kullanım sıklığı kontrol yöntemi uygulanmadığı ve Simpson Hetter Stratejisi kullanıldığı koşullarda, sağa çarpık ve sola çarpık dağılımlarda, test örtüşme indekslerinde azalma olsa da normal ve tek biçimli yetenek dağılımlarına göre test örtüşme indeksleri daha yüksek olduğu için test güvenliğinin daha olumsuz yönde etkilendiği yorumu yapılabilir. Ayrıca Aşamalı Düşürme Stratejisi kullanıldığında *örneklem büyüklüğü arttıkça* test örtüşme indeksleri artsa da diğer madde kullanım sıklığı kontrol yöntemlerinde hesaplanan test örtüşme indekslerine göre, her bir

koşulda manidar bir şekilde test örtüşme indeksleri daha küçük olduğundan Aşamalı Düşürme Stratejisi kullanıldığında test güvenliğinin daha olumlu yönde etkileneceği ifade edilebilir.

Diğer ölçme koşulları sabit tutulduğunda, madde kullanım sıklığı kontrol yöntemi uygulanmadığı ve Sympson Hetter Stratejisi kullanıldığı koşullarda *test uzunluğu arttığında*, test örtüşme indeksleri sağa ve sola çarpık dağılımda artmakta iken normal ve tek biçimli dağılımda büyük farklılık bulunmamıştır. O halde test uzunluğu arttıkça; sağa ve sola çarpık yetenek dağılımlarında hesaplanan test örtüşme indeksleri ile normal ve tek biçimli yetenek dağılımlarında hesaplanan test örtüşme indeksleri arasındaki genişlik daha da arttığından, sağa ve sola çarpık yetenek dağılımlarında test güvenliğinin daha olumsuz yönde etkilendiği çıkarımında bulunulabilir. Aşamalı Düşürme Stratejisi kullanıldığında *test uzunluğu arttıkça* test örtüşme indeksleri artsa da diğer madde kullanım sıklığı kontrol yöntemlerinde hesaplanan test örtüşme indekslerine göre, her bir koşulda manidar bir şekilde daha küçük olduğu için Aşamalı Düşürme Stratejisi kullanıldığında test güvenliğinin daha olumlu yönde etkilendiği söylenebilir.

Yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça, test örtüşme indeksinin azaldığı görülmüştür. Aşamalı Düşürme Stratejisi kullanıldığında sırasıyla *sağa çarpık*, *sola çarpık*, *tek biçimli* ve *normal dağılıma doğru* test örtüşme indeksi azalmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında, *normal dağılımlarda* test örtüşme indeksi en küçük iken; madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında *tek biçimli dağılımlarda* test örtüşme indeksi en küçüktür.

Farklı ölçme koşullarında test örtüşme indekslerine ilişkin elde edilen bulgular genel olarak incelendiğinde, Aşamalı Düşürme Stratejisi kullanıldığında en düşük test örtüşme indeksi, kısa test - küçük örneklem - normal dağılım koşulunda elde edilmiştir. Madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında, kısa test - küçük örneklem - tek biçimli dağılım koşulunda en düşük test örtüşme indeksi bulgusuna ulaşılmıştır. Fakat tüm koşullarda da Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan test örtüşme indeksleri diğer stratejilere göre manidar şekilde daha küçük bulunmuştur. Dolayısıyla BOBUT uygulamalarında, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında, herhangi iki bireyin karşılaştıkları aynı maddelerin sayısı

azaldığından hem test güvenliği daha az zedelendiği hem de madde havuzunun yıllarca kullanılabilir olmasına imkân sunduğu belirtilebilir.

Madde kullanım sıklığı oranının $r_i = .20$ olması, ilgili maddenin yaklaşık olarak her beş BOBUT uygulamasından birinde uygulandığı anlamına gelmektedir (Weiss ve Guyer, 2012). Dolayısıyla maddenin kullanım sıklığı oranı arttıkça, ilgili maddenin ifşâ edilerek psikometrik özelliklerinin yapay olarak düşmesine neden olmaktadır (Segall, 2004; Revuelta ve Ponsoda, 1998). Bu nedenle, madde havuzundaki maddelerin ne çok sık kullanılması ne de hiç kullanılmaması istenen durumdur. Bu bilgiler doğrultusunda, madde havuzunun etkili kullanılma/işlevsellik derecesini belirleyebilmek için farklı ölçme koşullarında madde havuzunun kullanımına ilişkin dağılıma Çizelge 11’de yer verilmiştir.

Çizelge 11

Farklı Ölçme Koşullarında Madde Havuzunun Kullanımına İlişkin Dağılım

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrol Yöntemi	$r_i=0$	$0<r_i\leq.2$	$.2<r_i\leq.4$	$.4<r_i\leq.6$	$.6<r_i\leq.8$	$.8<r_i\leq 1$
Kısa (25 madde)	Küçük (n=250)	Sağa çarpık	YOK	787	168	18	20	6	1
			SH	785	165	28	20	2	0
			AD	613	387	0	0	0	0
		Normal	YOK	784	177	33	3	3	0
			SH	783	173	36	8	0	0
			AD	587	413	0	0	0	0
		Tek Biçimli	YOK	776	191	28	4	1	0
			SH	776	181	42	1	0	0
			AD	595	405	0	0	0	0
		Sola Çarpık	YOK	795	168	13	15	6	3
			SH	790	166	19	23	2	0
			AD	608	392	0	0	0	0
	Geniş (N=1000)	Sağa Çarpık	YOK	771	186	19	19	4	1
			SH	770	180	28	22	0	0
			AD	608	392	0	0	0	0
		Normal	YOK	773	186	32	6	3	0
			SH	772	184	35	9	0	0
			AD	581	419	0	0	0	0
		Tek Biçimli	YOK	774	195	26	4	1	0
			SH	773	185	41	1	0	0
			AD	596	404	0	0	0	0
		Sola Çarpık	YOK	780	182	16	15	5	2
			SH	771	184	21	23	1	0
			AD	598	402	0	0	0	0
			SH	633	286	31	32	18	0
			AD	344	656	0	0	0	0

Çizelge 11’in Devamı

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrol Yöntemi	$r_i=0$	$0<r_i\leq.2$	$.2<r_i\leq.4$	$.4<r_i\leq.6$	$.6<r_i\leq.8$	$.8<r_i\leq 1$
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	YOK	668	251	28	26	24	3
			SH	660	253	37	26	24	0
			AD	390	610	0	0	0	0
		Normal	YOK	667	226	95	9	3	0
			SH	670	222	98	10	0	0
			AD	403	597	0	0	0	0
		Tek Biçimli	YOK	634	279	82	4	1	0
			SH	630	274	95	1	0	0
			AD	412	588	0	0	0	0
		Sola Çarpık	YOK	666	261	26	16	28	3
			SH	662	259	29	26	24	0
			AD	360	640	0	0	0	0
	Geniş (N=1000)	Sağa Çarpık	YOK	635	280	33	31	18	3
			SH	633	279	36	32	20	0
			AD	374	626	0	0	0	0
		Normal	YOK	633	270	85	9	3	0
			SH	631	272	81	16	0	0
			AD	390	610	0	0	0	0
		Tek Biçimli	YOK	631	283	81	4	1	0
			SH	630	274	95	1	0	0
			AD	412	588	0	0	0	0
		Sola Çarpık	YOK	639	287	28	20	23	3
			SH	633	286	31	32	18	0
			AD	344	656	0	0	0	0

Çizelge 11 incelendiğinde, 100 replikasyon sonucunda her bir BOBUT uygulamasının hiçbirinde kullanılmayan ($r_i = 0$) madde sayısı madde kullanım sıklığı kontrol yöntemi uygulanmadığında 631 ile 795; Sympson Hetter Stratejisi'nde 630 ile 790; Aşamalı Düşürme Stratejisi'nde 344 ile 613 aralığında değişmektedir. İlgili koşullarda madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında her bir BOBUT uygulamasının hiçbirinde kullanılmayan ($r_i = 0$) madde sayısı bakımından büyük farklılıklar bulunmamıştır. Aşamalı Düşürme Stratejisi kullanıldığında ise, diğer stratejilere göre tüm koşullarda kullanılmayan ($r_i = 0$) madde sayısı bakımından daha küçük olduğu sonucuna ulaşılmıştır.

Her bir BOBUT uygulamasının hiçbirinde kullanılmayan ($r_i = 0$) madde sayısının bir başka ifadeyle, madde havuzundan yararlanmama (*EK-4 Grafik 10; EK-5 Grafik 15; EK-6 Grafik 20*) bakımından uzun test ve geniş örneklemelerde diğer ölçme koşullarına göre daha düşük değerlerde olduğu belirtilebilir. Benzer olarak *EK-4 Grafik 11; EK-5 Grafik 16* ve *EK-6 Grafik 21* incelendiğinde de madde havuzunun kullanılma derecesi bakımından uzun test ve geniş örneklemelerde madde havuzunun kullanılma derecesinin daha yüksek olduğu görülmüştür.

Bütün maddeler için istenilen tek biçimli kullanım sıklık oranları (Tay, 2015) kısa test için .025, uzun test için .05'dir. Bunun yanı sıra ilgili literatürde BOBUT uygulamalarında hedeflenen kullanım sıklık oranı .20 (Eignor, Stocking, Way ve Steffen, 1993; Eggen, 2001; Veldkamp ve van der Linden, 2008) olarak tanımlanmaktadır. Bu bilgiler doğrultusunda madde havuzunun verimli/işlevsel kullanımına ilişkin, $.00 < r_i \leq .20$ madde kullanım sıklığı aralığındaki değerler (*EK-4 Grafik 12; EK-5 Grafik 17* ve *EK-6 Grafik 22*) incelendiğinde, test uzunluğu ve örneklem büyüklüğü arttıkça tek biçimli dağılım koşullarında büyük farklılıklar bulunmamakta iken diğer tüm ölçme koşullarında kullanılan madde sayısının arttığı bulgusuna ulaşılmıştır.

Aşamalı Düşürme Stratejisi kullanıldığında, hedeflenen kullanım sıklık oranından daha yüksek ($r_i > .20$) madde kullanım sıklığı aralığında kullanılan madde sayısı bulunmamaktadır. Sympson Hetter Stratejisi kullanıldığında hedeflenen kullanım sıklık oranından daha yüksek madde kullanım sıklık oranlarının dağılımı (*EK-5 Grafik 18*) incelendiğinde, uzun test-küçük örneklem-sola çarpık dağılımda hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı fazla olduğu için ilgili maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin en fazla ($Ky = -1.86$) olduğu görülmüştür. Sympson Hetter Stratejisi kullanıldığı - tek biçimli ve uzun testlerde hem küçük örneklem hem de geniş örneklem koşullarında hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı az olduğu için ilgili maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin en az ($Ky = 2.00$) olduğu belirtilebilir.

Madde kullanım sıklığı kontrol yöntemi uygulanmadığında, hedeflenen kullanım sıklık oranından ($r_i = .20$) daha yüksek madde kullanım sıklık oranlarının dağılımları (*EK-4 Grafik 13*) incelendiğinde, uzun test-küçük örneklem-sağa çarpık

dağılımda hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı fazla olduğu için ilgili maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin en fazla ($K_y = -1.09$) olduğu görülmüştür. Bunun yanı sıra, uzun test tek biçimli dağılımda hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı az olduğu için ilgili maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin en az ($K_y = 1.15$) olduğu sonucuna ulaşılmıştır.

Madde kullanım sıklığı kontrol yöntemi uygulanmadığında sağa ve sola çarpık dağılımlarda (EK-4 Grafik 14) ve Sympson Hetter Stratejisi-uzun test sağa ve sola çarpık dağılım koşullarında (EK-5 Grafik 19) hedeflenen kullanım sıklık oranından ($r_i = .20$) yüksek madde kullanım sıklıklarına ilişkin dağılımın çarpıklık katsayıları negatiftir. Dolayısıyla sağa ve sola çarpık dağılımlarda, madde kullanım sıklığı kontrol yöntemi uygulanmadığında hem kısa testlerde hem de uzun testlerde; Sympson Hetter Stratejisi kullanıldığında ise sadece uzun testlerde maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin arttığı ifade edilebilir.

BOBUT uygulama sürecinde farklı ölçme koşullarında ölçme duyarlılığı ve test güvenliğine ilişkin bulgular genel olarak değerlendirildiğinde, madde kullanım sıklığı kontrolü yapılmadığında en düşük RMSE ve sifra en yakın yanlılık değeri bulgusuna ulaşılsa da RMSE ve yanlılık değerleri bakımından madde kullanım sıklığı kontrolü yöntemleri arasında manidar farklılıklar bulunmadığı sonucuna ulaşılmıştır. Uyum katsayıları bakımından ise, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığı durumlara ilişkin uyum katsayıları arasında manidar bir farklılık bulunmamakta iken; sağa ve sola çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayılarına göre, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları lehine manidar bir farklılık bulunmuştur.

Ölçme duyarlılığına ilişkin ilgili değişkenler birlikte ele alınıp değerlendirildiğinde, farklı madde kullanım sıklığı kontrolü koşullarında ölçme duyarlılıkları arasında genel olarak büyük farklılıklar bulunmamıştır. Bu bulgu doğrultusunda, ilgili madde kullanım sıklığı kontrol yöntemleri kullanıldığında, ölçme duyarlılığının düşürülmeyeceği belirtilebilir.

BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin test güvenliğine etkisini inceleyebilmek için, madde kullanım sıklığı oranı, madde havuzundan yararlanmama, test örtüşme indeksi birlikte ele alınarak değerlendirilmektedir (Davis ve Dodd, 2005). Bu durumda, test güvenliğine ilişkin ilgili değişkenler birlikte ele alınıp değerlendirildiğinde, ilgili koşullarda Aşamalı Düşürme yöntemi seçildiğinde, test güvenliği daha iyi sağlanarak madde kullanım sıklığı dengeli dağılım göstermekte ve madde havuzu kullanımını daha verimli hale geldiği görülmüştür (Davis, 2002; Boyd, 2003). Bunların yanı sıra, Aşamalı Düşürme yöntemi seçildiğinde, bireyler için örtüşen maddelerin oranı (Chen, Ankenmann ve Spray, 1999) ve maksimum kullanım sıklık oranı düşük olduğundan madde havuzu ifşâ olmadan birçok uygulama için süreklilik sağlayacağı (Revuelta ve Ponsoda, 1998) ortaya çıkmıştır.

Madde kullanım sıklığı kontrol yöntemleriyle, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin sağlanması amaçlanmaktadır (Boyd, 2003; Boyd, Dodd ve Fitzpatrick, 2013; Davis ve Dodd, 2005; Pastor, Dodd ve Chang, 2002). Bu bilgi doğrultusunda, yapılan araştırmanın sonucunda, BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinden en yaygın kullanılan (Veldkamp, Vershoor ve Eggen, 2007) Sympson Hetter stratejisi yerine Aşamalı Düşürme Stratejisi seçilmesi durumunda, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin sağlandığı çıkarımında bulunulabilir. Bu bulgu, Davis (2002)'in belirttiği, Sympson Hetter stratejisinde her ne kadar BOBUT uygulamasından önce, hedeflenen kullanım sıklığı oranına ulaşılabilmesi için tekrarlı simülasyonlarla kullanım sıklığı kontrol parametreleri belirlense de madde örtüşme indeksi ve madde havuzunun kullanılma derecesi bakımından hedeflenen düzeyde madde kullanım sıklığı sonucuna ulaşılmadığına ilişkin ifadeyle benzerlik göstermektedir. Ayrıca madde seçme yöntemlerine ve farklı güçlük düzeyindeki madde havuzlarına göre, çeşitli madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı Boztunç-Öztürk (2014)'ün çalışmasında Sympson Hetter Stratejisi yerine Aşamalı Düşürme madde kullanım sıklığı kontrol yönteminin kullanılması durumunda test güvenliğinin arttığına ilişkin bulguyla paralellik göstermektedir.

BÖLÜM IV

4. SONUÇLAR VE ÖNERİLER

Bu bölümde, araştırmadan elde edilen bulgular doğrultusunda sonuçlara yer verilmiş ve bu sonuçlara dayalı olarak önerilerde bulunulmuştur.

4.1. Sonuçlar

Bu araştırmada, BOBUT uygulamalarında iki kategorili (1-0) puanlanan çoktan seçmeli maddelerde madde kullanım sıklığı kontrol yöntemlerinin farklı ölçme koşullarında (farklı yetenek dağılımlarında, test uzunluklarında ve örneklem büyüklüklerinde) ölçme duyarlılığına ve test güvenliğine etkisinin incelenmesi amaçlanmıştır. Bu doğrultuda, farklı ölçme koşullarında ölçme duyarlılığı ve test güvenliği indekslerine ilişkin bu araştırmadan elde edilen sonuçlara sırasıyla aşağıda yer verilmiştir.

4.1.1. Farklı Ölçme Koşullarında Ölçme Duyarlılığına İlişkin Sonuçlar

BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin ölçme duyarlılığına etkisini inceleyebilmek için, ölçme duyarlılığı arttığında, hata değerleri azalacağı için, her bir koşulda hata göstergeleri olan uyum, yanlılık ve Hataların Ortalama Karekökü (RMSE) katsayıları birlikte ele alınarak değerlendirilmiştir.

4.1.1.1. Farklı Ölçme Koşullarında Uyum Katsayılarına İlişkin Sonuçlar

- Her bir koşulda kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasında yüksek düzeyde uyum bulunduğu sonucuna ulaşılmıştır. Bunun yanı sıra diğer ölçme koşulları sabit tutulduğunda, yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça madde kullanım sıklığı kontrolü yöntemlerine ilişkin uyum katsayıları her ne kadar artsa da aralarında manidar bir farklılık bulunmadığı ortaya çıkmıştır.

- Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığı durumlara ilişkin uyum katsayıları arasında manidar bir farklılık bulunmamakta iken; sağa ve sola çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayılarına göre, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları lehine manidar bir farklılık bulunmuştur.
- Sağa çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayıları ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları arasında geniş düzeyde etki büyüklüğünün elde edildiği görülmüştür.
- Sola çarpık dağılımlarda, Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayıları ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları arasında kısa testlerde, orta düzeyde ya da orta düzeye yakın etki büyüklüğü elde edilirken; uzun testlerde ise, geniş düzeyde etki büyüklüğü elde edildiği belirlenmiştir. Bu durumda çarpık dağılımlarda madde kullanım sıklık yöntemlerinden Aşamalı Düşürme Stratejisi seçilmesi durumunda diğer koşullara göre kestirilen yetenek düzeyi ile gerçek yetenek düzeyi arasındaki uyum katsayısının önemli derecede azalacağı ve hatanın artacağı ifade edilebilir.

4.1.1.2. Farklı Ölçme Koşullarında Yanlılık Değerlerine İlişkin Sonuçlar

- Her koşulda yanlılık değerleri sıfıra yakın olduğu için, her bir koşulda kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasındaki ortalama farklılığın düşük olduğu saptanmıştır.
- Her bir ölçme koşulunda yanlılık değerleri arasında manidar farklılığın bulunmadığı sonucuna ulaşılmıştır.
- Diğer ölçme koşulları sabit tutulduğunda, genel olarak, örneklem büyüklüğü ve test uzunluğu arttıkça, yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme

Stratejisi kullanılmadığında gerçek yetenek puanı ile kestirilen yetenek puanı arasındaki ortalama farklılığın düştüğü görülmüştür.

4.1.1.3. Farklı Ölçme Koşullarında RMSE Değerlerine İlişkin Sonuçlar

- RMSE değerleri incelendiğinde, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin gerçek yetenek puanı ile kestirilen yetenek puanı arasındaki mutlak farklılıklar bakımından manidar bir değişim bulunmadığı ifade edilebilir.
- Farklı ölçme koşullarında RMSE değerleri arasında manidar farklılıklar bulunmamakla birlikte, diğer ölçme koşulları sabit tutulduğunda, genel olarak, örneklem büyüklüğü ve test uzunluğu arttıkça, yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça ve madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanılmadığında gerçek yetenek puanı ile kestirilen yetenek puanı arasındaki mutlak farklılığın azaldığı söylenebilir.

4.1.2. Farklı Ölçme Koşullarında Test Güvenliğine İlişkin Sonuçlar

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin test güvenliğine etkisini inceleyebilmek için ise madde kullanım sıklığı oranlarının standart sapma değerleri, maksimum madde kullanım sıklığı oranları, madde havuzundan yararlanmama, test örtüşme indeksi hesaplanmıştır.

4.1.2.1. Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Oranlarının Standart Sapma Değerlerine İlişkin Sonuçlar

- Farklı ölçme koşullarında madde kullanım sıklığı oranlarının standart sapma değerlerine ilişkin elde edilen bulgular genel olarak değerlendirildiğinde, SH ile YOK arasında genelde manidar farklılık bulunmamakta iken; SH ile AD arasında ve YOK ile AD arasında manidar farklılık bulunmuştur.
- Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan madde kullanım sıklığı oranlarının standart sapma değerleri, diğer iki yöntemle göre manidar bir şekilde daha küçük olduğu görülmüştür. Dolayısıyla Aşamalı Düşürme Stratejisi kullanıldığında madde kullanım sıklığı oranlarının diğer stratejilere göre manidar bir şekilde daha homojen dağılım gösterdiğinden, madde

kullanım sıklığı dengeli dağılım gösterdiği ve madde havuzunun daha verimli kullanıldığı ortaya çıkmıştır.

4.1.2.2. Farklı Ölçme Koşullarında Maksimum Madde Kullanım Sıklığı Oranlarına İlişkin Sonuçlar

- Maksimum madde kullanım sıklığı oranları bakımından incelendiğinde, YOK ile SH arasında; SH ile AD arasında ve YOK ile AD arasında manidar farklılık bulunmuştur.
- Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan maksimum madde kullanım sıklığı oranlarına göre, en düşükten en yükseğe doğru Aşamalı Düşürme Stratejisi, Sympson Hetter Stratejisi, madde kullanım sıklığı kontrolünün yapılmadığı koşul olarak sıralandığı görülmüştür.
- Maddelerin kullanım sıklığı oranı arttıkça maddelerin ifşâ olma ihtimalleri arttığı için ilgili koşullarda Aşamalı Düşürme Stratejisi kullanıldığında, maddelerin ifşâ olarak psikometrik özelliklerini kaybetme riskinin diğer koşullara göre manidar bir şekilde azaldığı ortaya çıkmıştır.

4.1.2.3. Farklı Ölçme Koşullarında Madde Havuzundan Yararlanmama İndekslerine İlişkin Sonuçlar

- Madde havuzundan yararlanmama indeksleri incelendiğinde, farklı ölçme koşullarında madde havuzundan yararlanmama oranları arasında manidar farklılığın bulunmadığı sonucuna ulaşılmıştır.
- Diğer ölçme koşulları sabit tutulduğunda, örneklem büyüklüğü ve test uzunluğu arttıkça, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde havuzundan yararlanmama oranlarının azaldığı görülmüştür.
- Örneklem büyüklüğü ve test uzunluğu arttıkça, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde havuzunun kullanılma derecesinin arttığı görülmektedir.

4.1.2.4. Farklı Ölçme Koşullarında Madde Kullanım Sıklık Oranlarının Çarpıklık Katsayısına (χ^2 değerlerine) İlişkin Sonuçlar

- Madde kullanım sıklık oranlarının çarpıklık katsayısı (χ^2 değerleri) incelendiğinde, farklı ölçme koşullarında hesaplanan χ^2 değerlerine göre, en düşükten en yükseğe doğru Aşamalı Düşürme Stratejisi, Sympson Hetter Stratejisi, madde kullanım sıklığı kontrol yöntemi uygulanmadığında hesaplanan χ^2 değerleri olarak sıralandığı ortaya çıkmıştır.
- En ideal madde kullanım sıklık yöntemi olduğuna karar verilebilmesi için kullanılan karşılaştırma ölçüsü F istatistiklerine ilişkin her bir koşulda $F_{\chi^2_{AD}, \chi^2_{SH}} < 1$ ve $F_{\chi^2_{YOK}, \chi^2_{SH}} > 1$ sonuçlarına ulaşılmıştır.
- Her bir koşulda, madde kullanım sıklığı oranlarının genel olarak dengelenmesi bakımından, madde kullanım sıklığı kontrolü yapılmamasının, Sympson Hetter Stratejisine göre daha kullanışsız bir yöntem olduğu; Aşamalı Düşürme Stratejisinin ise Sympson Hetter Stratejisine göre daha iyi bir yöntem ($F_{\chi^2_{AD}, \chi^2_{SH}} < 1$) olduğu saptanmıştır.

4.1.2.5. Farklı Ölçme Koşullarında Test Örtüşme İndekslerine İlişkin Sonuçlar

- Madde kullanım sıklığı kontrol yöntemlerine ilişkin test örtüşme indeksleri incelendiğinde, Aşamalı Düşürme Stratejisi kullanıldığında en düşük test örtüşme indeksi, kısa test - küçük örneklem - normal dağılım koşulunda elde edildiği görülmüştür.
- Madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında, en düşük test örtüşme indeksi, kısa test - küçük örneklem - tek biçimli dağılım koşulunda elde edilmiştir. Fakat tüm koşullarda da Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan test örtüşme indeksleri diğer stratejilere göre manidar bir şekilde daha küçük bulunmuştur.
- BOBUT uygulamalarında, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında, herhangi iki bireyin karşılaştıkları aynı maddelerin sayısı azaldığından hem test güvenliğinin daha az zedelendiği hem de madde havuzunun yıllarca kullanılabilir olmasına imkân sunduğu ortaya çıkmıştır.

4.1.2.6. *Farklı Ölçme Koşullarında Madde Havuzunun Kullanımına İlişkin Hedeflenen Kullanım Sıklık Oranından Daha Yüksek Madde Kullanım Sıklık Oranlarının Dağılımları*

- Farklı ölçme koşullarında madde havuzunun kullanımına ilişkin hedeflenen kullanım sıklık oranından daha yüksek madde kullanım sıklık oranlarının dağılımları incelendiğinde, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında, hedeflenen kullanım sıklık oranından daha yüksek oranda kullanılan maddenin bulunmadığı sonucuna ulaşılmıştır.
- Sympson Hetter Stratejisi kullanıldığında, hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı en fazla uzun test-küçük örneklem-sola çarpık dağılımda; en az tek biçimli ve uzun test koşullarında görülmüştür.
- Madde kullanım sıklığı kontrol yöntemi uygulanmadığında, hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı en fazla uzun test-küçük örneklem-sağa çarpık dağılımda; en az uzun test-tek biçimli dağılım koşulunda olduğu görülmüştür.
- Sağa ve sola çarpık dağılımlarda, madde kullanım sıklığı kontrol yöntemi uygulanmadığında hem kısa testlerde hem de uzun testlerde; Sympson Hetter Stratejisi kullanıldığında ise sadece uzun testlerde hedeflenen kullanım sıklık oranından yüksek madde kullanım sıklıklarına ilişkin dağılımın çarpıklık katsayıları negatif olduğundan maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin arttığı görülmüştür.

BOBUT uygulama sürecinde, madde kullanım sıklığı kontrol yöntemlerinin farklı yetenek dağılımlarında, test uzunluklarında ve örneklem büyüklüklerinde ölçme duyarlılığına ve test güvenliğine etkisi genel olarak değerlendirildiğinde, farklı madde kullanım sıklığı kontrolü koşullarında ölçme duyarlılıkları arasında genel olarak büyük farklılıklar bulunmadığı belirtilmiştir. Dolayısıyla, ilgili madde kullanım sıklığı kontrol yöntemleri kullanıldığında, ölçme duyarlılığının düşürülmeyeceği ifade edilmiştir. İlgili koşullarda Aşamalı Düşürme yöntemi seçildiğinde, test güvenliği daha iyi sağlanarak madde kullanım sıklığı dengeli dağılım gösterdiğinden madde havuzu kullanımının daha verimli hale geldiği ve madde havuzu ifşâ olmadan birçok uygulama için süreklilik sağlayabildiği ortaya çıkmaktadır. Bu doğrultuda, yapılan

araştırmanın sonucunda, BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi seçildiğinde, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin sağlanabileceği belirtilmiştir.

4.2. Öneriler

Araştırmadan elde edilen bulgular çerçevesinde bu bölümde uygulayıcılara ve araştırmacılara yönelik önerilere yer verilmiştir.

- Bu çalışma kapsamında madde kullanım sıklığı kontrol yöntemlerine ilişkin karşılaştırmalarda, kullanım sıklığının kontrol edilmediği durum referans olarak alınmıştır. BOBUT uygulamalarında, madde kullanım sıklığı kontrol yöntemlerini kullanmamak yerine, test güvenliğini ve ölçme duyarlılığını olumsuz yönde etkileyecek durumların üstesinden gelebilmek için BOBUT algoritmasında kısıtlamalar yaparak, madde kullanım sıklığı kontrol yöntemlerinin kullanılması önerilebilir.
- BOBUT uygulamalarında, madde kullanım sıklığı kontrol yöntemlerinden BOBUT uygulamasından önce, tekrarlı simülasyonlarla kullanım sıklığı kontrol parametrelerinin belirlendiği Sympson Hetter stratejisi yerine, BOBUT uygulanma sürecinde kullanım sıklığı kontrol parametrelerinin eş zamanlı olarak belirlendiği Aşamalı Düşürme Stratejisi tercih edilmelidir. Çünkü, madde kullanım sıklığı kontrol yöntemlerinden Sympson Hetter stratejisi yerine, Aşamalı Düşürme Stratejisi tercih edildiğinde, farklı ölçme koşullarında ölçme duyarlılığını düşürmeden madde havuzu daha dengeli kullanılarak test güvenliği daha iyi sağlanmaktadır.
- Bu araştırmada, BOBUT uygulamalarında farklı ölçme koşullarında, iki kategorili puanlanan çoktan seçmeli maddelerde madde kullanım sıklığı kontrol yöntemleri karşılaştırılmıştır. Çok kategorili puanlanan maddelerde madde kullanım sıklığı kontrol yöntemlerinin farklı yetenek dağılımlarında, test uzunluklarında ve örneklem büyüklüklerinde ölçme duyarlılığına ve test güvenliğine etkisi üzerine bir araştırma yapılabilir.
- Örneklem yanlılığı oluşmaması için araştırma kapsamında, her bir koşulda 100 tekrar kullanılmıştır. Tekrar sayılarının madde kullanım sıklığı kontrol

yöntemlerinin ölçme duyarlılığına ve test güvenliğine etkisini incelemek amacıyla yetenek dağılımları, test uzunlukları ve örneklem büyüklükleri sabit tutulup tekrar sayıları üzerinde farklı koşullar oluşturarak bir araştırma yapılabilir.

- Bu araştırma, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi ve Sympson Hetter Stratejisiyle sınırlandırılmıştır. Maddelerin kullanım sıklık oranını azaltmaya yönelik bir önsel dağılım kullanılarak, maddelerin farklı testlerde uygulandığı dönüşümlü madde havuzu yöntemi ile BOBUT uygulanma sürecinde kullanım sıklığı kontrol parametrelerinin eş zamanlı olarak belirlendiği Aşamalı Düşürme Stratejisi kullanılarak, farklı yetenek dağılımlarında, test uzunluklarında ve örneklem büyüklüklerinde ölçme duyarlılığı ve test güvenliklerinin nasıl etkilendiği karşılaştırılabilir.
- Farklı ölçme koşulları kapsamında, yetenek dağılımları, test uzunlukları ve örneklem büyüklüklerinin yanı sıra madde havuzu genişliklerini de tabakalandırarak madde kullanım sıklığı kontrol yöntemleri karşılaştırılabilir.
- Araştırmada simülatif BOBUT uygulaması için CATSim (Weiss ve Guyer, 2012) yazılım programında ve catR (Magis ve Barrada, 2017); catIRT (NyDick, 2014) paketlerinde madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi'ne yer verilmediğinden SimulCAT programı kullanılmıştır. Aşamalı Düşürme Stratejisini de kapsayan farklı paketler geliştirilip simülatif BOBUT uygulaması yapılarak paketlerden/programlardan elde edilen yetenek kestirimlerine ilişkin ölçme duyarlılıkları ve test güvenlikleri karşılaştırılabilir.
- Bu çalışmanın sınırlılığı olarak, BOBUT uygulaması için, başlatma kuralı $-.50 < b < .50$, yetenek kestirim yöntemi BSD, sonlandırma kuralı olarak sabit test uzunluğu stratejisi kullanılmıştır. Farklı stratejiler kullanılarak yapılan simülatif BOBUT uygulamalarında farklı yetenek dağılımlarında, test uzunluklarında ve örneklem büyüklüklerinde madde kullanım sıklığı kontrol yöntemlerine ilişkin ölçme duyarlılıkları ve test güvenlikleri karşılaştırılabilir.

KAYNAKÇA

- Agresti A. and Coull B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, 52(2), 119-126.
- Alpar, R. (2011). *Uygulamalı çok değişkenli istatistiksel yöntemler*. Ankara: Detay Yayıncılık.
- Ariel, A., Veldkamp, B. P., and van der Linden, W. J. (2004). Constructing rotating item pools for constrained adaptive testing. *Journal of Educational Measurement*, 41(4), 345–359.
- Arıcıgil Çılan, Ç. (2009). *Sosyal bilimlerde kategorik verilerle ilişki analizi*. Ankara: Pegem Akademi.
- Aytuğ Koşan, A. M. (2013). *Tıp eğitiminde gelişim sınavı soru bankası oluşturulması ve benzetim verileri ile bilgisayar uyarlamalı test uygulaması* (Doktora tezi). Ankara Üniversitesi, Ankara.
- Bailey, K. D. (1994). *Methods of social research*. New York: The Free Press.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi.
- Boyd, M. A. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems* (Doctoral dissertation). The University of Texas, Austin.
- Boyd, A. M., Dodd, B., and Fitzpatrick, S. (2013). A comparison of exposure control procedures in CAT systems based on different measurement models for testlets. *Applied Measurement in Education*, 26 (2), 113– 135.
- Boztunç - Öztürk, N. (2014). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı kontrol yöntemlerinin incelenmesi* (Yayınlanmış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, N.J.: Guilford Press.
- Bulut, O., ve Kan, A. (2012) Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eurasian Journal of Educational Research*, 49, 61-80.
- Burt, W. M., Kim. S. J., Davis, L. L., and Dodd, B. G. (2003, April). *A comparison of item exposure control procedures using a CAT system based on the generalized partial credit model*. Paper presented at the Annual meeting of the American Educational Research Association. Chicago, IL.
- Byrne, B. M. (1998). *Structural equation modeling with lirs, prelis, and simplis: Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum Associates.
- Chang, S. W., and Twu, B. Y. (1998). *A comparative study of item exposure control methods in CAT*. Research Report 98- 3. Iowa City, IA: ACT.
- Chang, H.-H., and Ying, Z. (1999). A-stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, 23, 211-222.

- Chang, H.-H., Qian, J., and Ying, Z. (2001). a-stratified multistage computerized adaptive testing with b blocking. *Applied Psychological Measurement*, 25, 333-341.
- Chen S. Y., Ankenmann, R. D. and Spay, J. A. (1999). *Exploring the relationship between item exposure rate and test overlap rate in computerised adaptive testing*. (ACT Research Report Series), ACT-RR-99-5.
- Clauser, B. E., and Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice*, 17, 31-44.
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.
- Cook, L. L., and Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational measurement: Issues and Practice*. 10 (3), 37-45.
- Cömert, M. (2008). *Bireye uyarlanmış bilgisayar destekli ölçme ve değerlendirme yazılımı geliştirilmesi* (Yayınlanmamış Yüksek Lisans Tezi). Bahçeşehir Üniversitesi, İstanbul.
- Crocker, L. and Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- Çetin, E. (2009). *Dikey ölçeklemede klasik test ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması* (Doktora tezi). Hacettepe Üniversitesi, Ankara.
- Çıkrıkçı-Demirtaşlı, N. (1999). Psikometride yeni ufuklar: bilgisayar ortamında bireye uyarlanmış test. *Türk Psikoloji Bülteni*, 5 (13), 31-36.
- Davey, T., and Parshall, C.G. (1995, April). *New algorithms for item selection and exposure control with computerized adaptive testing*. Paper presented at the annual meeting of the American Educational Research Association (AERA), San Francisco, CA.
- Davis, L. L. (2002). *Strategies for controlling item exposure in computerized adaptive testing with polytomously scored items* (Doctoral dissertation). The University of Texas, Austin.
- Davis, L. L. (2004). Strategies for controlling item exposure in computerized adaptive testing with the generalized partial credit model. *Applied Psychological Measurement*, 28 (3), 165–185.
- Davis, L., and Dodd, B. (2003). Item exposure constraints for testlets in the verbal reasoning section of the MCAT. *Applied Psychological Measurement*, 27 (3), 335–356.
- Davis, L. L., and Dodd, B. G. (2005). *Strategies for controlling item exposure in computerized adaptive testing with partial credit model* (PEM Research Report No. 05-01). Austin, TX: Pearson Educational Measurement.
- Davis, J. P., Eisenhardt, K. M., and Bingham, C. B. (2007) Developing theory through simulation methods. *Academy of Management Review*, 32(2), 480-499.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guildford Press.

- DeMars, C. E. (2006). Application of the Bi-Factor multidimensional item response theory model to Testlet-Based tests. *Journal of educational measurement*, 43 (2), 145–168.
- Eggen, T. J. H. M. (2001). *Overexposure and underexposure of items in computerized adaptive testing*. Arnhem, The Netherlands: CITO.
- Eignor, D. R., Stocking, M. L., Way, W. D., and Steffen, M. (1993). *Case studies in computer adaptive test design through simulation* (RR-93-56). Princeton, NJ: Educational Testing Service.
- Embretson, S. E., and Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- EPODDER. (2017). *Temel eğitimden ortaöğretime geçiş konusundaki basın bildirisi*. Online <http://www.epodder.org/duyuru> adresinden 6.10.2017 tarihinde alınmıştır.
- Eroğlu, M. G. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması* (Doktora tezi). Hacettepe Üniversitesi, Ankara.
- Evans, J. J. (2010). *Comparability of examinee proficiency scores on Computer Adaptive Tests using real and simulated data* (Doctoral dissertation). The State University of New Jersey.
- French, B. and Thompson T. (2003). *The evaluation of exposure control procedures for an operational CAT*. The annual meeting of the American Educational Research Association (AERA), Chicago, IL.
- Georgiadou, E., Triantafillou, E., and Economide, A. A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *The Journal of Technology, Learning, and Assessment*, 5 (8), 3-37.
- Glas, G. A. W. (2002). Item calibration and parameter drift. W.J. van der Linden, ve G. A. W. Glas (Ed.). *Computerized adaptive testing: Theory and practice* (ss.183 - 199). New York: Kluwer Academic Publishers.
- Goldman, S. H., and Raju, N. S. (1986). Recover of one and two parameter logistic item parameters: An empirical study. *Educational and Psychological Measurement*, 24, 11-21.
- Gökçe, S. (2012). *Comparison of linear and adaptive versions of the Turkish pupil monitoring system (pms) mathematics assessment*. (Doctoral dissertation). Middle East Technical University, Ankara.
- Green, S. B., and Salkind, N. J. (2007). *Using SPSS for windows and macintosh analyzing and understanding data*. New Jersey.
- Grubbs, F. E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments, *Technometrics*, 15, 53-66.
- Gu, L. and Reckase, M.D. (2007). Designing optimal item pools for computerized adaptive tests with Sympton-Hetter exposure control. In D. J. Weiss (Ed.), *Proceedings of the 2007 GMAC Conference on Computerize Adaptive Testing*. Retrieved [28.06.2016] from www.psych.umn.edu/psylabs/CATCentral/

- Guyer, R. and Thompson, N. (2011). *Item response theory parameter recovery using Xcalibre 4.1*. St. Paul MN: Assessment Systems Corporation.
- Guyer, R., and Thompson, N.A. (2012). *User's manual for Xcalibre item response theory calibration software, version 4.1.6*. St. Paul MN: Assessment Systems Corporation.
- Hair, J., Anderson, R., Tatham, R. ve Black, W. (1998). *Multivariate data analysis*. New Jersey: Prentice Hall, Upper Saddle River.
- Hambleton, R. K., and Jones, R. W. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12, 38–47.
- Hambleton, R. K., and Swaminathan, H. (1985). *Item response teory: principles and applications*. USA: Kluwer Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications Inc.
- Han, K. T. (2009). A gradual maximum information ratio approach to item selection in computerized adaptive testing. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Retrieved [25.11.2015] from [www.psych.umn.edu/psylabs/ CATCentral/](http://www.psych.umn.edu/psylabs/CATCentral/)
- Han, K. T. (2011). *User's manual: SimulCAT: Windows software for simulating computerized adaptive test administration*. Graduate Management Admission Council.
- Han, K. T. (2012). An efficiency balanced information criterion for item selection in computerized adaptive testing. *Journal of Educational Measurement*, 49 (3), 225– 246.
- Harwell, M. R., and Janosky, J. E. (1991). An empirical study of the effects of small datasets and varying prior variances on item parameter estimation in BILOG. *Applied Psychological Measurement*, 15(3), 279-291.
- Harwell, M., Stone, C. A., Hsu, T., and Kirisci, L. (1996). Monte carlo studies in item response theory. *Aplied Psychological Measurement*, 20 (2), 101-125.
- Hooper, D., Coughlan, J., Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronical Journal of Business Research Methods*. 6(1), 53-60
- Howell, C. D. (2010). *Statistical methods for psychology*. USA: Cengage Wadsworth.
- Huang, H. Y., Chen, P. H. and Wang, W. C. (2012). Computerized adaptive testing using a class of high-order item response theory models. *Applied Psychological Measurement*, 36 (8), 689-706.
- Hulin, C. L., Lissak, R. L., and Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: a monte carlo study. *Applied Psychological Measurement*, 6(3), 249–260.
- IACAT. (2017). *Some current issues in CAT*. Online <http://iacat.org/node/491> adresinden 18.01.2017 tarihinde alınmıştır.

- İşeri, A. I. (2002). *Assessment of students' mathematics achievement through computer adaptive testing procedures* (Unpublished doctoral dissertation). Middle East Technical University, Ankara.
- Jackson, J. L., Dezee, K., Douglas, K., ve Shimeall, W. (2005). *Introduction to structural equation modeling (path analysis)* [Electronic Version]. Precourse PA08. Society of General Internal Medicine (SGIM), Washington, D.C.
- Jöreskog, K. and Sörbom, D. (1993), *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software.
- Kalaycı, Ş. (2010). *SPSS uygulamalı çok değişkenli istatistik teknikleri*. Ankara: Asil Yayın Dağıtım.
- Kalender, İ. (2009). Başarı ve yetenek kestirimlerinde yeni bir yaklaşım: Bilgisayar ortamında bireyselleştirilmiş testler (Computerized adaptive tests-CAT). *CITO Eğitim Kuram ve Uygulama*, (5), 39-48.
- Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability* (Yayınlanmamış doktora tezi). Middle East Technical University, Ankara.
- Kaptan, F. (1993). *Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Kaskatı, O. T. (2011). *Rasch modelleri kullanarak romatoid artirit hastaları özürüllük değerlendirimi için bilgisayar uyarlamalı test yöntemi geliştirilmesi* (Yayınlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Kelecioğlu, H. (2001). Örtük özellikler teorisindeki b ve a parametreleri ile klasik test teorisindeki p ve r istatistikleri arasındaki ilişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, (20), 104-110.
- Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. United States of America: Sage Publications.
- Kezer, F. (2013). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması* (Yayınlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Khatree, R., and Naik, D. N., (2000). *Applied multivariate statistics with SAS software*. SAS Institute Inc., Cary, NC.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Publications.
- Köse, İ. A. (2010). *Madde tepki kuramına dayalı tek boyutlu ve çok boyutlu modellerin test uzunluğu ve örneklem büyüklüğü açısından karşılaştırılması* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi. Ankara.
- Lee, H. Y., and Dodd, B. G. (2012). Comparison of exposure controls, item pool characteristic, and population distributions for cat using the partial credit model. *Educational and Psychological Measurement*, 72 (1), 159- 175.
- Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28(4), 989–1020.

- Leroux, A. J., Lopez, M., Hembry, I., and Dodd, B. G. (2013). A comparison of exposure control procedures in CATs using the 3PL model. *Educational and Psychological Measurement*, 73, 857-874.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lord, F., and Stocking, M. (1988). Item response theory. J. P. Keeves (Ed.). *Educational research, methodology, and measurement: An international handbook*. New York: Pergamon Press.
- Lunz, M. E., and Stahl, J. A. (1998). *Patterns of item exposure using a randomized CAT algorithm*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Magis, D. and Barrada, J. R. (2017). Computerized Adaptive Testing with R: Recent Updates of the Package catR. *Journal of Statistical Software, Code Snippets*, 76(1), 1-18.
- Magis, D., and Raîche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software*, 48 (8), 1-31.
- Mazmanoğlu, A. (2016). *Herkes için temel istatistik yöntemleri ve uygulamaları*. Ankara: Nobel Yayıncılık.
- McDonald, P. L. (2002). *Computer adaptive test for measuring personality factors using item response theory* (Unpublished doctoral dissertation). The University Western of Ontario, London.
- McLachlan, G. J. (1999). Mahalanobis Distance, *RESONANCE*, 20-26. Online <http://www.ias.ac.in/article/fulltext/reso/004/06/0020-0026> adresinden 12.01.2017 tarihinde alınmıştır.
- Nydyck S. W. (2014). *Package 'catIrt'*. Online <https://cran.r-project.org/web/packages/catIrt/catIrt.pdf> adresinden 23.02.2017 tarihinde alınmıştır.
- Orlando, M., and Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64.
- Özbaşı, D. (2014). *Bilgisayar okuryazarlığı testinin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliğine ilişkin bir araştırma* (Doktora tezi). Ankara Üniversitesi, Ankara.
- Öztuna, D. (2008). *Kas-iskelet sistemi sorunlarının özürllülük değerlendirmesinde bilgisayar uyarlamalı test yönteminin uygulanması* (Yayınlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Maidenhead: Open University Press/McGraw-Hill
- Pastor, D. A., Dodd, B. G., and Chang, H. H. (2002). A comparison of item selection techniques and exposure control mechanisms in CATs using the generalized partial credit model. *Applied Psychological Measurement*, 26 (2), 147-163.
- Patsula, L. N., and Gessaroli M. E. (1995, April). *A comparison of item parameter estimates and ICCs produced with TESTGRAF and BILOG under different*

- test lengths and sample sizes*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Revuelta, J., and Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35 (4), 311-327.
- Ross, S. M. (2013). *Simulation*. USA:Elsevier
- Rudman, H. C. (1987). The future of testing is now. *Educational Measurement: Issues and Practice*, 4, 5-11.
- Scullard, M. G. (2007). *Application of item response theory based computerized adaptive testing to the strong interest inventory* (Unpublished Doctoral Dissertation). University of Minnesota, USA.
- Segall, D. O. (2004). Computerized adaptive testing. Kempf-Leanard (Eds.). *The encyclopedia of social measurement*, (ss. 429 – 438). San Diego, CA: Academic Press.
- Smits, N., Cuijpers, P., and Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry Research*, 188, 145-155.
- Speron, E. (2009). *A comparison of metric linking procedures in item response theory*. (Unpublished doctoral dissertation) Chicago, Illinois: Illinois Institute of Technology.
- Stocking, M. L., & Lewis, C. (1995). *A new method of controlling item exposure in computerized adaptive testing* (Research Report 95-25). Princeton, NJ: Educational Testing Service.
- Stocking, M. L. and Lewis C. (2002). Methods of controlling the exposure of items in CAT. W.J. van der Linden, ve G. A. W. Glas (Ed.). *Computerized adaptive testing: Theory and practice* (ss.163 - 182). New York: Kluwer Academic Publishers.
- Sulak, S. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında kullanılan madde seçme yöntemlerinin karşılaştırılması* (Yayımlanmış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Sümer, N. (2000). *Yapısal eşitlik modelleri: temel kavramlar ve örnek uygulamalar*. Türk Psikoloji Yazıları, 3(6) 49-74.
- Şahin, A., ve Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice*, 17, 321–335.
- Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlik*. Ankara: Seçkin Yayıncılık.
- Şimşek, Ö. F. (2007). *Yapısal eşitlik modellemesine giriş: Temel ilkeler ve LISREL uygulamaları*. Ankara: Ekinoks Yayınları.
- Tabachnick, B. G., and Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Allyn and Bacon.

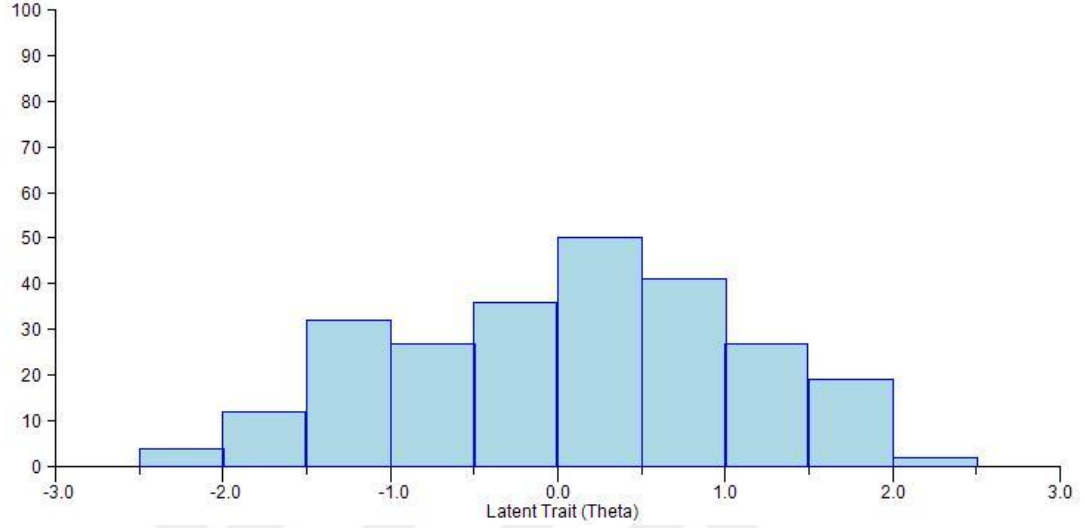
- Tang, K. L., Way, W. D., and Carey, P. A. (1993). *The effect of small calibration sample sizes on TEOFL IRT-based equating (TOEFL technical report TR-7)*. Princeton, NJ: Educational Testing Service.
- Tay, P. H. (2015). *On-the-fly assembled multistage adaptive testing* (Doctoral dissertation). The University of Texas, Urbana.
- Thissen, D., and Steinberg, L. (2009). Item response theory. R. Millsap and A. Maydeu-Olivares (Ed.) *The sage handbook of quantitative methods in psychology*. London: Sage Publications.
- Thissen, D., and Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika*, 47 (4), 397-412. doi:10.1007/BF02293705.
- Thompson, T. (2002, April). *Employing new ideas in CAT to a simulated reading test*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- Ünver, Ö., Gamgam, H., ve Altunkaymak, B. (2016). *Temel istatistik yöntemler olasılık – hipotez testleri – regresyon analizi*. Ankara: Seçkin Yayıncılık.
- Vaughn, B. K., and Wang, Q. (2010). DIF trees: using classifications trees to detect differential item functioning. *Educational and Psychological Measurement*, 70 (6), 941–952.
- Veldkamp, B. P., Verschoor, A. J. and Eggen T. J. H. M. (2007). *A multiple objective test assembly approach for exposure control problems in computerized adaptive testing*. Measurement and Research Department Reports 2007-1. Arnhem:Cito.
- Veldkamp, B. P., and van der Linden, W. J. (2008). Implementing Sympon-Hetter item-exposure control in a shadow-test approach to constrained adaptive testing. *International Journal of Testing*, 8(3), 272–289.
- Veldkamp, B.P., and van der Linden. W.J. (2010). Designing item pools for adaptive testing. van der W. J. Linden, and C.A.W. Glas (Ed.). *Elements of adaptive testing*. New York: Springer.
- Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement. Issues and Practice*, 17, 17-27.
- Way, W.D. (2005). *Practical questions in introducing computerized adaptive testing for k-12 assessments*, (PEM Research Report 05-03). Pearson Educational Measurement.
- Wang, T., and Vispoel, W. P., (1998). Properties of ability estimation methods computerized adaptive testing. *Journal of Educational Measurement*, 35 (2), 109-135.
- Wang, H. P., Kuo, B. C., Tsai, Y. H., and Liao, C. H. (2012). A Cefr-Based computerized testing system for chinese proficiency. *TOJET: The Turkish Journal of Educational Technology*, 11 (4), 1–12.
- Weiss, D. J. (1988). Adaptive testing. J. P. Keeves (Ed.). *Educational research, methodology, and measurement: An international handbook*. New York: Pergamon Press.

- Weiss, D. J. (2004). Computerized Adaptive Testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37 (2), 70-84.
- Weiss, D. J., and Guyer, R. (2012). *Manual for CATSim: Comprehensive simulation of computerized adaptive testing*. St. Paul MN: Assessment Systems Corporation.
- Weiss, D. J., and Von Minden, S. (2012). *A comparison of item parameter estimates from Xcalibre 4.1 and Bilog-MG*. St. Paul MN: Assessment Systems Corporation.
- Wetzel, E., Hell, B., and Passler, K. (2011). Comparison of different test construction strategies in the development of a gender fair interest inventory using verbs. *Journal of Career Assessment*, 20 (1), 88–104.
- Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*, 52(2), 275–291. doi:10.1007/BF02294241.
- Yen, W. M. (1993). Scaling performance assessment: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30 (3), 187-213.
- Yoes, M. (1995). *An updated comparison of microcomputer-based item parameter estimation procedures used with the 3-parameter IRT model*. Saint Paul, MN: Assessment Systems Corporation.
- Zitny, P., Halama, P., Jelinek, M., and Kveton, P. (2012). Validity of cognitive ability tests—comparison of computerized adaptive testing with paper and pencil and computer-based forms of administrations. *Studia Psychologica*, 54 (3), 181–194.
- Zheng, Y., ve Chang, H-H. (2014). On-the-fly assembled multistage adaptive testing. *Applied Psychological Measurement*. 39(2), 105-118.

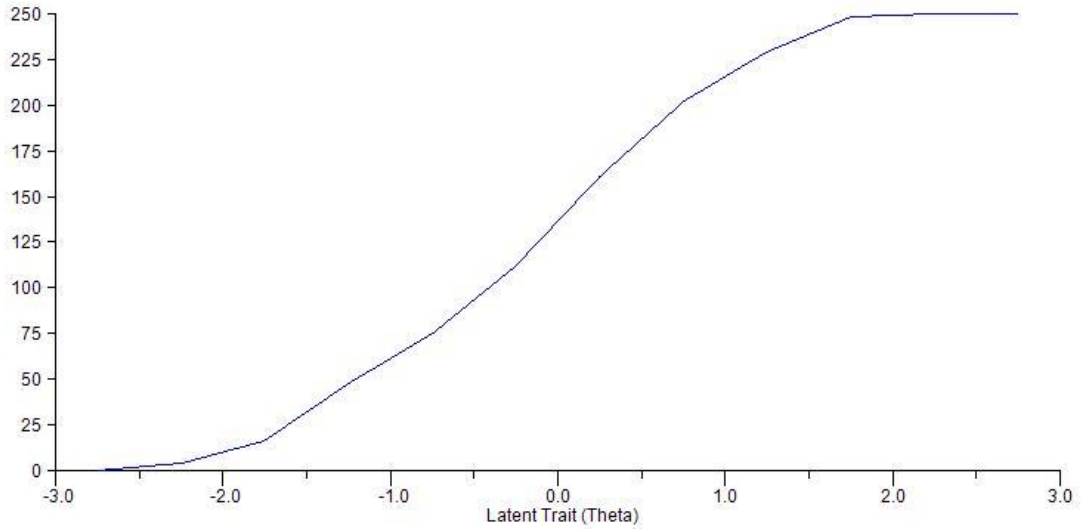
EKLER

EK-1

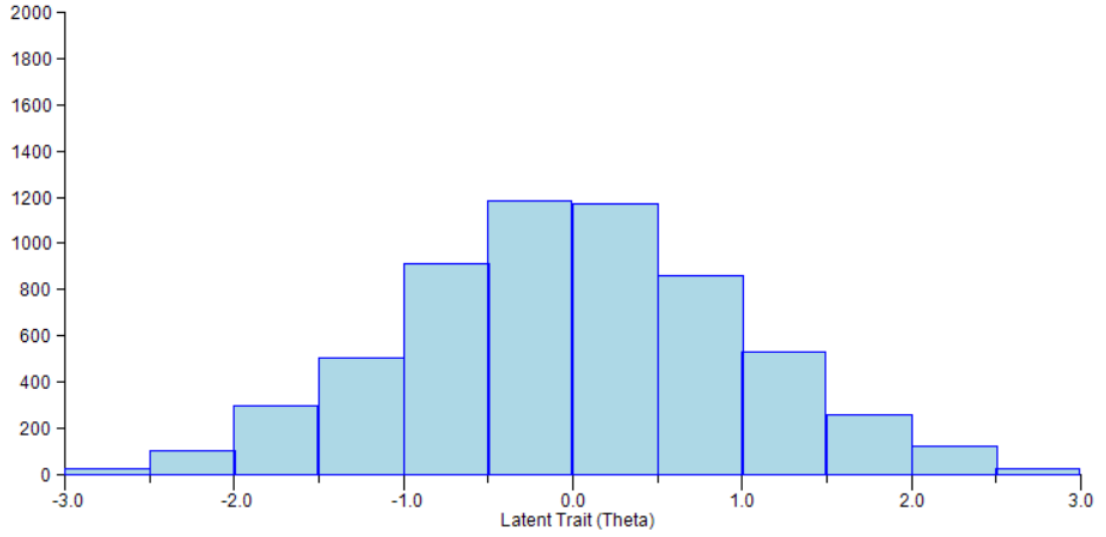
BOBUT Uygulamasına ve Sympson Hetter Örneklemine İlişkin Kümülatif ve Histogram Grafikleri



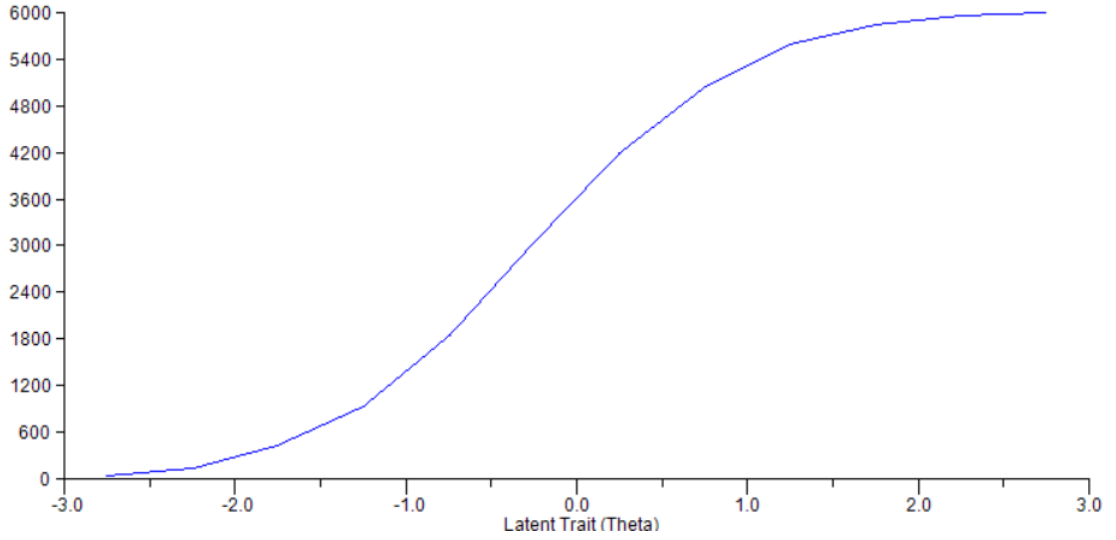
Grafik 1. BOBUT Uygulamasında Yetenekleri Normal Dağılım Gösteren Küçük Örnekleme İlişkin Histogram Grafiği



Grafik 2. BOBUT Uygulamasında Yetenekleri Normal Dağılım Gösteren Küçük Örnekleme İlişkin Kümülatif Grafik

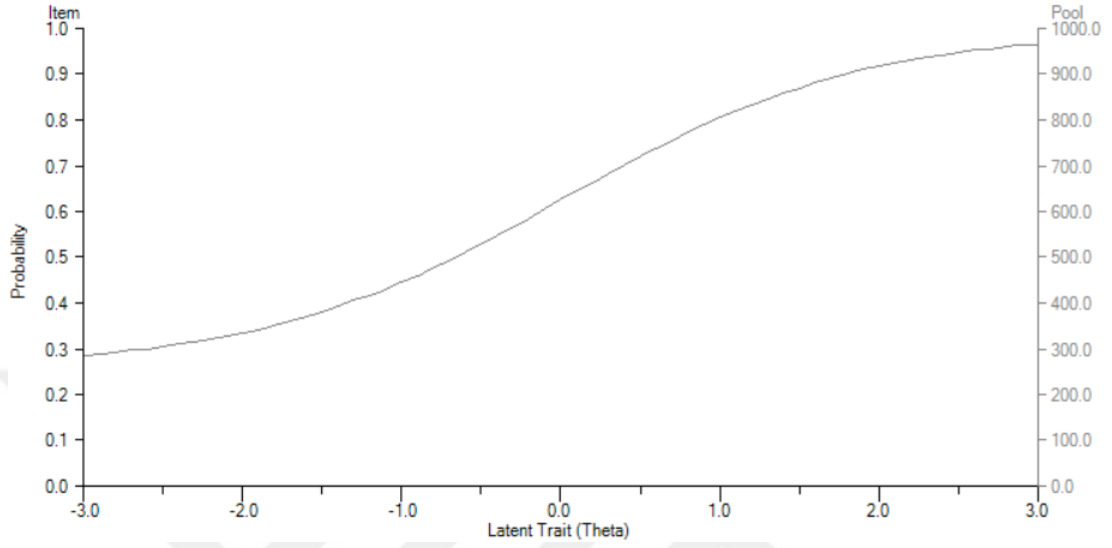


Grafik 3. Sympson Hetter Örneğine İlişkin Histogram Grafiği

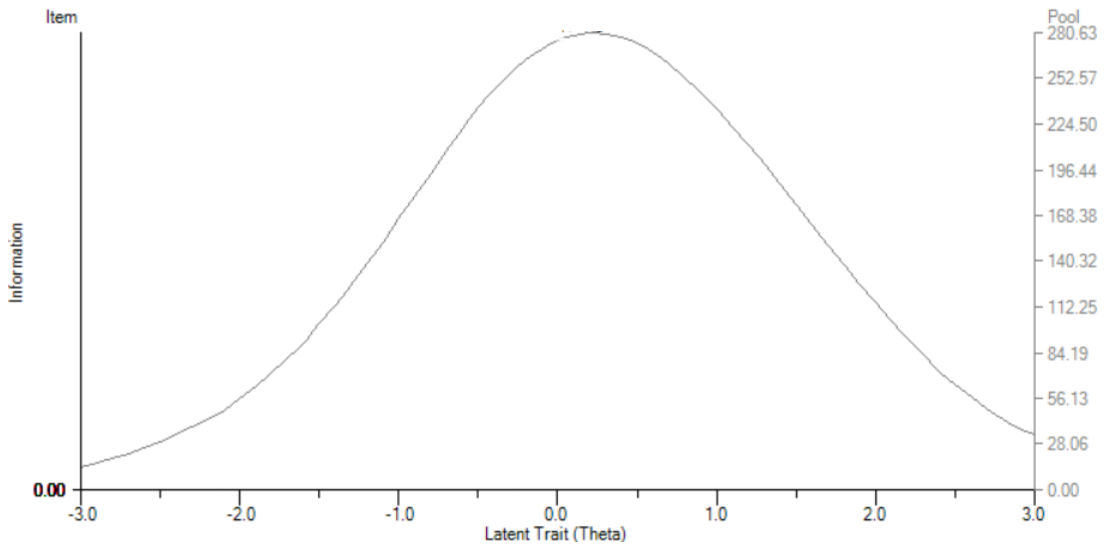


Grafik 4. Sympson Hetter Örneğine İlişkin Kümülatif Grafik

EK-2
Madde Havuzuna İlişkin Yanıt ve Bilgi Fonksiyonu Grafikleri



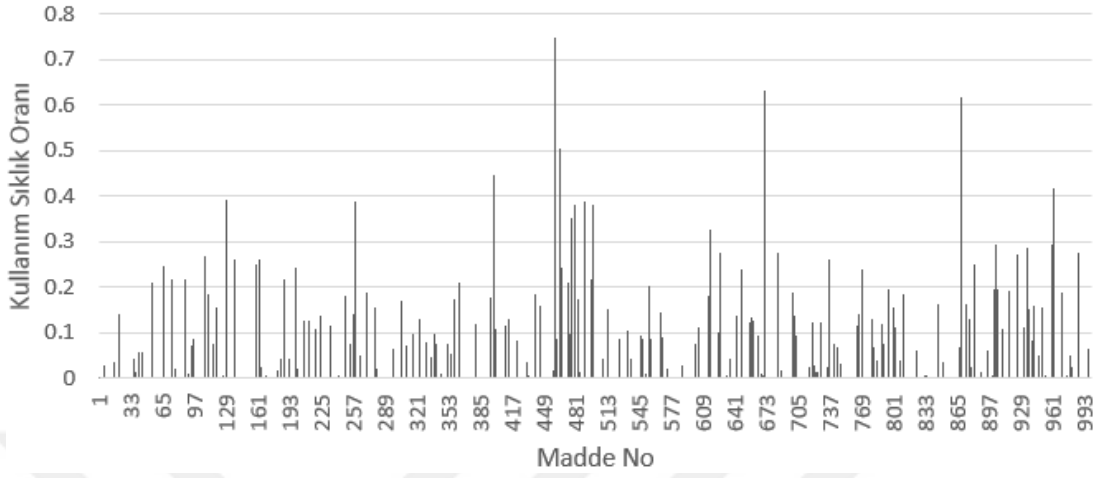
Grafik 5. *Madde Havuzuna İlişkin Yanıt Fonksiyonu Grafiği*



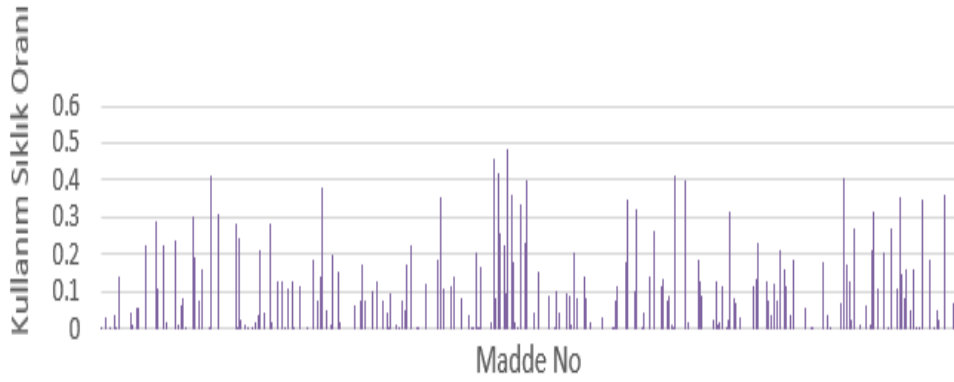
Grafik 6. *Madde Havuzuna İlişkin Bilgi Fonksiyonu Grafiği*

EK-3

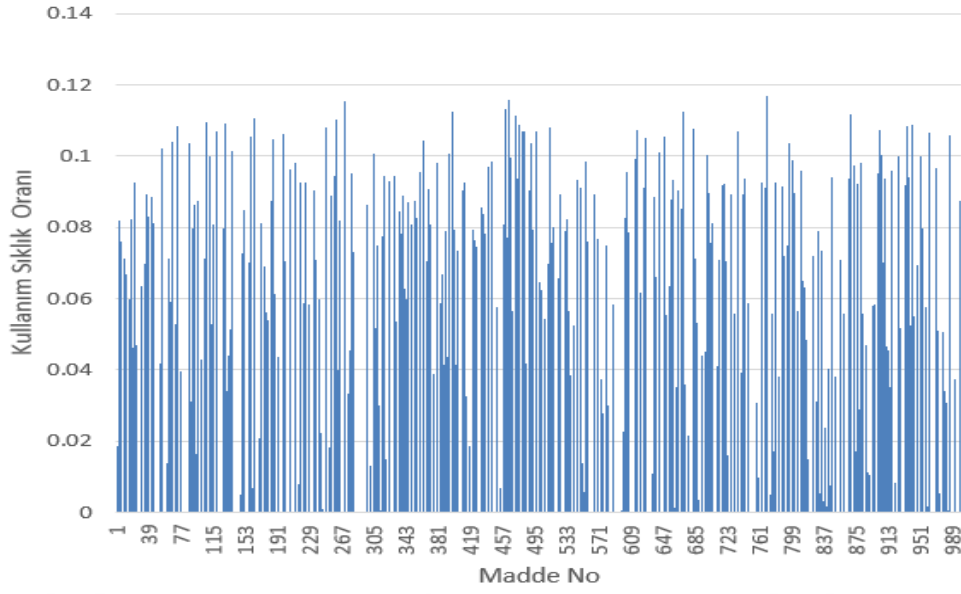
Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Oranlarının Dağılımı



Grafik 7. Küçük örneklem, normal yetenek dağılımı, kısa test ve madde kullanım sıklığı kontrolü yapılmadığında madde kullanım sıklığı oranlarının dağılımı



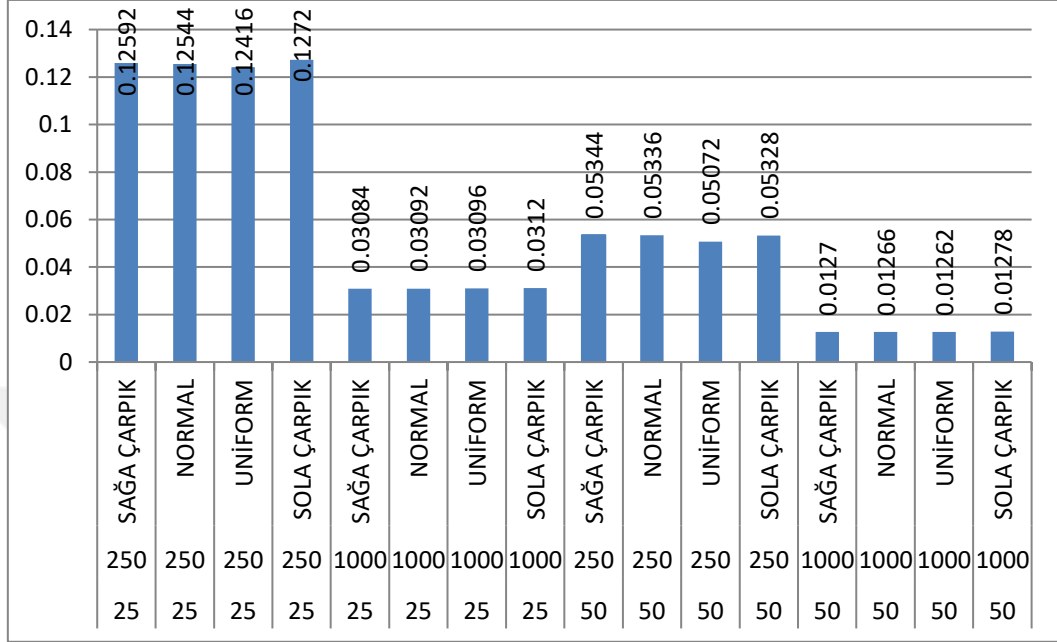
Grafik 8. Küçük örneklem, normal yetenek dağılımı, kısa test ve SH Stratejisi kullanıldığında madde kullanım sıklığı oranlarının dağılımı



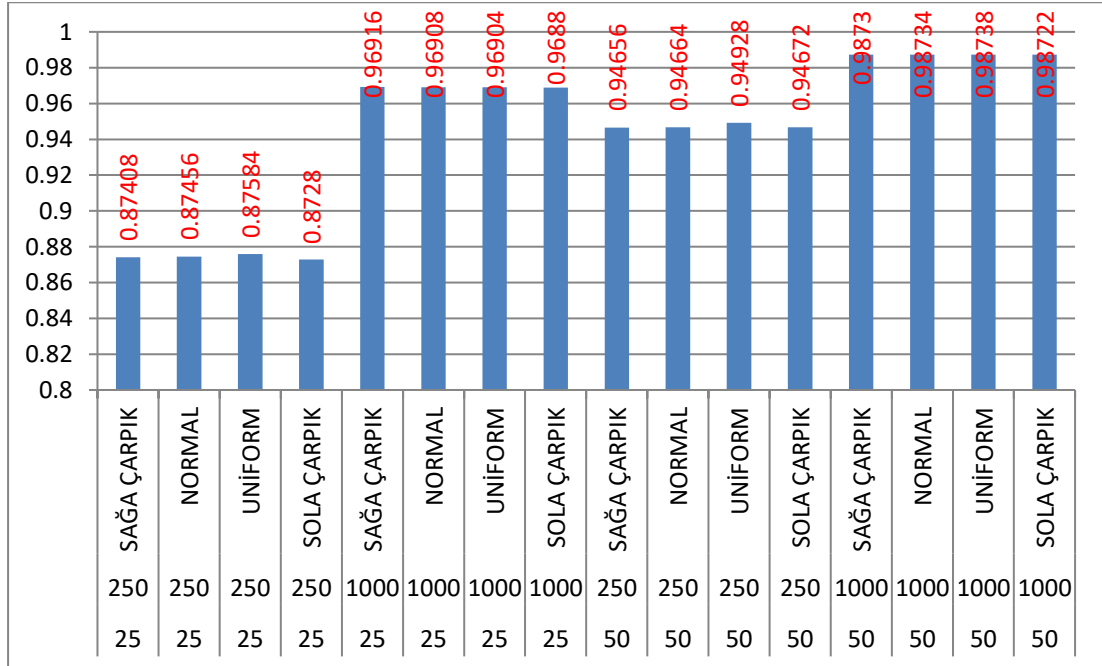
Grafik 9. *Küçük Örneklem, Normal Yetenek Dağılımı, Kısa Test ve Aşamalı Düşürme Stratejisi Kullanıldığında Madde Kullanım Sıklığı Oranlarının Dağılımı*

EK-4

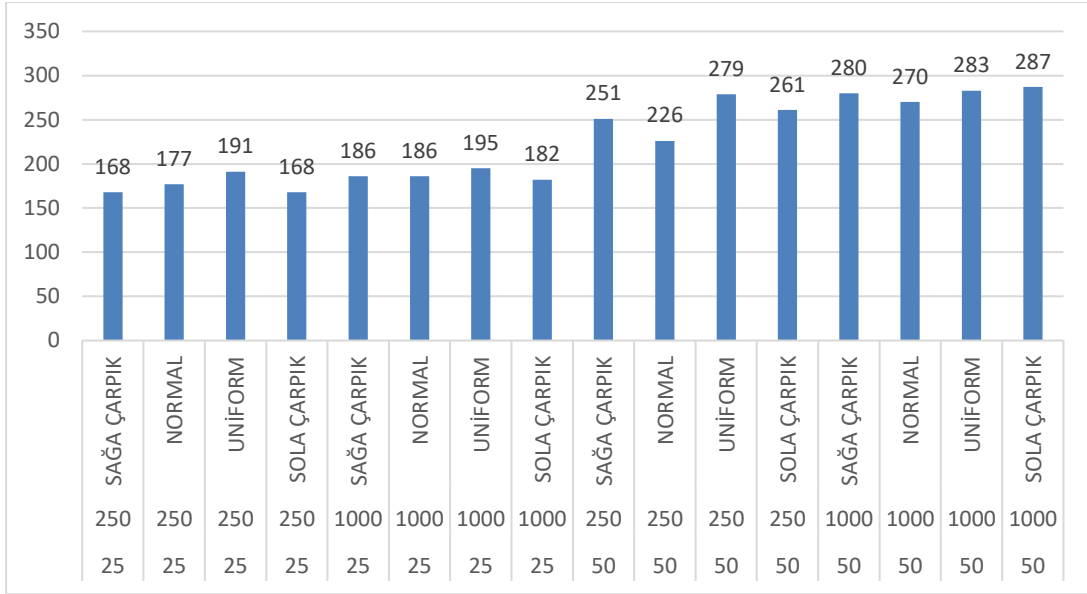
Madde Kullanım Sıklığı Kontrol Yöntemi Kullanılmadığında Madde Kullanım Sıklıklarının Dağılımı



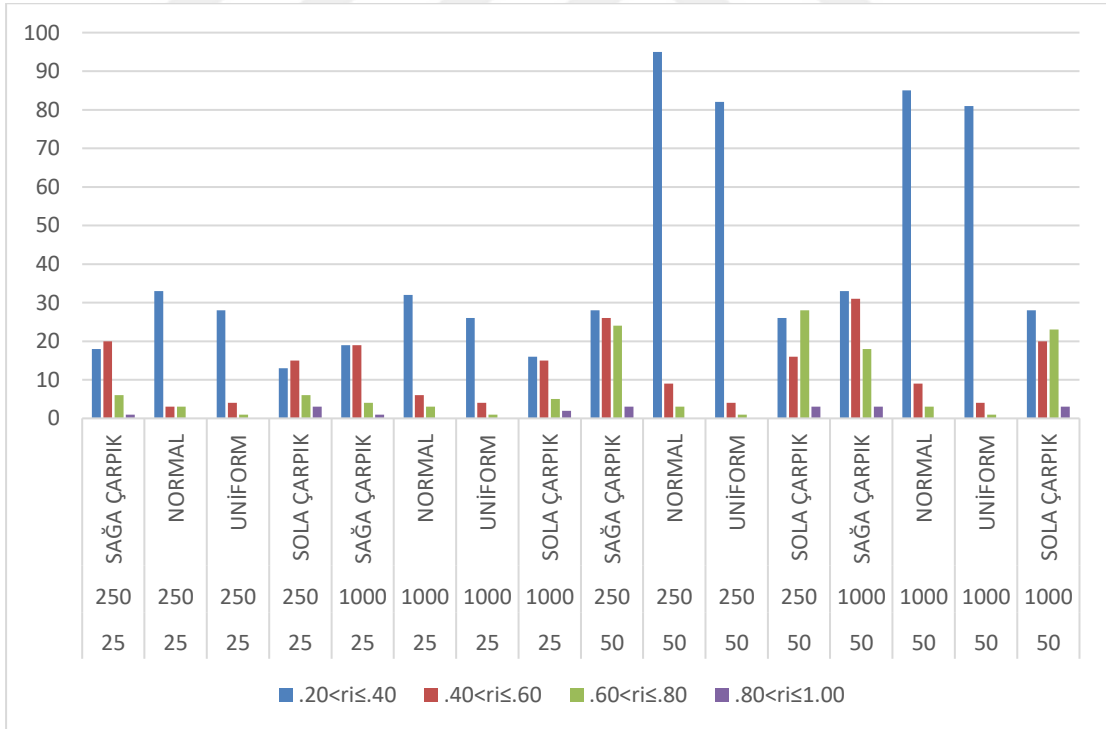
Grafik 10. Madde Havuzundan Yararlanmamaya İlişkin Dağılım



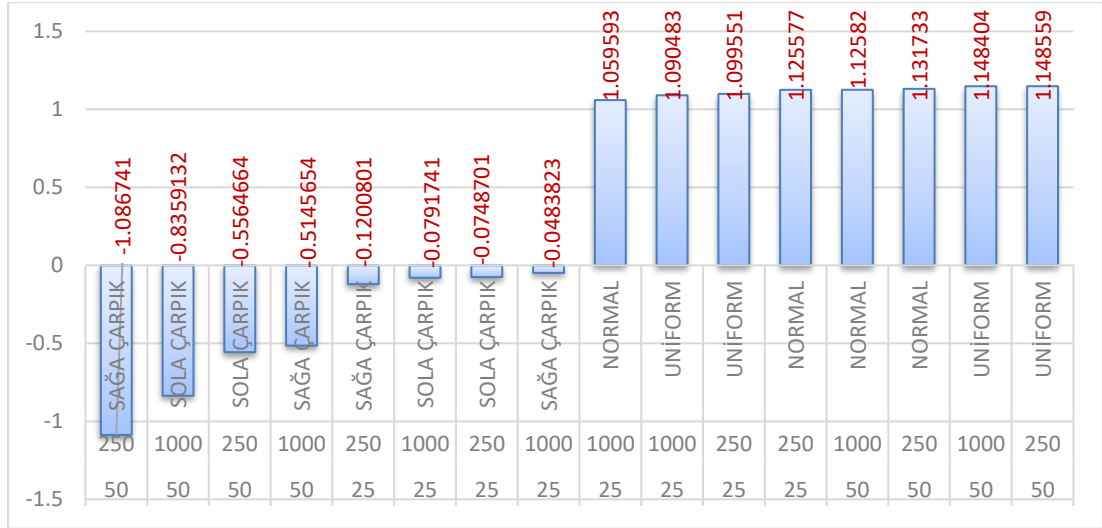
Grafik 11. Madde Havuzunun Kullanılma Derecesine İlişkin Dağılım



Grafik 12. Hedeflenen Kullanım Sıklık Oranından ($0 < r_i \leq .20$) Düşük Oranda Kullanılan Madde Sayısının Dağılımı



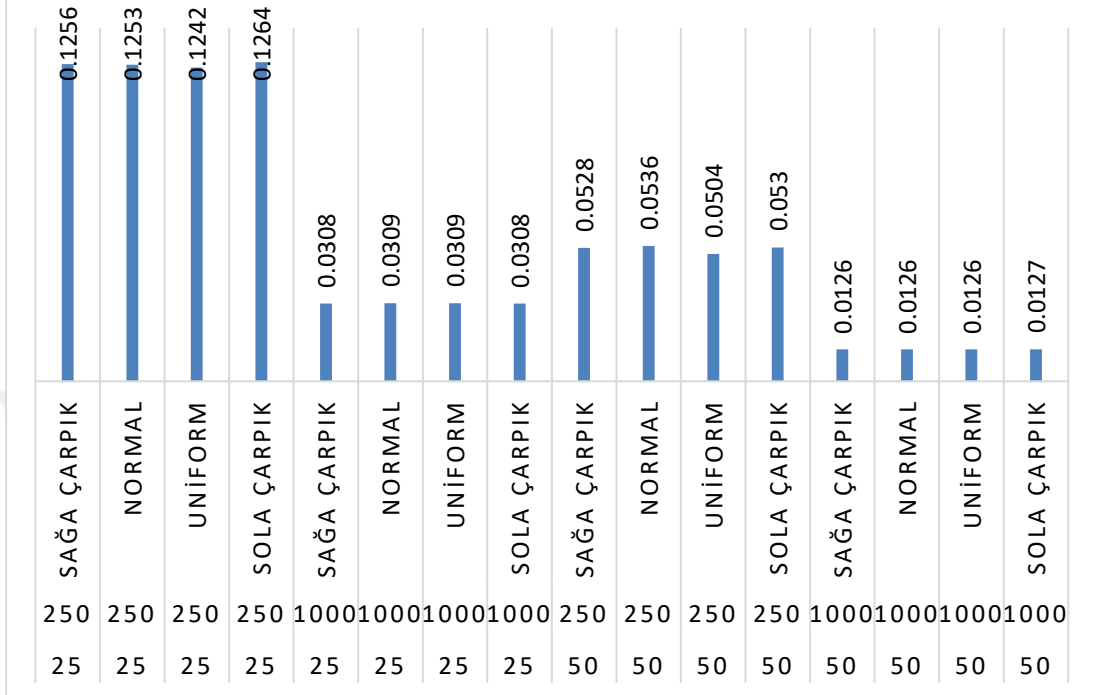
Grafik 13. Hedeflenen Kullanım Sıklık Oranından ($r_i > .20$) Yüksek Oranda Kullanılan Madde Sayısının Dağılımı



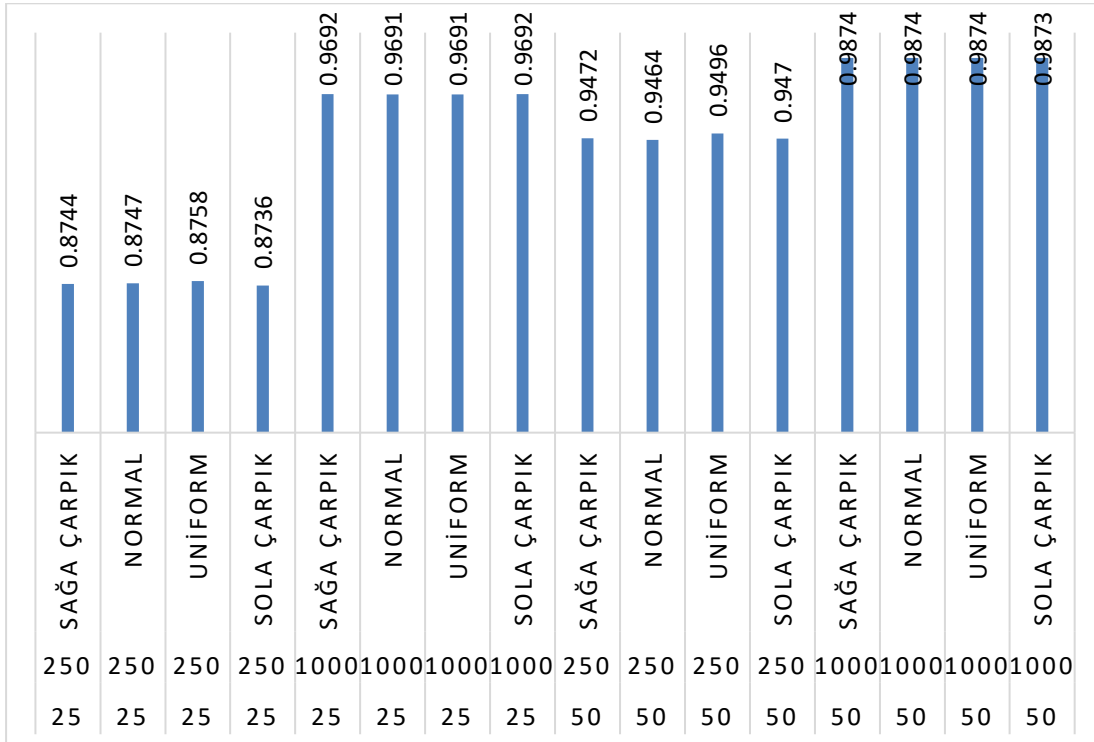
Grafik 14. Hedeflenen Kullanım Sıklık Oranından ($r_i = .20$) Yüksek Oranda Madde Kullanım Sıklıklarına İlişkin Çarpıklık Katsayıları Dağılımı

EK-5

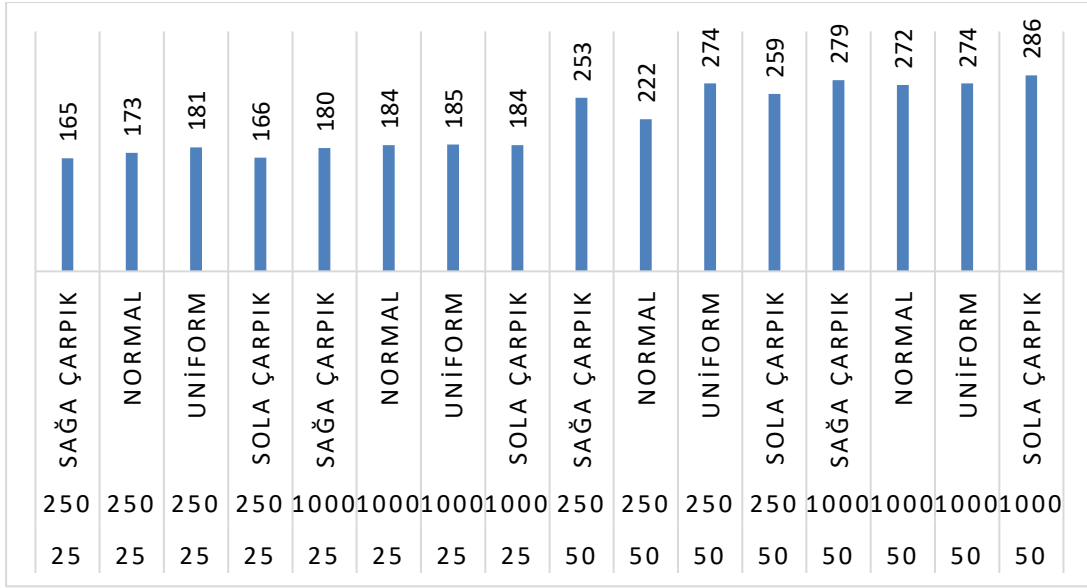
**Sympson Hetter Stratejisi Kullanıldığında
Madde Kullanım Sıklıklarının Dağılımı**



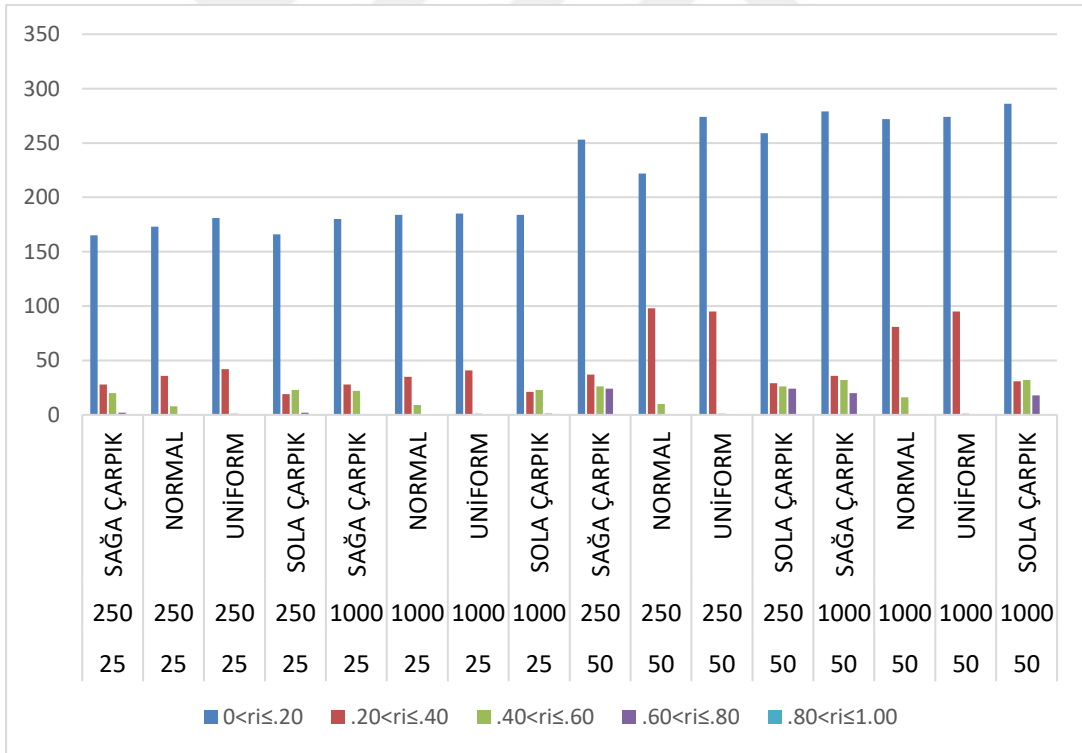
Grafik 15. Madde Havuzundan Yararlanmamaya İlişkin Dağılım



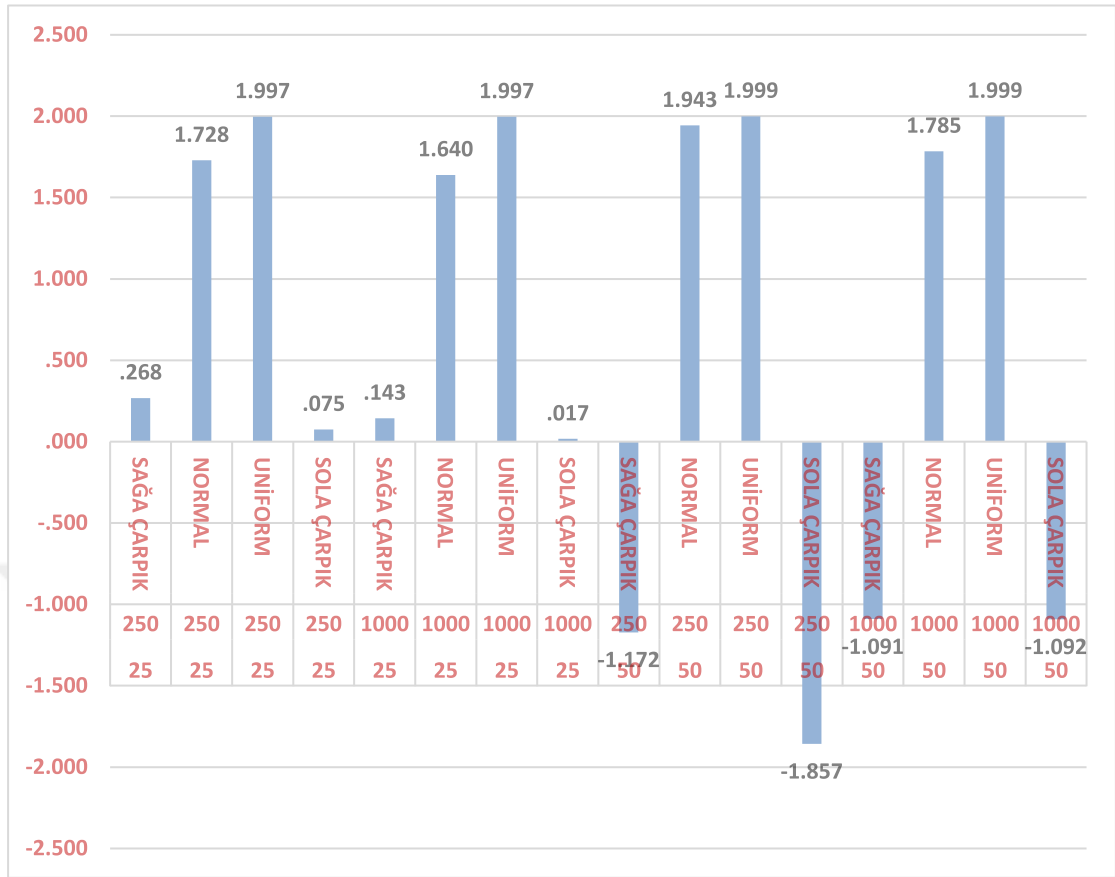
Grafik 16. Madde Havuzunun Kullanılma Derecesine İlişkin Dağılım



Grafik 17. Hedeflenen Kullanım Sıklık Oranından ($0 < r_i \leq .20$) Düşük Oranda Kullanılan Madde Sayısının Dağılımı

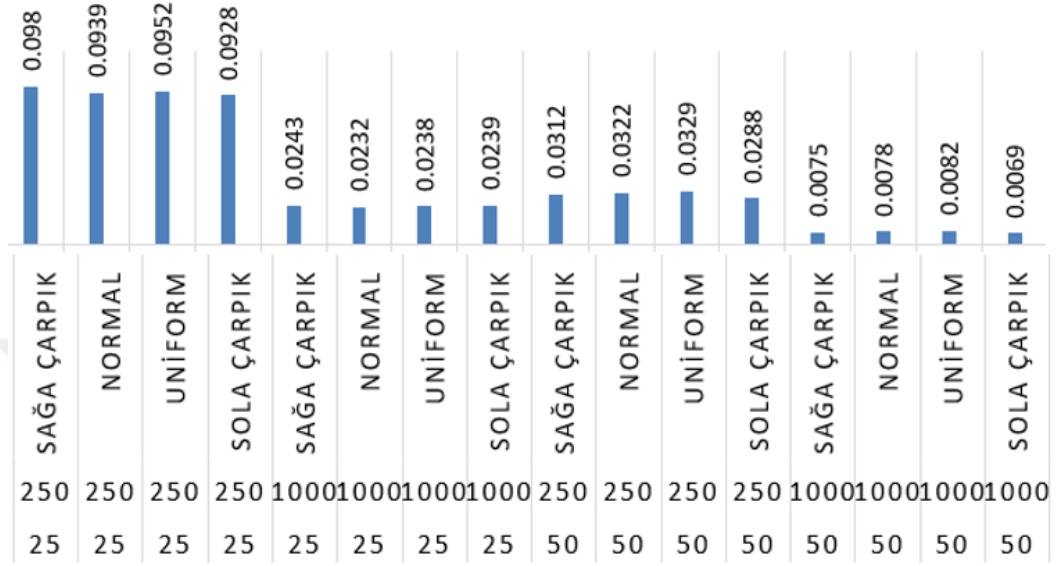


Grafik 18. Hedeflenen Kullanım Sıklık Oranından ($r_i > .20$) Yüksek Oranda Kullanılan Madde Sayısının Dağılımı

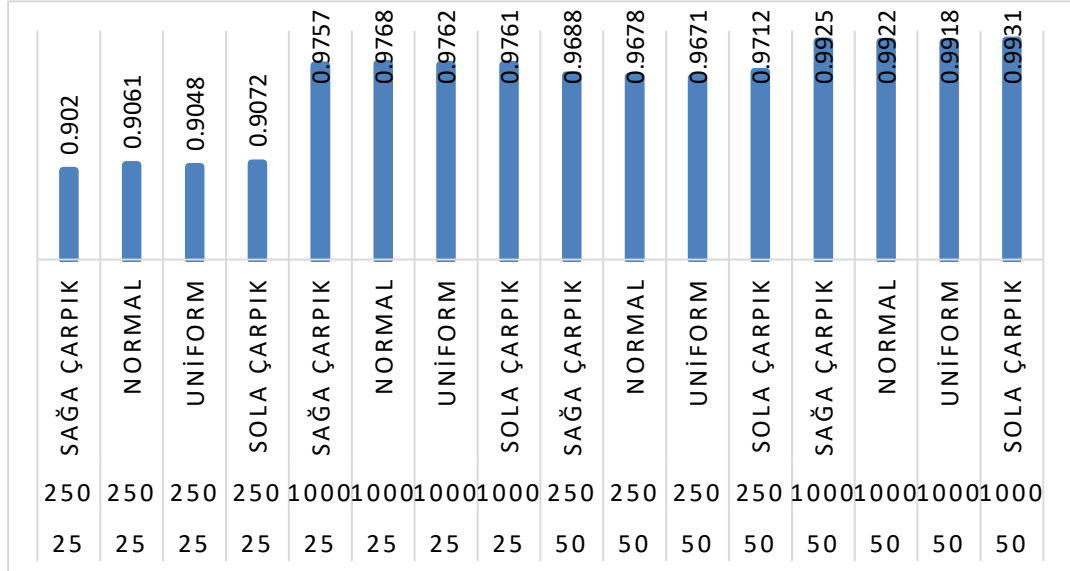


Grafik 19. Hedeflenen Kullanım Sıklık Oranından ($r_i = .20$) Yüksek Oranda Madde Kullanım Sıklıklarına İlişkin Çarpıklık Katsayıları Dağılımı

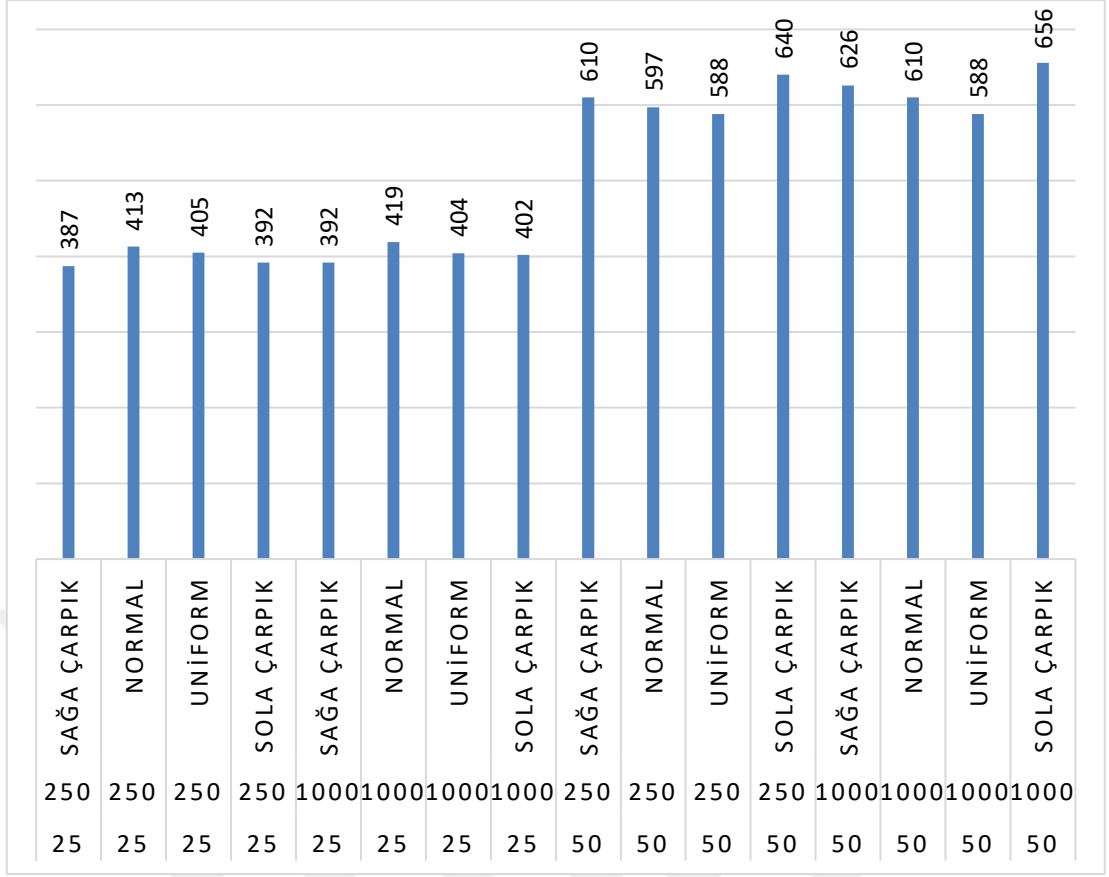
EK-6
Aşamalı Düşürme Stratejisi Kullanıldığında
Madde Kullanım Sıklıklarının Dağılımı



Grafik 20. Madde Havuzundan Yararlanmamaya İlişkin Dağılım



Grafik 21. Madde Havuzunun Kullanılma Derecesine İlişkin Dağılım



Grafik 22. Hedeflenen Kullanım Sıklık Oranından ($0 < r_i \leq 20$) Düşük Oranda Kullanılan Madde Sayısının Dağılımı

EK-7
Test Örtüşme İndeksleri Arasındaki Farklara İlişkin
z Değerleri

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	YOK-SH	SH-AD	YOK-AD
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	0.998816	6.650419	7.716719
		Normal	0.434794	4.718236	5.151285
		Tek Biçimli	0.524202	3.549492	4.068186
		Sola Çarpık	1.498506	7.373083	9.009985
	Geniş (N=1000)	Sağa Çarpık	1.891014	12.83317	14.83875
		Normal	0.874213	9.716817	10.59054
		Tek Biçimli	1.056632	6.991687	8.038333
		Sola Çarpık	2.775815	14.06539	17.06137
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	0.501695	8.038838	8.597521
		Normal	0.202735	3.853052	4.056204
		Tek Biçimli	0.235293	2.819017	3.053578
		Sola Çarpık	0.719234	8.12565	8.929501
	Geniş (N=1000)	Sağa Çarpık	0.922919	15.43542	16.45214
		Normal	0.309987	8.070622	8.38182
		Tek Biçimli	0.479468	5.558775	6.036951
		Sola Çarpık	1.266392	15.47479	16.87065

EK-8

**Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerine
İlişkin Uyum Katsayılarının r' Dönüştürülmüş Değerleri**

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Kontrolü Yapılmadığında	Sympson Hetter Stratejisi	Aşamalı Düşürme Stratejisi
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	3.800201	3.70899	3.036272
		Normal	4.402363	4.402363	4.402363
		Tek Biçimli	4.49154	4.530902	4.424516
		Sola Çarpık	3.911823	3.852906	3.383421
	Geniş (N=1000)	Sağa Çarpık	3.800201	3.752521	3.13200
		Normal	4.402363	4.402363	4.402363
		Tek Biçimli	4.60512	4.60512	4.60512
		Sola Çarpık	3.978614	3.978614	3.597344
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	4.055714	4.055714	3.159083
		Normal	4.951719	4.951719	4.820529
		Tek Biçimli	4.951719	4.904061	4.856403
		Sola Çarpık	4.258497	4.258497	3.597344
	Geniş (N=1000)	Sağa Çarpık	4.16953	4.204084	3.30315
		Normal	4.783473	4.783473	4.783473
		Tek Biçimli	4.783473	4.783473	4.783473
		Sola Çarpık	4.402363	4.402363	3.752521

EK-9

Yanlılık Değerlerine İlişkin Mahalanobis Uzaklıklarının Hesaplanmasında Kullanılan Syntax

```
> X1<-c(.1236, -.0025, .0111, -.0922, .11126, .00095, .0095, -.0881, .06676, -.00035, .00557, -.0522, .05873, -.00049, .0059, -.0465)

> X2<-c(.1249, -.0035, .0094, -.0963, .11349, -.00132, .00968, -.08949, .06899, -.00029, .0044, -.0524, .05969, -.00013, .00503, -.04673)

> X3<-c(.211, -.0044, .1277, -.1420, .18442, -.00161, .01245, -.1263, .152597, -.0025, .0076, -.0985, .12754, -.00021, .0074, -.0816)

> #size (n x p) #Means

> x<-cbind(x1, x2, x3)

> Mean<-colMeans(x)

> Sx<-cov(x)

> MAH.Bias<-mahalanobis(x, Mean, Sx)

> MAH.Bias

[1] 4.1880260 0.1997771 11.5117900 2.6745272 3.6159625 0.2678247

[7] 0.3164295 2.6184185 2.2434349 0.2545251 0.2323996 1.4157013

[13] 14.0409076 0.1937917 0.2594363 0.9670480

>
```

EK-10**RMSE Değerlerine İlişkin Mahalanobis Uzaklıklarının Hesaplanmasında Kullanılan
Syntax**

```

> y1<-c(.15234, .0428, .1014, .11554, .00447, .00154 , .0032, .00356, .08632,
.02414, .0546, .06814, .002479, .00082, .001755, .001962)

> y2<-c(.1584, .0434, .09868, .1201, .00456, .001535, .0032, .0036, .08927, .0242,
.05435, .06766, .002497, .00086, .001766, .00197)

> y3<-c(.2715, .0497, .1112, .18223, .0076, .0017, .003651, .0051, .20432, .02853,
.06382, .1321, .00556, .00091, .00204, .00354)

> #size (n x p) #Means
> y<-cbind(y1, y2, y3)
> Mean<-colMeans(y)
> Sy<-cov(y)
> MAH.RMSE<-mahalanobis(y, Mean, Sy)
>MAH.RMSE

[1] 7.3195572 0.9203546 10.6390363 5.6550434 0.6098550 0.6994168

[7] 0.6405209 0.6329877 8.4398631 0.2880340 1.1594230 5.2031502

[13] 0.6717483 0.7412237 0.6956466 0.6841390

>

```

EK-11

BOBUT Uygulamalarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Farklı Ölçme Koşullarında Ölçme Duyarlılığına Etkisi

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Madde Kullanım Sıklığı Kontrol Yöntemleri								
			YOK			SH			AD		
			Uyum	Yanlılık	RMSE	Uyum	Yanlılık	RMSE	Uyum	Yanlılık	RMSE
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.9990	.1236	.1523	.9988	.1249	.1584	.9954	.2111	.2715
		Normal	.9997	-.0025	.0428	.9997	-.0035	.0434	.9997	-.0044	.0497
		Tek Biçimli	.9997	.0111	.1014	.9998	.0094	.0987	.9997	.1277	.1112
		Sola Çarpık	.9992	-.0922	.1155	.9991	-.0963	.1201	.9977	-.1420	.1822
	Geniş (N=1000)	Sağa Çarpık	.9990	.1113	.0045	.9989	.1135	.0046	.9962	.1844	.0076
		Normal	.9997	.0009	.0015	.9997	-.0013	.0015	.9997	-.0016	.0017
		Tek Biçimli	.9998	.0095	.0032	.9998	.0097	.0032	.9998	.0125	.0036
		Sola Çarpık	.9993	-.0881	.0036	.9993	-.0895	.0036	.9985	-.1263	.0051
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.9994	.0668	.0863	.9994	.0689	.089	.9964	.1526	.2043
		Normal	.9999	-.0004	.0241	.9999	-.0003	.0242	.9999	-.0025	.0285
		Tek Biçimli	.9999	.0056	.0546	.9999	.0044	.0544	.9999	.0076	.0638
		Sola Çarpık	.9996	-.0522	.0681	.9996	-.0524	.0677	.9985	-.0985	.1321
	Geniş (N=1000)	Sağa Çarpık	.9995	.0587	.0025	.9996	.0597	.0025	.9973	.1275	.0055
		Normal	.9998	-.0005	.0008	.9998	-.0001	.0009	.9998	-.0002	.0009
		Tek Biçimli	.9998	.0059	.0018	.9998	.0051	.0018	.9998	.0074	.0020
		Sola Çarpık	.9997	-.0465	.0020	.9997	-.0467	.0020	.9989	-.0816	.0035

EK-12

BOBUT Uygulamalarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Farklı Ölçme Koşullarında Test Güvenliğine Etkisi

Test Uzunluğu	Örneklem	Yetenek Dağılımı	Madde Kullanım Sıklığı Kontrol Yöntemleri														
			YOK					SH					AD				
			SS	Max	Yarar	χ^2	Örtüşme	SS	Max	Yarar	χ^2	Örtüşme	SS	Max	Yarar	χ^2	Örtüşme
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.097	.834	.1259	375.06	.396	.091	.622	.1256	330.99	.353	.048	.193	.0980	90.78	.112
		Normal	.074	.750	.1254	218.66	.239	.071	.486	.1253	201.84	.223	.037	.117	.0939	54.50	.076
		Tek Biçimli	.068	.677	.1242	183.84	.205	.064	.448	.1242	165.06	.187	.038	.152	.0952	58.49	.079
		Sola Çarpık	.101	.922	.1272	406.71	.429	.092	.642	.1264	341.64	.363	.044	.184	.0928	77.97	.099
	Geniş (N=1000)	Sağa Çarpık	.095	.830	.0308	362.35	.386	.090	.594	.0308	321.22	.346	.047	.189	.0243	89.39	.114
		Normal	.075	.739	.0309	224.52	.249	.072	.492	.0309	208.19	.232	.037	.117	.0232	54.17	.078
		Tek Biçimli	.068	.683	.0310	182.74	.207	.064	.454	.0309	164.07	.188	.038	.151	.0238	58.23	.082
		Sola Çarpık	.099	.915	.0312	387.70	.412	.091	.623	.0308	328.23	.352	.044	.181	.0239	76.78	.101
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.142	.840	.0534	403.50	.452	.138	.727	.0528	381.86	.429	.063	.200	.0312	79.54	.126
		Normal	.098	.761	.053	193.19	.239	.096	.508	.0536	185.85	.232	.054	.151	.0322	58.51	.105
		Tek Biçimli	.089	.686	.051	158.61	.205	.087	.463	.0504	150.22	.197	.055	.183	.0330	60.25	.107
		Sola Çarpık	.143	.9294	.0533	406.23	.454	.137	.733	.0530	375.25	.422	.060	.196	.0288	72.14	.119
	Geniş (N=1000)	Sağa Çarpık	.139	.836	.0127	389.23	.439	.136	.721	.0126	369.19	.419	.063	.197	.0075	78.38	.128
		Normal	.099	.749	.0127	197.60	.247	.098	.523	.0126	191.56	.241	.054	.151	.0078	57.01	.106
		Tek Biçimli	.089	.689	.0126	157.27	.206	.086	.461	.0126	148.79	.198	.055	.181	.0082	59.82	.109
		Sola Çarpık	.139	.921	.0128	387.87	.437	.134	.714	.0127	359.85	.409	.060	.192	.0069	71.50	.121