

**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

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

**TEST TAHRİFATININ BELİRLENMESİNDE SİLME
BELİRLEME İNDEKSİNİN I. TİP HATA VE
İSTATİSTİKSEL GÜÇ AÇISINDAN
İNCELENMESİ**

DOKTORA TEZİ

ESRA KINAY ÇİÇEK

**ANKARA
TEMMUZ, 2020**



**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

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

**TEST TAHRİFATININ BELİRLENMESİNDE SİLME
BELİRLEME İNDEKSİNİN I. TİP HATA VE
İSTATİSTİKSEL GÜÇ AÇISINDAN
İNCELENMESİ**

DOKTORA TEZİ

ESRA KINAY ÇİÇEK

DANIŞMAN: PROF. DR. ÖMAY ÇOKLUK BÖKEOĞLU

**ANKARA
TEMMUZ, 2020**

ETİK İLKELERE UYGUNLUK BİLDİRİMİ

Tez içindeki bütün bilgileri akademik yazım kurallarına uygun biçimde raporlaştırdığımı ve bunları etik ilkelere (atıfta bulunulan tüm yapıtlara kaynaklarda yer verilmesi, tezde kullanılan bilgi ve belgelere resmi yollarla ulaşılması ve bunların aslı bozulmadan kullanılması vb.) uygun olarak elde ettiğimi ve sunduğumu bildiririm.



Esra KINAY ÇİÇEK

ÖZET

TEST TAHRİFATININ BELİRLENMESİNDE SİLME BELİRLEME İNDEKSİNİN I. TİP HATA VE İSTATİSTİKSEL GÜÇ AÇISINDAN İNCELENMESİ

KINAY ÇİÇEK, Esra

Doktora Tezi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanı: Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU

Temmuz, 2020, xiv + 175 sayfa

Bu araştırmada, test tahrifatının belirlenmesinde kullanılan Silme Belirleme İndeksi'nin (SBI) farklı koşullar (örneklem büyüklüğü, yetenek düzeyi, madde sayısı ve hileli silme oranı) altında, I. Tip hata oranı ve istatistiksel güç bakımından nasıl değiştiğinin belirlenmesi amaçlanmaktadır.

Bu amaç doğrultusunda 250, 2500, 25000 ve 49995 test alan için 50 ve 160 madde üzerinden Sınıflamalı Tepki Modeline göre simülatif yanıt örüntüleri üretilmiştir. Veri setlerinde iyicil silmelerden tesadüfi, kaydırma ve dizi sonu silmeleri simüle edilmiştir. İyicil silmeleri içeren veri setlerine, ayrı ayrı sabit silme ve değişken silme yapılarak yanıtlar tahrifata uğratılmıştır. Sonrasında veri setleri birleştirilerek iyicil ve hileli silmeleri içeren veri setleri analize hazır hale getirilmiştir.

Süreklilik düzeltmesi uygulanan ve uygulanmayan SBI'nin ürettiği I. Tip hata oranları ve istatistiksel güç değerleri; 500, 5000, 50000 ve 99990 örneklem büyüklüklerinde beş farklı yetenek düzeyine göre, 50 ve 160 maddelik veri setlerinde, %10, %20 ve %30 oranlarında sabit silme ve değişken silme için .00001, .0001, .0005, .001, .005, .01 ve .05 α seviyelerinde incelenmiştir.

Süreklilik düzeltmesi uygulanmadığında I. Tip hata oranları genellikle iyi kontrol edilememiştir. Süreklilik düzeltmesi uygulanması tüm alfa seviyelerinde I. Tip hata oranlarını düşürürken istatistiksel gücün de bir miktar düşmesine neden olmuştur. I. Tip hata oranları 50 maddelik veri setlerinde 160 maddelik veri setlerine göre daha iyi kontrol edilmiştir. Örneklem büyüklüğü arttıkça I. Tip hata oranlarında azalma görülmüştür. Sabit silme oranları arttıkça tahrifata uğrayan test alanlar daha doğru tespit edilmiş ve

istatistiksel güç artmıştır. Yetenek düzeyi arttıkça istatistiksel gücün düştüğü belirlenmiştir. Araştırma sonucunda Türkiye’de uygulanan yüksek riskli sınavlarda SBİ ile tahrifatın belirlenmesine yönelik çalışmaların yürütülmesi önerilmektedir.

Anahtar Sözcükler: Test Tahrifatı, Silme Belirleme İndeksi, İyicil Silme, Hileli Silme, Öğretmen Kopyası



ABSTRACT

ANALYSIS OF THE ERASURE DETECTION INDEX IN TERMS OF TYPE I ERROR AND STATISTICAL POWER IN DETERMINING TEST TAMPERING

KINAY ÇİÇEK, Esra

PhD Dissertation, Department of Measurement and Evaluation

Supervisor: Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU

July 2020, xiv + 175 pages

This study aims to examine Type I error rates and statistical power of the Erasure Detection Index (EDI) when used in determining test tampering under different conditions (sample size, ability level, number of different items and fraudulent erasure rates).

For this purpose, simulative response patterns for 250, 2500, 25000 and 49995 examinees were produced using the Nominal Response Model including 50 and 160 items. Three types of benign erasures, random, misalignment, and string-end erasures, were simulated in the datasets. Responses were tampered by making individually fixed and variable erasure to the datasets containing benign erasures. Then, combining datasets, datasets containing both types of benign erasure and fraudulent erasure were prepared for analysis.

According to five different ability levels of 500, 5000, 50000 and 99990 sample sizes, in data sets of 50 and 160 items, Type I error rates and statistical power values produced by EDI with and without continuity correction were examined in alpha levels of .00001, .0001, .0005, .001, for 10%, 20% and 30% fraudulent fixed erasure and variable erasure.

Without continuity correction, Type I error rates were generally not well controlled. The application of continuity correction decreased the Type I error rates at all alpha levels and caused a decrease in the statistical power. Type I error rates were better controlled in the 50-item datasets than the 160-item datasets. As the sample size increased, Type I error rates decreased. As the fixed erasure rates increased, the statistical power increased and the examinees were detected more accurately. It was determined that

the statistical power increased as the ability level decreased. According to the findings of this study, it is recommended that large scale testing organizations should carry out studies to determine the test tampering with EDI in high-stakes tests in Turkey.

Key Words: Test Tampering, Erasure Detection Index, Benign Erasure, Fraudulent Erasure, Teacher Cheating



ÖNSÖZ

Sonuçları hem bireysel hem de toplumsal olarak büyük önem taşıyan geniş ölçekli test uygulamalarının geçerlik ve güvenilirliğinin sağlanması son derece önemlidir. Test puanlarının geçerlik ve güvenilirliğinin sağlanmasının temelinde yer alan ve uluslararası alanyazını ile dünya kamuoyunda sıklıkla gündeme gelen sınav güvenliğine ilişkin çalışmaların ülkemizde eksik olduğu görülmüştür. Özellikle, kopyanın bir türü olarak adlandırılan ve sınav güvenliğini tehdit eden test tahrifatına yönelik ülkemizde bir çalışma yapılmamış olması bu çalışmanın yürütülmesinin temel çıkış noktası olmuştur. Bu çalışma ile test tahrifatının belirlenmesinde kullanılan yöntemlerden biri olan Silme Belirleme İndeksinin farklı koşullarda etkililiği değerlendirilerek uygulamaya dönük öneriler elde edilmeye çalışılmıştır.

Doktora tez sürecinde yönlendirmeleriyle bana destek olan, çalışma motivasyonumu artıran, bana güvenen ve destekleyen değerli tez danışmanım Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU'na teşekkürlerimi sunarım. Tez izleme komitesinde yer alan ve değerli bilgileriyle tezime katkıda bulunan Prof. Dr. Duygu ANIL, Doç. Dr. Hamide Deniz GÜLLEROĞLU ve tez sürecinde emekli olan Prof. Dr. Nizamettin KOÇ'a; tez jürimde yer alan ve tezime katkı sağlayan Prof. Dr. Nilüfer KAHRAMAN ve Doç. Dr. Seher YALÇIN'a; bölümdeki değerli hocalarıma; tez çalışmamda destek aldığım Prof. Dr. Özlem İLK DAĞ'a; akademik disiplinini örnek aldığım ve bilgisine her zaman güvendiğim Dr. Öğr. Üyesi Ünal TÖNGÜR'e teşekkür ederim.

Yüksek lisansta tanıştığım günden beri hem akademik hem de sosyal hayatımın bir parçası olan, iyi günümde ve zor günümde yanımda olan sevgili arkadaşım Dr. Öğr. Üyesi Selda ÖRS ÖZDİL ve bitmek bilmez soru ve sorunlarımla her daim ilgilenen canım arkadaşım Araş. Gör. Dr. Cansu AYAN'a akademik ve psikolojik olarak hep yanımda oldukları, tezime görüş ve önerileriyle katkı sağladıkları için teşekkür ederim.

Bana her daim inanan ve güvenen, girdiğim çıkmazlarda beni yüreklendiren, ümitsizliğe kapıldığım günlerde sabrıyla destek olan, hayatı paylaştığım sevgili eşim Ali Doğan ÇİÇEK'e; her zaman yanımda olduklarını hissettiren, iyi ki varlar dediğim sevgili ablalarıma; sonsuz sabrı, sevgisi, emeğiyle beni yetiştirerek bugünlere gelmemi sağlayan, her koşulda bana güvenen, tezimi tamamladığım günü benden daha büyük bir heyecanla bekleyen canım annem Belhuzar KINAY'a sonsuz teşekkürlerimi sunarım.

İÇİNDEKİLER

	Sayfa
ETİK İLKELERE UYGUNLUK BİLDİRİMİ.....	iii
ÖZET	iv
ABSTRACT	vi
ÖNSÖZ.....	viii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ.....	xi
ŞEKİLLER DİZİNİ	xiii
KISALTMALAR/SİMGELER	xiv
BÖLÜM 1.....	1
GİRİŞ.....	1
Problem.....	1
Amaç.....	19
Önem.....	20
Sayıtlılar.....	21
Sınırlılıklar	21
Tanımlar.....	22
BÖLÜM 2.....	23
YÖNTEM.....	23
Araştırma Modeli.....	23
Simülasyon Deseni.....	23
Sabit Koşullar.....	23
Manipüle Edilen Koşullar	24
Verilerin Üretilmesi	27
Verilerin Silinme Simülasyonu.....	29
Verilerin Çözümlemesi	38
Üretilen Verilerden Elde Edilen Yetenek Parametrelerinin Normal Dağılım Varsayımının Kontrolü.....	38
Üretilen Verilerden Elde Edilen Yetenek Parametrelerinin Tek Boyutluluk Varsayımının Kontrolü.....	39
Üretilen Verilerde Madde Yanıtlarının Yerel Bağımsızlık Varsayımının Kontrolü	40
Yanıtları Silinmemiş Maddelere Dayalı Yetenek Kestirimi	42
Bireysel Yetenek Kestiriminde Yanlılık (Bias) ve Hataların Ortalama Karekökü (RMSE) Değerlerinin Hesaplanması.....	42
SBİ Hesaplanması	45
Testin İstatistiksel Gücünün ve I. Tip Hata Oranının Hesaplanması	47

BÖLÜM 3.....	49
BULGULAR VE YORUMLAR	49
500 Örneklem Büyüklüğü İçin SBI'nin İstatistiksel Güç ve I. Tip Hata Bulguları..	50
5000 Örneklem Büyüklüğü İçin SBI'nin İstatistiksel Güç ve I. Tip Hata Bulguları	67
50000 Örneklem Büyüklüğü İçin SBI'nin İstatistiksel Güç ve I. Tip Hata Bulguları	86
99990 Örneklem Büyüklüğü İçin SBI'nin İstatistiksel Güç ve I. Tip Hata Bulguları	105
BÖLÜM 4.....	125
SONUÇLAR VE ÖNERİLER	125
Sonuçlar	125
Öneriler	127
Uygulayıcılara Öneriler.....	127
Araştırmacılara Öneriler.....	128
KAYNAKLAR.....	131
EKLER	139
EK 1. Etik Kurul Onayı	140
EK 2. 50 ve 160 Maddelik Veri Setlerine Ait R'da Üretilen Seçenek Eğim ve Kesişim Parametreleri.....	141
EK 3. 50 Madde 50000 Örneklem Büyüklüğü İçin Silme Belirleme İndeksi Algoritması	147
EK 4. Silinmemiş Maddeler Üzerinden Yetenek Kestirimine Ait Yanlılık ve RMSE Bulguları	152
BENZERLİK BİLDİRİMİ	173
ÖZGEÇMİŞ.....	175

TABLolar DİZİNİ

Tablo	Sayfa
Tablo 1. Eğitici Kopyasının Taksonomisi	3
Tablo 2. Simülasyon Deseni	26
Tablo 3. Madde Sayısı ve Örneklem Büyüklüğü Koşullarına Göre Değişken Silme İçin Belirlenen Gerçek Puanlar ve Kullanılan Kesme Puanları	34
Tablo 4. Örneklem Büyüklüğü ve Silme Türlerine Göre Silme Yapılması Beklenen Tahmini Birey Sayısı	35
Tablo 5. Örneklem Büyüklüğü ve Silme Türlerine Göre Silme Yapılan Tahmini Birey Sayısı (Revize).....	35
Tablo 6. Sabit ve Değişken Silmeli Veri Setleri Birleştirildiğinde Örneklem Büyüklüğü ve Silme Türlerine Göre Silme Yapılan Tahmini Birey Sayısı	37
Tablo 7. Örneklem Büyüklüğü ve Tahrifat Türüne Göre Tahrifatlı Birey Yüzdesi.....	37
Tablo 8. DETECT Analizi Sonuçları İçin Referans Değerler	39
Tablo 9. Üretilen Veriler İçin Hesaplanan Betimsel İstatistik, Tek Boyutluluk ve Yerel Bağımsızlık Analizlerine Ait Sonuçlar	41
Tablo 10. Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi Yapılmaksızın SBİ'nin I. Tip Hataları (n= 500)	51
Tablo 11. Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi İle SBİ'nin I. Tip Hataları (n= 500).....	54
Tablo 12. Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBİ'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 500)	57
Tablo 13. Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBİ'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 500)	61
Tablo 14. Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi Yapılmaksızın SBİ'nin I. Tip Hataları (n= 5000)	68
Tablo 15. Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi İle SBİ'nin I. Tip Hataları (n= 5000).....	71
Tablo 16. Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBİ'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 5000)	75
Tablo 17. Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBİ'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 5000)	79
Tablo 18. Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi Yapılmaksızın SBİ'nin I. Tip Hataları (n= 50000)	87
Tablo 19. Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi İle SBİ'nin I. Tip Hataları (n= 50000).....	90

Tablo 20. Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 50000)	94
Tablo 21. Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 50000)	98
Tablo 22. Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi Yapılmaksızın SBI'nin I. Tip Hataları (n= 99990)	106
Tablo 23. Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi İle SBI'nin I. Tip Hataları (n= 99990).....	109
Tablo 24. Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 99990)	113
Tablo 25. Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 99990)	117
Tablo 26. 50 Maddelik Veri Setinde 49995 Örneklem Büyüklüğüne Ait Eğim (λ) ve Kesişim (ζ) Parametreleri	141
Tablo 27. 160 Maddelik Veri Setinde 49995 Örneklem Büyüklüğüne Ait Eğim (λ) ve Kesişim (ζ) Parametreleri	143
Tablo 28. 50 ve 160 Madde Üzerinden 500 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü	152
Tablo 29. 50 ve 160 Madde Üzerinden 500 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü.....	155
Tablo 30. 50 ve 160 Madde Üzerinden 5000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü	157
Tablo 31. 50 ve 160 Madde Üzerinden 5000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü.....	160
Tablo 32. 50 ve 160 Madde Üzerinden 50000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü	162
Tablo 33. 50 ve 160 Madde Üzerinden 50000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü.....	165
Tablo 34. 50 ve 160 Madde Üzerinden 99990 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü	167
Tablo 35. 50 ve 160 Madde Üzerinden 99990 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü.....	170

ŞEKİLLER DİZİNİ

Şekil	Sayfa
Şekil 1. Yanıt Kâğıtlarında Oluşabilecek İşaret ve Silme Örnekleri.....	6
Şekil 2. $N= 50$ ve $p= .02$ Olan Binominal Dağılımdan Elde Edilen Tesadüfi Silme Sayılarının Olasılık Grafiği	30
Şekil 3. $N= 160$ ve $p= .02$ Olan Binominal Dağılımdan Elde Edilen Tesadüfi Silme Sayılarının Olasılık Grafiği	31
Şekil 4. $N= 50$ ve $p= .25$ Olan Binominal Dağılımdan Elde Edilen Kaydırma Yapılan Madde Sayılarının Olasılık Grafiği	32
Şekil 5. $N= 160$ ve $p= .25$ Olan Binominal Dağılımdan Elde Edilen Kaydırma Yapılan Madde Sayılarının Olasılık Grafiği	32

KISALTMALAR/SİMGELER

MTK	Madde Tepki Kuramı
θ	Yetenek Düzeyi
STM	Sınıflamalı Tepki Modeli (Nominal Response Model. NRM)
SBİ	Silme Belirleme İndeksi (Erasure Detection Index. EDI)
RMSE	Root Mean Square Error (Hata Ortalamalarının Karekökü. HOK)
Bias	İstatistiksel Yanlılık
λ	STM'de Eğim Parametresi
ζ	STM'de Kesişim Parametresi
2PL	İki Parametrelili Lojistik Model
3PL	Üç Parametrelili Lojistik Model
$P_i(\theta)$	Maddeye Doğru Yanıt Verme Olasılığı
$SE(\theta)$	Tahminin Standart Hatası
X_j	Gözlenen Puan
T_j	Gerçek Puan
E	Hata Puanı
$I_{E,j}$	Silme yapılmış maddeler
$\hat{\theta}_{j[i \notin I_{E,j}]}$	Silme yapılmamış maddeler üzerinden kestirilen yetenek düzeyi
$X_{j,I_{E,j}}$	Silme yapılmış maddelerin gözlenen değeri
θ_j	Tüm maddeler üzerinden kestirilen yetenek düzeyi
$\hat{\epsilon}_{j[i \notin I_{E,j}]}$	Silme yapılmamış maddeler üzerinden kestirilen yetenek düzeyinin hata değeri
$\hat{\theta}_j$	İyicil silmeleri de içeren (tahrifatlı maddeler hariç) maddeler üzerinden kestirilen yetenek düzeyi

BÖLÜM 1

GİRİŞ

Bu bölümde araştırmanın problem durumu açıklanmış; amaç, önem, sayıtlar, sınırlılıklar ve tanımlara yer verilmiştir.

Problem

Öğrenme düzeyini belirlemek üzere uygulanan sınavların, değerlendirme amacına hizmet edebilmesi için güvenilirlik ve geçerlik gibi psikometrik özelliklere sahip olması gerekmektedir. Güvenirlik en genel anlamıyla ölçme sonuçlarının hatalardan arınıklık derecesi olarak ifade edilirken, geçerlik ise ölçme aracının ölçülmek istenen özelliği başka özelliklerle karıştırmadan ölçebilme derecesidir (Turgut, 1983). Ölçme aracından elde edilen puanların geçerlik ve güvenilirliğinin sağlanması, ölçek geliştirme süreci kadar, ölçeğin uygun koşullarda uygulanması ve puanlanmasına da bağlıdır. Bu durum “sınav güvenliği” adı altında incelenmektedir. Sınav güvenliği (Fremer ve Ferrara, 2013),

- sınavın uygun şekilde tedarik edilmesinin sağlanması,
- sınavın uygun şekilde yönetiminin sağlanması,
- sınav öncesi, sınav esnası ve sonrasında kopyanın caydırılması,
- kopyanın belirlenmesi ve soruşturulması,
- uygulayıcı eğitiminin yapılması ve uygunsuzluk belirlendiğinde yaptırıma başvurulması,
- sınav materyallerinin gizliliğinin korunması

ile ilgili yönergeleri ve işlemleri içermektedir.

Sınav güvenliğinin temelini kopyanın önlenmesi ve kopyanın belirlenmesi oluşturmaktadır. Cizek (1999) kopyayı “aldatma veya sahtekârlık yoluyla kişinin gerçekte sahip olmadığı bilgiye sahip olduğunu gösterme girişimi” olarak tanımlamış ve kopya yöntemlerini üç ana başlıkta sınıflandırmıştır:

- Bilgiyi diğer bireylerden alma ya da diğerlerine verme
- Yasak olan eşyalar ya da bilgi kaynaklarını kullanma
- Değerlendirme sürecini etkileme

Cizek (1999), sınav yönetimi sırasında yöneticiler tarafından sınıf içinde kopyayı caydırmak için kullanılan işlemleri ve yapılan uygulamaları incelemiştir. Test güvenliğinden emin olmak ve kopyayı önlemek için geniş ölçekli test uygulamalarına yönelik öneriler sunarak bu konuyu geniş bir çerçevede ele almıştır. Önerilerin bir kısmı test programlarının yöneticilerine yöneliktir (testleri yönetmek için gerekli olan durumlar dışında, okul personelinin gizli test materyallerini görmesini yasaklamak vb.). Diğer kısmı ise test yayıncılarına yöneliktir (örneğin, bireysel test kitapçıklarını paketlemek ve mühürlemek, şüpheli silmeler için yanıt kâğıtlarını taramak ve test gruplarında yanıt örüntülerinin benzerliğini analiz etmek vb.).

California Eğitim Departmanı (California Department of Education, CDE) test sonuçlarında gözlenen normal olmayan durumları “test usulsüzlükleri (test irregularities)” olarak ifade etmiş ve test usulsüzlüklerinin öğrenci ve yetkili (test uygulayıcısı, test yöneticisi, öğretmen vb.) olmak üzere iki temel kaynağı olduğunu belirtmiştir. Öğrenciden kaynaklanan usulsüzlükler öğrencinin test süresince kopya çekmesi olarak ifade edilirken, yetkililerden kaynaklanan usulsüzlükler ise test süresince öğrencilere uygunsuz yardım sağlanması, öğrenci yanıt kâğıtlarına müdahale (yanıtların değiştirilmesi ya da boş yanıtların doldurulması) ve test öncesinde öğrencilerin uygunsuz şekilde teste hazırlanması olarak ifade edilmiştir (CDE, 2001; Akt. Haick, 2003). Thiessen (2006) kopyadan söz edildiğinde çoğunlukla öğrencilerin kopya çekme davranışlarına odaklanıldığını ancak son zamanlarda birçok eğitimcinin kopya çektiğinin belirlendiğini ve bu nedenle sınavlarda öğrencileri için kopya çeken öğretmenlere ve müdürlere daha fazla dikkat edilmesi gerektiğini ifade etmektedir.

Kopyanın sadece öğrenciler tarafından çekilmediği, öğretmenlerin de test puanlarını manipüle edebildiği bilinmektedir. Örneğin Amerika’da Hiçbir Çocuk Geride Kalmasın Hareketi (2001) kapsamında testler, hesapverebilirlik ölçütlerinin bir parçası olarak bireylerin işlerini ne kadar iyi yaptıklarını, öğretmenleri, yöneticileri, okulları, bölgeleri ve eyaletleri değerlendirmek için kullanılmaktadır. Son yıllarda birçok büyük test programı, teste katılan birçok bireyin puanlarının geçerliğinin sorgulanmasına neden olan geniş çaplı organize kopya skandalları ile uğraşmak zorunda kalmıştır. Kopya skandalları sıklıkla medyaya da yansımaktadır. Örneğin, 2003 yılında Teksas Bilgi ve Beceri Değerlendirmesi (Texas Assessment of Knowledge and Skills, TAKS) sınavında bir okulda öğretmenlerin üçte ikisinin öğrencilere yardım ederek birçok öğrencinin mükemmele yakın puanlar almasını sağladıkları belirlenmiştir (Wollack ve Fremer, 2013).

Thiessen (2006), gazete haberleri ile üç çalışmanın (Gay, 1990; Nichols ve Berliner, 2004; Shepard ve Dougherty, 1991) bulgularından yola çıkarak testlerde hile yaparken eğitimcilerin kullandıkları yöntemlere ilişkin Tablo 1’de yer alan Eğitici Kopyasının Taksonomisi’ni geliştirmiştir.

Tablo 1

Eğitici Kopyasının Taksonomisi

	Shepard & Dougherty (1991)	Gay (1990)	Nichols & Berliner (2004)
Yanıt Kâğıtlarının Manipülasyonu			
Öğrencilerin yanıtlarını değiştirme	%6.1	%1.5	6 olay
Öğrencilere yanıtları verme	%8.3		8 olay
Boş bırakılan maddeleri işaretleme		%10	11 olay
Test Yönetim Sürecinin Manipülasyonu			
Öğrencilere test maddeleriyle ilgili ipuçları verme	%22.7	%10	7 olay
Öğrenciler için test maddelerini tekrar ifade etme	%18.0		
Öğrencilere ek süre sağlama	%19.6	%15	3 olay
Öğrencilerin okuması gereken maddeleri okuma	%14.1		
Öğrencileri teste dâhil etmeme	%7.4-13.3		41 olay
Test içeriği ile ilgili soruları yanıtlama	%11.7		
Öğrencilere yanıtlanmamış bir madde için belirli bir yanıt işaretlemelerini söyleme			
Öğrencilere kolaylık sağlama ya da uygunsuz şekilde yerleştirme			1 olay
Puanların Raporlaştırılma Sürecinin Manipülasyonu			
Öğrenci puanlarını resmi kayıttan çıkarma			1 olay
Öğrencilere sahte kimlik numaraları vererek yanıtlarını geçersiz saydırma			1 olay
Veri setini yanlış sunma			8 olay
Yeterlilik kriterlerini değiştirme			21 olay
Tutarsız hesapverebilirlik derecelendirmeleri			15 olay
Öğretim Sürecinin ya da Öğretim Felsefesinin Manipülasyonu			
Öğrencilere spesifik test maddelerini öğretme ya da test süresince öğretme	%30.4	%5	4 olay
Gerçek test maddeleri ile alıştırma yapma	%11.3		17 olay
Testi öğretme			15 olay
Öğretim programının daraltılması			13 olay

Kaynak. Thiessen, 2006, 6.

Tablo 1’de ikinci ve üçüncü sütunlar, her bir özel hile yöntemini kullandığını kabul eden eğitimcilerin yüzdesini göstermektedir. Son sütunda, Nichols ve Berliner’in her hile yöntemini kullanan eğitimcileri tanımladığı gazete haberlerinin sayısı yer almaktadır. Thiessen (2006), eğitimci kopyalarının çoğunun, test yönetim süreçlerini, puanların raporlaştırılmasını veya öğretim süreçlerini manipüle etmeyi içerdiğini ifade etmektedir.

Ayrıca Tablo 1’de her üç araştırmacı tarafından da yanıt kâğıtlarının manipülasyonuna ilişkin bulguların yer aldığı görülmektedir.

Yakınındaki (sağında, solunda, önünde arkasında ya da çaprazında konumlanan) bireyden kopya çekmek olarak tanımlanan yanıt kopyalama en yaygın kopya türlerinden biridir. Yanıt kopyalamanın farklı yetenek düzeylerindeki bireyler, farklı uzunluklardaki testler ve farklı miktar ve türdeki kopya için belirlenmesine yönelik çeşitli istatistiksel yöntemler geliştirilmiş ve bu yöntemlerin I. Tip hata oranları ve istatistiksel gücüne yönelik yeterli bilgi paylaşılmıştır. Ancak bu tarz bir kopya girişimi, yalnızca kopya çekenin puanlarının geçerliğini düşürme eğilimindedir. Kopya çekecek olan kişi, başarılı bir kopya kaynağının yakınına oturma şansını yakalayamadığı durumda, kopya çekme bireyin puanında önemli artışa neden olmayacaktır. Bu nedenle, yanıt kopyalama olarak adlandırılan bu tür kopya davranışı, hala ciddi bir geçerlik problemi olmasına rağmen, kopyanın diğer türlerinden daha az zararlı olarak nitelendirilebilir. Birçok birey için sınav geçerliğini ciddi biçimde tehdit eden organize kopya biçimleri (test içeriği hakkında yasadışı yönlendirme, test salonu dışından sürekli iletişim sağlayan sistemler, testin ardından test yanıtlarında tahrifat vb.), yanıt kopyalamadan çok daha ciddi ve kaygı vericidir (Wollack ve Fremer, 2013). Yanıt kopyalamada etik sorun testi alan birey ile ilgili iken, tahrifat (öğrenci yanıtlarının bir otorite tarafından değiştirilmesi) durumlarında etik sorun öğrenci ile ilgili değil, yanıtı değiştiren (tahrif eden) kişi/kişiler ile ilgilidir (Wollack, Cohen ve Eckerly, 2015). Eğitimsel hesapverebilirlik sistemleri çerçevesinde değerlendirme sonucunda eğiticiler ve okullar kaynak kısıtlamasına tabi tutulabilmekte, hak kaybına uğrayabilmekte ve hatta okulun kapanması durumu söz konusu olabilmektedir. Bu durum eğiticilerin hesapverebilirlik sistemlerinin sonuçlarına yönelik kaygı düzeylerinin artması ve test usulsüzlüklerinin gerçekleşmesi ile sonuçlanabilmektedir. Örneğin, Atlanta Devlet Okulları’nda yaşanan skandalda öğretmenlerin, testler tamamlandıktan sonra öğrenci yanıtlarını değiştirdiği belirlenmiş ve öğretmen/yöneticiler tarafından yapılan tahrifatın işaretlerinin aranması gerekliliği ortaya çıkmıştır. Test tahrifatı çoğunlukla, eğitimcilerin test tamamlandıktan sonra öğrenci yanıtlarını değiştirebileceği kâğıt kalem testleriyle ilişkilidir. Test tahrifatı durumunda resmi ya da özel test uygulamasının organizasyonunu yapan kurum, kuruluş ya da şirketlerin, görevi kötüye kullanma durumu olup olmadığını belirlemek için veriyi silme analizleri gibi bazı analizlere tabi tutması gerekmektedir (Wollack vd., 2015).

Test tahrifatının belirlenmesinde genellikle iki yaklaşımdan biri kullanılmaktadır. Bunlardan ilki, öğrenme düzeylerinde beklenmedik büyük artışa odaklanma ile

modellenmektedir. Eğitimin amacının bireylerin öğrenme düzeylerini artırmak olduğu göz önünde bulundurulduğunda, bu yaklaşım doğası gereği tahrifatın belirlenmesinde zorluk oluşturmaktadır. Dolayısıyla artışın öğrenmeden mi, yoksa usulsüzlükten mi kaynaklandığının ayırt edilmesi önemli hale gelmektedir. İkinci yaklaşım ise test tahrifatı sonucu öğrenme düzeylerinde gerçekleşen olağan dışı artışın büyük oranda yanlış yanıtların üzerinde yapılan silmeler ile gerçekleşmesine dayanmaktadır. Bu silmeler özellikle yanlış yanıtların doğru yanıtlarla değiştirildiği (yanlıştan doğruya) silmeler şeklindedir. Bu nedenle test tahrifatının belirlenmesindeki ikinci yaklaşımda, silmenin işaretçisi olan yanlış yanıt kağıdı üzerindeki ikincil izler incelenmekte ve analiz edilmektedir (Wollack vd., 2015).

Mroch, Lu, Huang ve Harris (2012) tarafından silme, test alan kişinin yanlış yanıt kağıdı üzerinde yanıtını bir ya da daha fazla kez silerek değiştirmesi olarak ifade edilmektedir. Ölçme Anonim Şirketi (Measurement Incorporated, MI) tarafından yayımlanan New Jersey Beceri ve Bilgi Durum Belirlemeleri (New Jersey Assessment of Skills and Knowledge, NJ ASK) Silme Analiz Raporu'nda ise silme, bir yanlış yanıt silindikten sonra yanlış yanıt kağıdı üzerinde kalan artık işaretleri ifade etmektedir (MI, 2012). Valilik Öğrenci Başarıları Ofisi (The Governor's Office of Student Achievement, GOSA), optik form üzerinde bir seçenek silindiğinde, aslında silmenin hiçbir zaman tamamen gerçekleşmediğini belirtmektedir (GOSA, 2016). Bu izler, bilgisayar programları kullanılarak analiz edilmekte ve silmenin ardından kalan iz, daha önceden tanımlanmış bir aralık içinde kalıyorsa, silme olarak nitelendirilmekte ve öğrencinin nihai yanlış yanıtı olarak tanımlanmamaktadır (MI, 2012).

van der Linden ve Jeon (2012) silme davranışını, tüm nihai yanlış yanıtlara iki aşamalı bir süreç yoluyla ulaşıldığını varsayarak modellemiştir. Birinci aşamada, her yanlış yanıtlanmıştır ve ikinci aşamada maddelere tekrar dönülerek yanlış yanıt ya değiştirilmiş (silme üretilmiş) ya da korunmuştur (silme üretilmemiştir). Buna bağlı olarak da silmelerin üç farklı türü tanımlanmıştır. Bunlar; doğru yanıtta yanlış yanıtla değiştirme, yanlış yanıtta başka bir yanlış yanıtla değiştirme ve yanlış yanıtta doğru yanıtla değiştirmedir. Şekil 1'de silmelerin işaretçisi olan ikincil izler ve ne anlam ifade ettikleri yer almaktadır.

A.Y. Madde	Seçenek			D Yorum	
	A	B	C		
A	1.	●	○	○	○ Kutucuğu tamamen kaplayan silme izi olmayan koyu renk işaretleme
A	2.	●	○	●	○ Yanlıştan Doğruya Silme (YD [Wrong to Right Erasure - WR])
D	3.	○	○	●	● Doğrudan Yanlışa Silme (DY [Right to Wrong Erasure - RW])
C	4.	○	●	○	● Yanlıştan Yanlışa Silme (YY [Wrong to Wrong Erasure - WW])
A	5.	●	○	○	○ Çok açık silme izi ile YD silme
D	6.	○	●	○	● Daha koyu silme izi ile YD silme
A	7.	●	●	○	○ Çift İşaretleme (Çİ [Double Grid – DG]) olarak anlaşılabilir YD silme
D	8.	○	●	○	● Gerçek Çift İşaretleme (Her iki kutucukta da aynı koyulukta ve kapsamda işaretleme)
B	9.	●	○	○	○ Yanıt anahtarını kısmen kaplayan YD silme
D	10.	○	⊖	○	● YD silme olarak yanlış kodlanabilecek tesadüfi işaret
C	11.	○	○	○	○ Ulaşılamayan ya da boş bırakılan madde

Şekil 1. Yanıt Kâğıtlarında Oluşabilecek İşaret ve Silme Örnekleri

Kaynak. Bishop ve Egan, 2017, 196.

Test tahrifatının temelinde yer alan varsayım yanıtların yanlıştan doğruya değiştirilmesi durumudur. Dolayısıyla tahrifatın belirlenmesi üzerinde çalışılırken şüpheyle yaklaşılması gereken silme türü yanlıştan doğruya doğru yapılan silmelerdir (van der Linden ve Jeon, 2012). Mead, Andersen ve Korts'a göre silmeler üç şekilde açıklanabilmektedir (2010; Akt: Mroch vd., 2012):

- Tekrar düşünme: Testi alan kişi başlangıçta verdiği yanıtı tekrar gözden geçirerek değişiklik yapar.
- Kaydırma: Testi alan kişi dikkatsizlik sonucu yanıtını başka bir maddenin yanıtı yerine işaretler, hatayı fark eder ve yanıtlarını sıralayarak doğru işaretleme yapar.
- Usulsüzlük: Kopyanın meydana geldiği durumdur. Bu şekilde meydana gelen beklenmedik ya da usulsüz silmelerin nedenlerini belirlemek, varlıklarını belirlemekten daha zordur.

Wollack ve diğerleri (2015) ise silmelerin iki farklı türünü tanımlamıştır. Bunlar, yasal olarak kabul edilen, yanıtın tekrar gözden geçirilerek değiştirildiği iyicil silmeler ve test bittikten sonra yanıtların test yöneticileri (öğretmen, gözetmen, idareci vb.) tarafından silinerek değiştirildiği hileli silmelerdir. İyicil silmeler; tesadüfi (random) silmeler, kaydırma (misalignment) nedenli silmeler ve dizi sonu (string-end) silmeler olarak

tanımlanmıştır. Tesadüfi silmeler, öğrencinin yanlışlıkla yanlış yanıt işaretlediği ve bunu hemen fark ederek yanıtını istediği yanıtla değiştirdiği ya da maddeyi bir yoldan çözerek yanıtladığı ancak gözden geçirirken yanıtını değiştirdiği durumu ifade etmektedir. Kaydırma nedenli silmeler, öğrencinin yanıt kâğıdı üzerinde yanlışlıkla bir maddeyi bir önceki ya da bir sonraki sıraya işaretlemesinin ardından işaretlemeye devam ettiği için, hatanın yapıldığı yerden itibaren tüm yanıtlarının yanlış olmasına neden olan durumu ifade etmektedir. Silme işlemi, öğrencinin hatasını fark ederek yanlışlık yaptığı yerden itibaren tüm yanıtlarını silerek doğru yerlere işaretleme yapması ile gerçekleşmektedir. Dizi sonu silmeler ise, öğrencinin zamanının azaldığını fark ederek, testin sonunda yer alan sorulara erişemeyeceğini anlaması ve sonda kalan tüm soruları, yanıtlanmamış soru kalmaması için önce tesadüfi olarak işaretlemesi, ancak zamanı kalırsa, çözdüğü maddeler için bulduğu yanıtla değiştirmesi sonucu oluşmaktadır.

Silmeler, testi alan kişinin yanıtları tek bir yönde (yanlıştan doğruya) olduğunda olağan dışı olarak tanımlanabilmektedir (Mroch vd., 2012). Öğrenciler bireysel olarak birçok nedenle silme işlemi yapabilmektedir ancak beklenmedik düzeyde yüksek bir oranda ve tüm sınıf için yanıtların yanlıştan doğruya doğru yapıldığı silmeler test tahrifatının göstergesi olarak değerlendirilebilir. Bu nedenle, silme analizleri sıklıkla test tahrifatının belirlenmesinde kullanışlı bir yaklaşım olarak kabul edilmektedir (Wollack vd., 2015).

Yanlıştan doğruya silmelerin, test tahrifatının belirlenmesi için kullanımına örnek olarak NJ ASK uygulamasının Silme Analizi Raporu (MI, 2012) verilebilir. Raporda, yanıt örüntülerindeki düzensizlikleri araştırmak için, tüm New Jersey okullarında yanlıştan doğruya doğru yapılan silmelerin ortalaması alınmış ve tüm okullar bu ortalama ile karşılaştırılmıştır. Ortalamanın iki standart sapma üstü eşik değer olarak kabul edilmiş ve eşik değeri aşan okullar belirlenmiştir. Ancak bu raporda silmelerin her zaman uygunsuzluğa bir kanıt oluşturmadığı da belirtilmektedir. Silmeler için bazı yasal nedenlerin varlığına da (test puanlarına verilen önem arttıkça silmelerde de artış gözlenmesi, öğrencilerin test alma deneyimleri arttıkça onlara bazı sistematik stratejiler öğretilmesi ve öğrencilerin yanıtlarını kontrol edip değişiklik yapmaları konusunda teşvik edilmesi vb.) değinilmektedir. Yüksek oranda silme işlemi yapıldığı belirlenen okullardan, bunun altında yatan nedenlerin belirlenmesi için ek incelemelerin istenebileceği belirtilmektedir. Söz konusu okul ve/veya sınıflarda öğrencilere, geri dönüp yanıtlarını değiştirmelerinin öğretilip öğretilmediği, öğrencilerin test alma stratejileri (örneğin öğrencinin iki yanıt arasında kararsız kalması ve son kararını

vermeden önce her iki yanıtı da işaretlemesi vb.) gibi öneriler sunulup sunulmadığının incelenmesi önerilmektedir. Wollack ve diğerleri (2015), silinmiş maddelerin tahrifatın göstergesi olmadığını, ancak tahrifatlı maddelerin silmeler sonucu oluştuğunu; dolayısıyla silinmiş maddelere verilen yanıtların geçerliğinin, özellikle sınavda çok fazla sayıda silinmiş madde varsa şüpheyle izlenmesi gerektiğini belirtmektedir.

Wollack ve Fremer (2013), kopyanın her türünün gerçek olduğunu ve tüm test uygulamalarında ortaya çıkma olasılığı bulunduğunu; test uygulayıcılarının kendi uygulamalarını kopyaya karşı korunaklı olarak görmemeleri gerektiğini ya da kendilerini kopyanın kendi uygulamalarında gerçekleşmediğine inandırmamaları gerektiğini belirtmektedir. Bu görüşlerine dayanak olarak da bu konuda yapılan araştırmaların sonuçlarını göstermekte ve yüksek riskli sınavlarda kopya bulunamamasını, büyük olasılıkla sonuçların bu anlamda yeteri kadar iyi kontrol edilmemiş olmasına bağlamaktadır.

Karar almada kullanılan testler önemli sonuçlar doğurma potansiyeline sahip olduğu zaman “yüksek riskli” olarak tanımlanmaktadır (Kane, 2002). Başaran (2005) yüksek riskli sınavları öğrenciler için özel akademik gerekliliklerin kazanılıp kazanılmadığının belirlenmesi, bir üst sınıfa veya bir üst öğrenime geçiş, diploma alma, diğer bir deyişle iş dünyasına veya yükseköğrenime geçiş ve hayat için gerekli bazı bilgi ve becerilerin kazanılması anlamında “bir şeyler ifade eden” sınavlar olarak tanımlamıştır. Türkiye’de uygulanan üniversite giriş sınavları yüksek riskli sınavlar çerçevesinde değerlendirilmektedir (Kumandaş, 2013). Türkiye’de yüksek riskli sınavlar çerçevesinde değerlendirilebilecek merkezi sınavlar Milli Eğitim Bakanlığı (MEB) ve Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) tarafından yapılmakta ve sonuçları değerlendirilmektedir.

Türkiye’de de bazı dönemlerde merkezi sınavlarla ilgili kopya iddiaları basında yer bulmaktadır. Dinç ve Akşit (2015), 1998-2013 yılları arasında ortaöğretime geçiş sisteminde yapılan değişikliklerin belli başlı ulusal gazetelerdeki köşe yazılarında nasıl ele alındığını saptamaya yönelik bir çalışma yürütmüşlerdir. Çalışmalarında köşe yazarlarının, köşe yazılarında ele aldıkları konular arasında “giriş sınavlarında kopya” konusunun bulunduğu görülmektedir. “Köşe yazarlarının tespit ettikleri sorunlar” başlığı altında “kopya çekildiği iddiaları”, “köşe yazarlarının eleştiri ve değerlendirmeleri” başlığı altında ise “kopya çekildiği iddialarının havada kalması, kanıtlanamaması” gibi bulgular elde edilmiştir.

ÖSYM tarafından sıklıkla, kopya iddialarının asılsızlığına ilişkin basın açıklamaları yapılmaktadır. Ancak, 2010 yılında uygulanan Kamu Personeli Seçme Sınavı (KPSS) Eğitim Bilimleri Testi'nin “sınav sürecinde bazı usulsüzlüklerin meydana geldiği kanaatine varıldığı” gerekçesiyle iptal edildiği belirtilirken (ÖSYM, 2010a), bir başka basın açıklamasında söz konusu sınavın “kopya olayı” nedeniyle iptal edildiği açıklanmıştır (ÖSYM, 2015). 2012 yılı mayıs ayında yapılan Avukatlar İçin Adli Yargı Sınavı'nın sonuçlarının istatistiki analizler ve adayların soru kitapçıkları üzerinde yapılan incelemeler sonucunda olağan dışı sonuçlar ürettiğinin görüldüğü ve sınavın ÖSYM Yönetim Kurulu kararı ile iptal edilerek yeniden yapılmak istendiği, ancak bu kararın idari yargı mercilerince iptal edilerek sınavın geçerli kılındığı belirtilmiştir (ÖSYM, 2015).

ÖSYM Mevzuatı Kanun ve Yönetmelikler (ÖSYM, 2014) dokümanı incelendiğinde kopya analizlerinin nasıl yapıldığına yönelik herhangi bir bilgiye rastlanmamıştır. ÖSYM tarafından yayımlanan Sınav Uygulama Yönergesi'nde (2016), adayların test sorularına verdikleri yanıtların dağılımlarının “bilgi işlem” yöntemleriyle inceleneceği, bu incelemelerden elde edilen bulguların bireysel ya da toplu olarak kopya çekildiğini göstermesi durumunda, kopya eylemine katılan adayların sınavının geçersiz sayılacağı belirtilmektedir. Ayrıca ÖSYM'nin resmi internet sitesi incelendiğinde 2007 Kamu Personeli Yabancı Dil Bilgisi Seviye Tespit Sınavı'nın (KPDS) Sınavda Uyulması Gereken Kurallar (ÖSYM, 2007a) ve Temel İlke ve Koşullar (ÖSYM, 2007b) dokümanlarında, “Adayların testlere verdikleri cevaplar bilgi işlem yöntemleriyle incelenmektedir. Bu incelemelerden elde edilen bulgular, ÖSYM tarafından değerlendirilecek, ikili veya toplu kopya çekildiği sonucuna varılan salonlarda, kopya eylemine karıştığı belirlenen adayların teste verdikleri cevaplar geçersiz sayılacaktır.” ifadeleri yer almaktadır. MEB Merkezi Sistem Sınav Yönergesi'nde (2011a) de cevap kâğıdının değerlendirmesi yapılırken sistemin ikili/toplu kopya taraması sonucunda kopya tespit edilmesi durumunda, adayın sınavının geçersiz sayılacağı belirtilmektedir.

Alanyazında sıklıkla kullanılan yanıt kopyalama analizlerinden Klasik Test Kuramı'na (KTK) dayalı olanların, K indeksi (Holland, 1996), K_1 ve K_2 indeksleri (Sotaridona ve Meijer, 2002), S_1 ve S_2 indeksleri (Sotaridona ve Meijer, 2003); Madde Tepki Kuramı'na (MTK) dayalı olanların ise Genelleştirilmiş Binominal Test (GBT) (van der Linden ve Sotaridona, 2006) ve ω indeksi (Wollack, 1996) olduğu görülmektedir. K indeksi, K_1 ve K_2 indeksleri, testi alan iki kişinin yanlış yanıtlarının uyumuna dayandırılarak; S_1 ve S_2 indeksleri, testi alan iki kişi arasındaki hem doğru hem de yanlış

yanıtların uyumunu temel alarak; ω indeksi ve GBT ise iki yanıt vektörü arasındaki benzerliği tüm yanıtları dikkate alarak hesaplamak üzere geliştirilmiştir (Zopluoğlu ve Davenport, 2012). ω indeksi ve GBT'nin "yüksek" belirleme oranları sağlarken, I. Tip hata oranını nominal seviyenin altında tuttuğu belirtilmiştir. K indeksi, K_1 ve K_2 indeksleri ile S_1 ve S_2 indekslerinin de "makul" belirleme oranları sağlarken, I. Tip hata oranını, yine nominal seviyenin altında tuttuğu ifade edilmiştir (Zopluoğlu, 2013). Yanıt kopyalama indeksleri bireylerin birbirinin yanıtlarını kopyaladığı durumda kullanılmakta ve kopyayı belirlemede etkili olmaktadır. Ancak öğretmen, test uygulayıcısı ya da test yöneticisi tarafından yapılan yanıt değişikliğini ortaya çıkarmak üzere bir belirleme yapmamaktadır.

Yapılan alanyazın taramasında Türkiye'de yanıt kopyalamanın belirlenmesi üzerine iki çalışmaya rastlanırken (Sünbül ve Yormaz, 2018; Yormaz, 2014), test tahrifatı ile ilgili hiçbir çalışmaya rastlanmamıştır. Yormaz (2014) yürüttüğü çalışmada, çoktan seçmeli maddelerden oluşan testlerde yanıt kopyalamanın belirlenmesinde kullanılan S_1 ve S_2 indeksleri ile Kappa istatistiğinin, farklı koşullar altında I. Tip hata oranlarını ve istatistiksel güçlerini ortaya koymayı amaçlamıştır. Çalışmada indekslerin, belirlenen α düzeylerine (0.0005, 0.001, 0.005, 0.01 ve 0.05) göre örneklem büyüklüğü, test uzunluğu ve kopya çekilen kaynağın yetenek düzeylerinin I. Tip hata oranlarına etkisi ve bu koşullara ek olarak kopya oranının kopya belirleme güçlerine etkisi incelenmiştir. Sünbül ve Yormaz'ın (2018) yürüttüğü çalışmada ise yanıt kopyalamanın belirlenmesinde kullanılan ω indeksinin, kaynak ve kopyacı bireyin yetenek düzeyi ile madde güçlük ve madde ayırt edicilik düzeylerine göre I. Tip hata oranları ve istatistiksel güç oranlarının farklı α düzeylerine göre (.05, .01, .001) değişimi incelenmiştir.

Test tahrifatının belirlenmesine yönelik yurtdışında yapılan çalışmalar incelendiğinde, Jacob ve Levitt'in (2003) test sonuçlarının öğretmen tarafından değiştirilmesini öğretmen kopyası olarak tanımlayan çalışmaları, öğretmen kopyasının sistematik olarak analiz edilmesine ilişkin ilk çalışma olarak ifade edildiği görülmektedir. Bu yöntem, çoktan seçmeli maddeler üzerinde olağandışı yanıt örüntülerinin, öğrencilerin yapılan değerlendirmelerde geçmiş ve gelecek yıllarda aldıkları puanlar ve öğrenci özellikleri (yaş, cinsiyet, ırk vb.) kullanılarak, çok kategorili lojistik regresyon analizi ile belirlenmesine dayanmaktadır. Yöntem temelde, benzer özelliklere sahip öğrencilerin olasılığına dayanarak yanıt seçimini değerlendirmektedir. Yapılan çalışmaya göre aynı öğretmenin öğrencileri için, yanıt kâğıtları üzerinde beklenmedik doğru yanıt

örüntülerinin varlığı, olası kopya durumu açısından dikkatle değerlendirilme gerekliliğini ortaya koymaktadır.

Qualls (2001) tarafından yürütülen bir başka araştırmanın temelini, yüksek riskli sınavlarda test tahrifatı olasılığının düşük riskli sınavlardakine oranla daha fazla olması beklentisi oluşturmaktadır. Bu nedenle normal sayılabilecek silme sayılarının belirlenmesi amacıyla düşük riskli bir sınavın optik formları incelenmiş ve silme davranışının normal sınırlarının belirlenmesi amaçlanmıştır. Araştırma sonuçları, öğrencilerin 40 maddelik testte ortalama bir yanıt değişikliği; %5'inden azının ise beş veya daha fazla yanıt değişikliği yaptığını göstermiştir. Öğrencilerin genel olarak testteki maddelerin %6-7'sinden daha azında yanıt değişikliği yaptıkları belirlenmiştir. Öğrencilerin testteki maddelerin %15'inden fazlasına verdikleri yanıtları değiştirmelerine ise nadir rastlandığı belirtilmiştir. Buradan hareketle silme sayılarının ortalama ve standart sapmalarına bağlı olarak bir güven aralığı belirlemenin makul olacağı ifade edilmiştir. Yanıt değişikliklerinin tümünün yanlıştan doğruya olmadığını, bir yanıt değişikliği için yanlıştan doğruya silme oranının %55'i aşmadığını, bu nedenle birden çok silme olması durumunda, tüm silmelerin yanlıştan doğruya olma olasılığının çok daha düşük olacağı belirtilen çalışmada, bu sınırları aşan yanlıştan doğruya silmeleri olan test alanların anormal olarak işaretlenmesi gerektiği vurgulanmıştır.

Mroch ve diğerleri (2012), silinmeleri analiz etmek için tanımlayıcı bir yaklaşım kullandıkları bir araştırma yapmışlardır. Araştırmada anormal silme davranışına sahip olan test alanların belirlenmesi için görece zayıf istatistiksel varsayımlar yapılmıştır. Çalışmanın amacı, büyük ölçekli bir değerlendirmeden elde edilen verileri kullanarak bireysel düzeydeki silme analizinin potansiyel yararlılığını incelemek ve bireysel silme modellerinin olası özellikleri hakkında bilgi sağlamaktır. Çalışmada öncelikle silme davranışının test alanların yetenek düzeylerine, madde güçlüğüne ve maddelerin yerine göre değişiklik gösterip göstermediği incelenmiştir. Daha sonra ise bireysel düzeyde silme analizi kullanılabilecek indeksler incelenmiştir. Bunun için yanlıştan doğruya silmelere dayanan üç silme indeksi hesaplanmıştır. Birinci indeks (WRT), yanlıştan doğruya silinmelerin toplam madde sayısına bölünmesiyle elde edilmektedir. İkinci indeks (WRC), yanlıştan doğruya silme sayısının doğru yanıt sayısına bölünmesiyle elde edilmektedir. Üçüncü indeks (WRE), yanlıştan doğruya silinme sayısının toplam silme sayısına bölünmesiyle elde edilmektedir. Silme indekslerinin test alanları işaretlemeye kullanılması için olağandışı veya anormal silmelere karşılık gelen kesme noktası, indeksin belirli bir değerinin dışına düşen test alan yüzdesine dayandırılarak,

arařtırmacılar tarafından test alanların %5'i olarak belirlenmiřtir. Arařtırma sonucunda, yanlıřtan dođruya silmelerin ve diđer silmelerin nadir gerekleřtiđi belirtilmiřtir. Kullanılan dort formdan birinde (diđerlerine gore daha kolay olan form) gerekleřen silme sayıları diđerlerine gore farklılık gostermiřtir. “Zor” maddelerin “kolay” maddelerden daha fazla yanlıřtan dođruya silme sayısı ierdiđi ifade edilmiřtir. Düşük yetenekli test alanların genellikle orta veya yüksek yetenekli test alanlara kıyasla gorece daha az silmeye sahip olduđu ancak farkın büyük olmadığı belirtilmiřtir. Maddenin testteki yeri ile silme sayıları arasında net bir iliřki kurulamamıřtır. Ü silme indeksi birbirleriyle yüksek düzeyde tutarlılık gostermiřtir. Bununla birlikte WRT indeksi ile WRC indeksi arasındaki tutarlılık daha yüksek bulunmuřtur.

van der Linden ve Jeon (2012), tahrifatın belirlenmesi üzerine yaptıkları alıřmada, bir olasılık modeli ve test istatistiđi olarak yanlıřtan dođruya silmelerin sayısı ile bundan türetilen istatistiksel bir yöntem önermektedirler. Arařtırmacılar MTK'ya dayanarak test alan j 'nin yetenek düzeyinin test boyunca sabit kaldıđı varsayımı altında İki Ařamalı Modelleme Yönteminde veri setine uyumlu olan 2PLM iin madde parametrelerini kullanmaktadır. Model nihai yanıtların dort olası kořulu (test alanın ilk yanıtının dođru ya da yanlıř olması kořuluna gore nihai yanıtın dođru ya da yanlıř olması) iin türetilmiřtir. İlk yanıtın yanlıř olması kořulunda nihai yanıtın dođru olması kořulunun olasılıđı, yanlıřtan dođruya silme olasılıđını vermektedir. Her birey iin, maddelerin bařlangıta yanlıř yanıtlanmış olması kořuluna bađlı olarak, yanlıřtan dođruya silmelerin sayısının gerek olasılık dađılımını üretmek iin genelleřtirilmiş binom dađılımı kullanılmaktadır. Bu yaklařımda, yanlıřtan dođruya silme sayısının belirlenen sınırları ařtıđı gözlenen test alanlar tahrifatlı olarak belirlenmektedir. Geliřtirilen yöntem, 2555 öđrencinin 65 maddeye yanıtlarından oluřan büyük ölekli 3. sınıf matematik deđerlendirmesinden elde edilen gerek bir veri setine uygulanmıřtır. .05, .01 ve .001 α seviyelerinde sırasıyla test alanların %2.62, %1.29 ve %0.47'si tahrifatlı olarak iřaretlenmiřtir. Bu iki ařamalı MTK analizinin etkililiđini I. Tip hata oranları ve istatistiksel gü aısından deđerlendiren bir simülasyon alıřması yapılmamıřtır.

Belov (2015), anormal yanıt deđerikliđi olan test alanları belirlemek iin kullanılan, nihai ve ilk yanıtlar arasındaki farka dayanan iki popüler istatistik olduđunu belirtmiřtir. Bunlardan ilki yanlıřtan dođruya yanıt deđeriklerinin sayısının tespit edilmesi; ikincisi ise yanlıřtan dođruya yanıt deđerikliklerinin tüm yanıt deđerikliklerine bölünmesidir. Daha yüksek sayıda yanlıřtan dođruya yanıt deđerikliđi daha yüksek olasılıkla anormalliđe iřaret etmektedir. Bununla birlikte yüksek sayıda yanlıřtan dođruya

yanıt deęişikliğine neden olabilecek teknolojik (bilgisayarların silmeleri yanlış taraması) veya insan kaynaklı (yanıt deęişikliğini test alanın yapması) “anormal olmayan” birçok nedeni olduğunu belirtmiştir (Wollack ve Maynes, 2014; Akt. Belov, 2015). Bu durum yanıt deęişikliklerinin belirsizliği olarak ifade edilmektedir. Bu istatistikler, belirsizlikten dolayı I. Tip hatanın artmasına neden olabilmektedir. Bu nedenle yanıt deęişikliği verilerindeki belirsizliğe olası bir çözüm, nihai ve ilk yanıtlar arasındaki farkı görmezden gelmek olduğunu belirtmektedir. Belov (2014; 2015) yaptığı çalışmalarda yanıt deęişikliği ile ilgili bilgileri, testte uygulanan maddeleri, yanıt deęişikliklerinin oluştuğu ve yanıt deęişikliklerinin oluşmadığı iki alt teste bölünmesi için kullanmaktadır. Yeni bir istatistik olarak tanımlanan D, belirsizlikten kaçınmak için, sadece nihai yanıtlar kullanıldığında bu alt testler arasındaki performans farkına dayanmaktadır. Bu yöntemin temelinde anormal test alanların, belirlenen bu iki alt test arasında benzer şekilde davranması gerektiği yatmaktadır. Örneğin, bir öğretmen düşük yetenekli bir test alanın bazı yanıtlarını düzeltirse, bu test alanın, ilk alt teste (yanıt deęişikliğinin olmadığı) kıyasla ikinci alt testte (yanıt deęişikliği yapılan) olağandışı bir şekilde çok sayıda doğru nihai yanıtının olması beklenmektedir. Burada alt testler arasındaki yetenek farkı, örtük özelliklerin Kullback-Leibler Ayrımı (KLD) ile ölçülmekte ve D istatistiği olarak ifade edilmektedir. D istatistiği ilk yanıtları göz ardı ettiği için belirsizliğe karşı hassas olmadığı, beklenmedik derecede büyük D değerinin, yanıt deęişikliği olan ve olmayan maddeler arasındaki beklenmedik büyük performans farkına işaret ettiği ifade edilmektedir. Buna karşın, yanlıştan doğruya silmelerin doğrudan analizinde, belirsizlik nedeniyle, büyük sayıda silmenin anormal yanıt deęişikliğine işaret edip etmediğinin belirsiz olduğu ifade edilmektedir. Belov (2015), simülatif veriler üzerinde yaptığı çalışmada istatistik D'nin, yanlıştan doğruya yanıt deęişikliklerine dayanan iki popüler istatistiğin aksine, belirsizliği giderme ve yanıt deęişikliklerini belirleme oranları için oldukça güçlü olduğu sonucuna ulaşmıştır. Bir alt testte yetersiz sayıda madde olması durumu ise D'nin genel bir sınırlılığı olarak belirtilmiştir.

Wollack ve diğerleri (2015), van der Linden and Jeon'a (2012) benzer bir yaklaşım izlemişlerdir. Simülatif veri üzerinde çalışarak silme örüntüsünü modellemek için MTK kullanmışlardır. Testi alan bir kişi için yetenek kestiriminin tahrifatlı maddeler içermesi durumunda, kestirilen yeteneğin yapay olarak yükseleceğini belirtmişlerdir. Bu nedenle, kullandıkları yöntemde, testi alan tüm bireyler için yetenek kestirimini, sadece tahrifatlı olmayan maddeleri kullanarak yapmışlardır. Daha sonra, yanlıştan doğruya silmelerin sayısını model altında beklenen sayıyla karşılaştırarak değerlendiren bir indeks

olan Silme Belirleme İndeksi'ni (SBİ [Ersasure Detection Index - EDI]) geliştirmişlerdir. SBİ süreklilik düzeltmesi için bir sabit ($C = -1/2$) eklenerek ya da eklenmeyerek kullanılabilir. 50 maddelik ve 250000 örneklem büyüklüğüne sahip veri setine öncelikle iyicil silmelerin üç türünü (kaydırma, dizi sonu ve tesadüfi) simüle etmişlerdir. Daha sonra iyicil silmeleri içeren veri setlerinde test alanların %0.4'ü tesadüfi olarak seçilerek ayrı ayrı sabit (5, 10 ve 15 silme) ve değişken tahrifatı simüle etmişlerdir. Sabit ve değişken tahrifatları içeren veri setlerini birleştirerek 1000000 örneklem büyüklüğüne sahip veri seti elde etmişlerdir. SBİ'nin süreklilik düzeltmesi uygulanan ve uygulanmayan formları için I. Tip hata oranlarını ve istatistiksel gücünü, 7 farklı α seviyesinde (.00001, .0001, .0005, .001, .005, .01 ve .05) ve 5 farklı yetenek düzeyine göre incelemişlerdir. Wollack ve diğerleri (2015) tarafından SBİ'nin uygulanan süreklilik düzeltmesiyle I. Tip hata oranını iyi kontrol ettiği ve küçük düzeydeki tahrifat miktarlarını yüksek güç ile belirlediği sonucuna ulaşılmıştır.

Qin (2016) tarafından yürütülen çalışmada, test tahrifatını belirlemek amacıyla çoktan seçmeli maddelerde silme analizi için EDI_WTR yöntemi, Diff₀ yöntemi ve silmelerin marjinal dağılımı (MD) yöntemi önerilmiştir. EDI_WTR yöntemi, yalnızca birden fazla yanlıştan doğruya silmeye sahip test alanlar için yanlıştan doğruya silmelere dayalı olarak SBİ'nin (EDI) düzenlenmiş halidir. Diff₀ yönteminde SBİ'nin yalnızca silinmemiş maddelere dayalı yetenek kestirimi ile silinmiş maddeleri de içeren tüm maddelere dayalı yetenek kestirimi arasındaki farkın normal dağılması gerektiği varsayımı yatmaktadır. Kestirilen yetenek düzeyleri arasındaki fark belirlenen manidarlık düzeyinde pozitif yüksek bir değer alıyorsa birey tahrifatlı olarak belirlenmektedir. MD, toplam veya yanlıştan doğruya silmelerin marjinal dağılımındaki aykırı değerlere odaklanmaktadır. Bu yöntemde, üçüncü çeyrekte en az 1.5 çeyrekler arası aralık (interquartiles ranges [IQRs]) büyük olan gözlemler aykırı değer olarak belirlenmektedir. Önerilen yöntemlerin sonuçları, mevcut yöntemler olan van der Linden ve Jeon'un (2012) iki aşamalı modelleme yöntemi (VJ) ve Wollack ve diğerlerinin (2015) SBİ (EDI) yöntemlerinin sonuçları ile karşılaştırılmıştır. Öncelikle bu yöntemler 45 maddelik ve 35280 örneklem büyüklüğünde gerçek veri seti üzerinden tahrifatın belirlenmesi amacıyla karşılaştırılmıştır. Daha sonra gerçek veri seti üzerinden, iyicil silmeler (tesadüfi, kaydırma ve dizi sonu silmeleri) simüle edilmiştir. İyicil silmeleri içeren veri setlerinde sabit tahrifat (5 ve 10 silme) ve değişken tahrifat yapılmıştır. Birey tahrifat yüzdesi (%25, %50, %100) ve okul tahrifat yüzdesi (%25, %50, %100) değiştirilerek,

farklı yöntemlerin I. Tip hata oranları ve güç oranları, farklı α seviyelerinde (.05, .01 ve .001) incelenmiştir. I. Tip hata oranları Bradley'in esnek referansına (1978) göre değerlendirilmiştir. I. Tip hata oranları referans aralığının içinde kalan koşullar için istatistiksel güç oranları raporlaştırılmıştır.

Qin'in (2016) araştırma bulgularına göre gerçek veri setinde tahrifatlı bireyler $Diff_{\theta}$ ile genellikle VJ, MD ve süreklilik düzeltmesi uygulanan EDI_WTR'ye göre daha az belirlenmiştir. Tahrifatlı bireylerin MD ile belirlenmesi, değişken tahrifat için $\alpha = .05$ düzeyinde süreklilik düzeltmesi uygulanan EDI_WTR'den daha düşüktür. MD için I. Tip hata oranlarının, simülasyon çalışmasında en küçük olduğu belirlenmiştir. VJ'nin tahrifatın bireysel olarak tespit edilmesi için en iyi yöntem olduğu belirtilmiştir. Simülasyon çalışmasında, hangi tahrifat stratejisi (sabit ya da değişken) kullanılırsa kullanılsın, test alanların ve okulların yüzdeleri ne olursa olsun, $\alpha = .05$ ve .01 düzeylerinde VJ'nin I. Tip hatayı kontrol ettiği ve en yüksek gücü (çoğunlukla .97'den daha yüksek) sağladığı ifade edilmiştir. Simülasyon çalışmasında SBİ, Wollack ve diğerlerinin (2015) de belirttiği gibi, süreklilik düzeltmesi uygulanmayan SBİ'nin I. Tip hata oranları yüksektir. Süreklilik düzeltmesi yapılmaksızın kullanılan SBİ'nin istatistiksel gücünün diğer istatistiklere göre en düşük olduğu belirlenmiştir. SBİ, tahrifatın bireysel düzeyde belirlenmesinde en zayıf belirlemeyi yapmakta ve $Diff_{\theta}$ 'ya benzer şekilde, çok sayıda yanlıştan doğruya silmesi olan ve ayrıca VJ ve MD tarafından işaretlenen bazı test alanları tespit edemediği ifade edilmiştir. Süreklilik düzeltmesi uygulanmayan EDI_WTR, $\alpha = .01$ düzeyinde I. Tip hata oranlarını kontrol etmiş, ancak hata oranı $\alpha = .05$ 'te alt sınırın altına düşmüştür. Simülasyon koşulları değiştiğinde I. Tip hata oranlarının net bir eğilimi gözlenmemiştir. I. Tip hata oranları kontrol edildiğinde, süreklilik düzeltmesi uygulanmayan EDI_WTR'nin gücü VJ'den düşüktür. Bu, EDI_WTR'nin gerçek verilerde VJ'ye oranla daha az şüpheli test alanı işaretlediği bulgusu ile tutarlı bulunmuştur. Süreklilik düzeltmesi uygulanan EDI_WTR $\alpha = .05$ düzeyinde I. Tip hatayı, süreklilik düzeltmesi uygulanmayan EDI_WTR'den daha iyi kontrol etmiş ve gücü daha yüksek bulunmuştur (.78'in üzerinde). Daha fazla sayıda madde, test alan ve okul tahrif edildiğinde güç azalmış ancak, sabit tahrifattan değişken tahrifata geçerken güç .97'nin üstüne çıkmıştır. $\alpha = .01$ ve .001 düzeyinde I. Tip hata oranı şişirilmiştir ve ortaya çıkan güç değerlendirilmemiştir. Süreklilik düzeltmesi uygulanan EDI_WTR'nin bireysel düzeyde tespit için kullanılabilir iyi bir seçenek olduğu ifade edilmiştir.

Sinharay ve Johnson (2016), test tahrifatını belirlemek amacıyla yanıt değişikliklerini analiz etmek için SBİ'den türetilen üç yöntem önermiştir. Önerilen yöntemler, sıfır hipotezinin normal dağıldığı varsayımı üzerine kurulmamıştır. SBİ'nin boş hipotezinin standart normal dağılım gösterdiğini varsaymak yerine, test tahrifatının olmadığı boş hipotezi test etmek için kesin olasılık hesaplamaları kullanılmıştır. Yöntemlerden, Genelleştirilmiş Binom Yöntemi (Generalized Binomial Method [GBM]) ve Kesin Yöntem (Exact Method [EM]), kesin olasılık hesaplamalarını içerirken Sonsal Yordayıcı Model Kontrolü (Posterior Predictive Model Checking [PPMC]) yönteminde bayesian yaklaşımı izlenmiştir. GBM Yöntemi'nde SBİ'nin normal dağılım izlediğini varsaymak yerine, van der Linden ve Sotaridona (2006) ve van der Linden ve Jeon (2012)'de olduğu gibi genelleştirilmiş binom modeli kullanılmakta; ancak buradaki gibi iki aşamalı MTK modeline uyumu gerektirmemektedir. Daha fazla hesaplama yoğunluğu olan EM'de test alanın belirli maddeler için puan kalıpları olasılığının, diğer maddelerde gözlenen puan kalıplarının olasılığından çok yüksek olduğu durumlar için p değerinin hesaplanmasına dayanmaktadır. PPMC'nde ise SBİ'nin boş hipotez dağılımının standart normal dağılım olduğunu varsaymak yerine, sonsal kestirim dağılımına (posterior predictive distribution) yaklaştırılmıştır. Sinharay ve Johnson (2016) SBİ ile bu üç yöntemin karşılaştırdıkları bu çalışmada hem I. Tip hata oranları hem de istatistiksel güç göz önüne alındığında, EM veya PPMC'nin diğerlerinden daha iyi olduğu sonucuna ulaşmışlardır. Bu iki yöntem arasında, EM'nin, I. Tip hata oranlarının daha büyük olduğunu ve PPMC'ye kıyasla gücünün daha yüksek olduğunu belirtmişlerdir. Ancak SBİ'nin uygulanmasının daha kolay olduğunu ve uygulama kolaylığı istendiğinde tercih edilmesi gerektiğini vurgulamışlardır.

Sinharay, Duong ve Wood (2017) yürüttükleri çalışmada, Olabilirlik Oran Testine (OOT [Likelihood Ratio Test – LRT]) dayanan yeni bir istatistik (L -indeks) önermektedir. L -indeks'i, anormal yanıt değişikliğinin olmadığı sıfır hipotezi altında asimtotik bir standart normal dağılıma sahiptir. Yanıt değişikliği yapılan maddeler üzerinden kestirilen yetenek düzeyinin yanıt değişikliği yapılmayan maddeler üzerinden kestirilen yetenek düzeyinden daha büyük olması durumunda, büyük bir değer olan L -indeksi, sıfır hipotezinin reddedilmesine ve anormal yanıt değişiklikleri için test alanın tahrifatlı olarak işaretlenmesine yol açmaktadır. Sinharay ve diğerlerinin yaptıkları simlasyon çalışmasında, L -indeks'in I. Tip hata oranının nominal seviyeye çok yakın olduğu ve istatistiksel gücün, anormal yanıt değişikliklerini tespit etmek için mevcut bazı istatistiklerinkine kıyasla tatmin edici olduğu ifade edilmiştir. L -indeks'in I. Tip hata

oranlarını ve istatistiksel gücünü, D -indeksi D_1 , D_2 , D_3 istatistikleri (Belov, 2015), SBİ (EDI; Wollack vd., 2015); D^* istatistiği (Cubbellotti ve Belov, 2015); sonsal değişiklik istatistiği (PSS; Belov, 2016) ve Optimal İndeks (OI) istatistikleri ile farklı koşullar altında karşılaştırmıştır. Araştırmacılar 100 madde ve 100000 örneklem büyüklüğü için veri üretmiş ve yetenek düzeylerini 3PLM'ye göre kestirmiştir. İyicil silmelerin üç türünü de (kaydırma, dizi sonu ve tesadüfi) veri setinde simüle ettikten sonra test alanların %5'i 4 ve 8 maddede tahrifata uğratılmıştır. Süreklilik düzeltilmesi uygulanan SBİ, D_1 ve L -indeks'in I. Tip hata oranlarını ve istatistiksel gücünü .05, .01 ve .001 α düzeylerinde raporlaştırmıştır. SBİ için I. Tip hata oranlarının üç α düzeyinde de kontrol edildiği görülmüştür. 4 silme için de istatistiksel gücün belirlenen alfa düzeylerinde kabul edilebilir düzeyde olmadığı; 8 silme için ise yalnızca .05 düzeyinde kabul edilebilir istatistiksel güç elde edildiği, ayrıca silme sayısı arttıkça istatistiksel gücün de arttığı görülmektedir. Geliştirilen L -indeks'in tüm α düzeylerinde hata kontrolü sağladığı ve SBİ'den az miktarda daha iyi olmak üzere benzer istatistiksel güç elde edildiği görülmektedir.

Wollack ve Eckerly (2017), van der Linden ve Jeon'un (2012) İki Aşamalı Modelleme Yöntemi ile Wollack ve diğerleri (2015) tarafından geliştirilen SBİ'nin, tahrifatın belirlenmesinin bilimsel zorluğu açısından önemli adımlar olmasına rağmen, test alanları bireysel olarak ve alışılmadık derecede yüksek silme sayıları ile belirlemek için tasarlanmış olmalarını önemli bir sınırlılık olarak ifade etmişlerdir. Belirli bir öğrencinin silme modeli ne kadar sıradışı olursa olsun, pratikte okulların, şehirlerin eyaletlerin öğrencileri bireysel olarak belirlemekle ilgilenmelerinin pek mümkün olmadığını belirtmişlerdir. Bunun yerine, test tahrifatında bir hipotez, eğitimcilerin birden fazla öğrenci için yanıt formlarını değiştirdikleri yönündedir. Bu nedenle SBİ'yi, okul veya bölge düzeyinde tahrifatı belirlemek üzerine EDI_g olarak yeniden düzenlemişlerdir. EDI_g'nin grup düzeyi analizi için test alanların sınıf, okul ve bölge bilgilerinin önceden bilinmesi gerekmektedir. Yaptıkları simülasyon çalışmasında iyicil silmelerin üç türü (tesadüfi, dizi sonu ve kaydırma) ve grup düzeyinde hileli silmeler simüle edilmiştir. Hileli silmeler için her okulda tahrifata uğramış sınıfların dört farklı oranı (%0, %33, %67 ve %100) simüle edilmiş ve EDI_g'nin farklı alfa düzeylerinde I. Tip hata oranı ve gücü incelenmiştir. Araştırma sonucunda tüm koşullar için tüm alfa düzeylerinde I. Tip hata oranlarının kontrol edildiği görülmüştür. EDI_g'nin gücünün tahrifatlı madde sayısı ve tahrifatlı birey sayısı arttıkça arttığı görülmüştür. 25 kişilik bir sınıfta yalnızca bir tahrifat mağduru olduğu durumda EDI_g'nin tahrifatlı sınıfları belirleme gücünün düşük,

en az üç tahrifat mağduru olduğu durumda orta ve beş veya daha fazla tahrifat mağduru olduğu durumda yüksek olduğu belirtilmiştir. 10 tahrifatlı maddeye sahip beş tahrifat kurbanı olduğu durumda güç, .90 olarak belirlenmiştir. Okul düzeyinde tahrifatın belirlenmesinde de sınıf düzeyindeki benzer bir sonuç bulunmuştur. Tahrifatlı sınıf sayısı arttıkça EDI_g'nin gücünün arttığı görülmüştür.

İlgili araştırmalar bir bütün olarak değerlendirildiğinde tahrifatın belirlenmesinde kullanılan yöntemlerin birbirlerine göre bazı üstün ve zayıf yönleri olduğu görülmektedir. Örneğin, van der Linden ve Jeon (2012), modern optik okuyucular ile silme analizlerinin yapılabildiğini; herhangi bir usülsüzlüğün olmadığı durumlarda da testi alanların bazen beklenmedik düzeyde silme yapabildiği ya da silme örüntüleri üretebildiğini, yasal ve yasal olmayan davranışlar arasında istatistiksel ayırım yapmanın tek olası yolunun, maddeler üzerindeki yasal silmelerin olasılıklarına yönelik olarak tatmin edici bir modele sahip olmak ve yanıt kâğıtlarını bu olasılıklar altında değerlendirmek olduğunu söyleyerek Jacob ve Levitt'in (2003) yöntemini eleştirmektedir. Jacob and Levitt (2003) yaklaşımı test alanların önceki ve sonraki puanları ile geçmiş bilgilerinin kullanımını gerektirmekte ve ayrıca Qualls (2001) ile Mroch ve diğerlerinde (2012) de olduğu gibi istatistiksel olasılık dağılımını içermemektedir. van der Linden ve Jeon'un (2012) ise yöntemlerinde genelleştirilmiş binom dağılımını kullanması hesaplama gücünün doğrudur. Sinharay ve Johnson (2016) ile Qin'in (2016) araştırmalarında SBİ ve SBİ'ye dayalı olarak türettikleri istatistikleri kullandıkları görülmüştür. Wollack ve diğerleri (2015) tarafından geliştirilen SBİ'nin istatistiksel olasılık dağılımı içermesi, simülasyon verisi ile çalışmaya olanak sağlaması, hesaplama kolaylığı sağlaması, alanyazında I. Tip hatanın kontrol edildiğine ve istatistiksel gücün yüksek olduğuna ilişkin bulguların yer alması nedeniyle tahrifatın belirlenmesinde izlenebilecek diğer yöntemlere göre daha avantajlı olduğu düşünülerek test tahrifatının belirlenmesi üzerine yürütülen bu çalışmada söz konusu yaklaşımın kullanılması uygun bulunmuştur.

Öğrenci yanıtlarının silinerek değiştirilmesiyle gerçekleştirilen test tahrifatı tüm değişikliklerin yanlıştan doğruya olmasını gerektirdiğinden, test tahrifatının puanların geçerliğini yanıt kopyalamadan daha çok etkilemesi beklenmektedir. ÖSYM tarafından uygulanan Sınav Uygulama Yönergesi'nde (ÖSYM, 2016), test tahrifatına işaret eden “sınav sonuç verilerini gerçeğe aykırı bir şekilde değiştirenler, sınav sonuçlarını adayın lehine veya aleyhine değiştirenler hakkında yasal işlemler yapılacaktır” ifadesi bulunmakta ancak sınav sonuçlarının analizinde hileli silmeler ve tahrifata ilişkin ayrıca bir analiz yapıldığına ilişkin bilgi yer almamaktadır. Yüksek riskli sınavların yapıları

gereği sınav sonuçları, bireylerin hayatında önemli kararların alınmasında kullanılmaktadır. Basında yer alan kopya haberleri kamuoyunda ÖSYM ve MEB tarafından yapılan yüksek riskli sınavların objektifliğine olan güvenin azalmasına neden olmaktadır. Türkiye’deki yüksek riskli sınavlarda tahrifat analizlerinin yapılmasının hem test puanlarının geçerlik ve güvenilirliğine ilişkin daha fazla kanıt sunulmasında hem de kamuoyunda sınav sonuçlarına olan güveninin artırılmasında etkili olabileceği düşünülmektedir. Araştırmalara dayanarak, test tahrifatının belirlenmesine yönelik çalışmaların alanyazında eksiklik olduğu tespit edilmiştir. Ayrıca alanyazında SBİ ile test tahrifatının belirlenmesinde örneklem büyüklüğünün ve madde sayısının değiştiği bir çalışmaya rastlanmamıştır. Yapılan tartışmalar doğrultusunda, çoktan seçmeli testlerde bireylerin yanıtları üzerinde değişiklik yapılmasıyla test puanlarının geçerliğini doğrudan etkileyen test tahrifatını belirlemek üzere kullanılan SBİ’nin, örneklem büyüklüğü, madde sayısı, hileli silme oranı ve yetenek düzeylerine göre, tahrifatlı bireyleri tespit etmedeki performansının incelenmesi bir gereklilik olarak görülmüştür.

Amaç

Bu araştırmanın genel amacı, test tahrifatını belirlemede kullanılan indekslerden biri olan SBİ’nin, simülatif olarak üretilen çoktan seçmeli test verileri üzerinde farklı koşullar (örneklem büyüklüğü, madde sayısı, hileli silme oranı ve yetenek düzeyi) altında I. Tip hata oranı ve istatistiksel güç bakımından nasıl değiştiğinin belirlenmesidir. Bu genel amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

Belirlenen α düzeylerinde (.05, .01, .005, .001, .0005, .0001 ve .00001) süreklilik düzeltmesi uygulanan ve uygulanmayan durumlarda SBİ’nin ürettiği,

1. I. Tip hata oranları,
 - a. farklı madde sayılarında (50 ve 160),
 - b. farklı örneklem büyüklüklerinde (500, 5000, 50000 ve 99990),
 - c. farklı yetenek düzeylerinde
2. İstatistiksel güç,
 - a. farklı madde sayılarında (50 ve 160),
 - b. farklı örneklem büyüklüklerinde (500, 5000, 50000 ve 99990),
 - c. farklı yetenek düzeylerinde,

d. farklı sabit tahrifat oranlarında (%10, %20 ve %30) ve deęişken tahrifatta nasıl deęişmektedir?

Önem

Yapılan arařtırmalar, eęitim alanında kullanılan yüksek riskli sınavlarda sınav güvenlięi problemlerine iřaret etmektedir. Belirlenmiř amaçlar için test puanlarının geçerlięinin garanti edilememesi önemli riskler taşımaktadır.

Test puanlarının doęruluęuna iliřkin yeterli kanıtın saęlanamaması, sonuçların geçerlięini de řüpheli hale getirmektedir. Dürüst olarak elde edilen puanların, hileli olarak elde edilenlerden ayırt edilememesi, vasıfsız adayların lisans ve sertifikalar elde etmeleriyle; üniversitelere ya da işlere yerleřtirilmeleriyle, dolayısıyla sınırlı sayıdaki istihdam olanaęının daha niteliksiz bireylerle doldurulması ile sonuçlanmaktadır. Bu durum, dürüst bir řekilde sınava katılan bireyler için adaletin saęlanamaması anlamına gelmektedir (Wollack ve Fremer, 2013).

Yurtdıřında test organizasyonu řirketleri, test usulsüzlüęü riskinin farkında olup, güvenlik ihlallerini önlemek ve usulsüzlüęü belirlemek için kaynak ayırmaktadırlar. Yanıt kopyalama analizleri ve silme sayıları analizleri gibi analizlerin hepsi, test güvenlięinin ihlalinin iřaretlerini göstermek üzere, test anormalliklerinin arařtırılması amacıyla rutin olarak test řirketleri tarafından yapılmaktadır (GOSA, 2016; MI, 2012). Türkiye’de de merkezi sınavları yürüten ÖSYM ve MEB’de yanıt kopyalama analizleri yapıldıęı belirtilmekte, ancak bu analizlerin nasıl yapıldıęı ile ilgili bilgiler paylařılmamaktadır. MEB, bazı sınavların ardından yayımlanan raporlarda “ikili ve çoklu kopya” nedeniyle sınavları geçersiz sayılan birey sayısı hakkında bilgi sunarken (MEB, 2011b), ÖSYM de bazı sınavlara iliřkin “tespit edilen kopya sayısı” adı altında bilgilere yer vermekte (ÖSYM, 2007c; ÖSYM, 2010b) ancak silme analizleri ve dolayısıyla test tahrifatı ile ilgili analizlere iliřkin bilgiler bulunmamaktadır.

Yapılan alanyazın taramasında Türkiye’de test tahrifatı üzerine yapılmıř bir çalıřmaya rastlanmamıřtır. Alanyazında eksiklik olarak belirlenen test tahrifatı üzerine yapılmıř olan bu çalıřmanın özgün deęeri ile test tahrifatı riskine dikkat çekilerek, tahrifat analizlerinin yüksek riskli sınavlarda uygulanabilirlięinin belirlenmesi ve bu yönüyle de çalıřmanın hem alanyazına, hem de uygulamaya katkı saęlaması beklenmektedir.

Sınavların türü, formatı ve nasıl yürütüldüęüne baęlı olmaksızın, tüm sınavlar güvenlik ihlallerine karřı hassastır. Özellikle bireylerin gelecekleri ile ilgili önemli

kararların alınmasında kullanılan yüksek riskli sınav uygulamalarında, sınav güvenliğinin sağlanması, hem test puanlarının geçerliğine ilişkin kanıtların ortaya konması hem de kamuoyunda testlerin objektif değerlendirilmesi ve sonuçlarına olan güvenin artması açısından önemlidir. Sınav öncesi ve sınav sırasında alınan güvenlik önlemlerinin yanı sıra sınav sonrasında da test verilerinin güvenliğinin sağlanması önemlidir. Bu nedenle sınav uygulamalarında güvenlik ihlallerinin yapılıp yapılmadığının belirlenmesi ve bu bilgilerin kamuoyu ile paylaşılması gerekmektedir. Tahrifat analizlerinin sonuçları, tahrifatın yapıldığının kesin kanıtını oluşturmamakla birlikte, şüpheli durumları ortaya çıkarmaktadır. Bu analizlerin şüpheli durumları ortaya koymada bir dayanak olduğu göz önünde bulundurulduğunda, bu çalışma sayesinde ulusal sınavlarda kopya analizlerinin yanı sıra tahrifat analizleri de yapılarak sonuçların raporlaştırılabileceği ve bu raporların olası şüpheli durumların incelenmesine dayanak oluşturabileceği düşünülmektedir. Yürütülen çalışmada alanyazındaki diğer çalışmalardan farklı olarak, farklı örneklem büyüklükleri, madde sayıları ve hileli silme oranlarının tahrifatın belirlenmesindeki rolünün incelenmesiyle, çeşitli koşullar altında karşılaştırma yapma olanağı sağlaması açısından da alanyazına katkı sağlaması beklenmektedir. Bu çalışma gerektiği durumlarda ayrıntılı incelemelere temel oluşturacak nitelikte çalışmaların planlanmasına öncülük etmesi açısından da önemli görülmektedir.

Sayıtlılar

Araştırmanın verileri, düzeltme formülünün kullanılmadığı beş seçenekli çoktan seçmeli bir sınav uygulamasında, öğrencilerin tüm maddeleri yanıtladıkları varsayılarak, yanıtlanmamış madde ya da kayıp veri olmayacak şekilde üretilmiştir.

Sınırlılıklar

Araştırma, beş seçenekli çoktan seçmeli maddelere ilişkin üretilen veriler ile sınırlandırılmıştır. Verilerin üretilmesinin ve analizinin zaman alıcı olması, kullanılan bilgisayar programlarında sorun yaşanmaması ve bulguların raporlaştırılma sürecinin dikkate alınması gibi gerekçelerle, örneklem büyüklüğü (500, 5000, 50000 ve 99990), madde sayısı (50 ve 160) ve hileli silme oranı koşulları (%10, %20, %30 ve değişken) ile sınırlı tutulmuştur. Ayrıca MULTILOG programının ulaşılan sürümü ticari sürüm olmadığından analizlerde örneklem büyüklüğü üst sınırı 99999 olarak belirlenmiştir.

Tanımlar

Test Alan: Bir test uygulamasına katılan ya da simülasyon ile test maddelerine ilişkin yanıtları üretilen bireydir.

Kopya: Sahtekârlık yoluyla kişinin kendisi ya da bir başkası için gerçekte sahip olmadığı bilgiye sahip olduğunu gösterme girişimidir.

Tahrifat/Hileli Silme: Bir testte bireyin yanlış yanıtlarının test uygulayıcısı/yönetici/öğretmen vb. tarafından doğru yanıtla değiştirilmesidir. Simülasyon verisi için ise verilerin üretilmesinin ardından yanlış yanıtların araştırmacı tarafından doğru yanıtlar ile değiştirilmesi şeklinde gerçekleştirilen işlemdir.

İyicil Silme: Test alan tarafından sınav esnasında fikir değiştirme, hatasını farketme vb. nedenlerle yapılan yanıt değiştirilmedir.

Silme Sayısı: Değiştirilen yanıt sayısıdır.

Kesme Puanı: Bireyin başarılı ve başarısız olarak sınıflandırılmasında (ayrıt edilmesinde) kullanılan yetenek düzeyine göre belirlenen gerçek puandır.

Yetenek Düzeyi: MTK'ya göre kestirilen yetenek düzeylerine göre küçükten büyüğe sıralanan test alanların, her birinde eşit sayıda test alan olmak üzere beş gruba bölünmesinden oluşan gruplardır.

Silme Belirleme İndeksi: Tahrifata uğrayan bireylerin belirlenmesi amacıyla geliştirilmiş olan indekstir.

BÖLÜM 2

YÖNTEM

Bu bölümde araştırma modeli, simülasyon deseni, verilerin üretilmesi, verilerin silinme simülasyonu ve verilerin çözümlenmesi ile ilgili açıklamalara yer verilmiştir.

Araştırma Modeli

Araştırmada üretilen simülatif veriler üzerinden, örneklem büyüklüğü, madde sayısı, yetenek düzeyi ve tahrifat oranı gibi çeşitli koşullar altında, test yanıtlarında tahrifat yapılan bireylerin belirlenmesinde kullanılan SBİ'nin etkililiği, I. Tip hata oranları ve istatistiksel güç değerleri aracılığıyla incelenmektedir. Araştırmada, söz konusu indeksin hangi koşullar altında daha düşük I. Tip hataya ve daha yüksek istatistiksel güce sahip olduğu hakkında bilgi edinilerek var olan bilgiye yenilerini katmak amaçlandığı için temel araştırma niteliğindedir (Karasar, 2007).

Simülasyon Deseni

Bu başlık altında verilerin üretilmesi ve yanıtların silinmesinin hangi koşullar altında yapıldığı ele alınmıştır. Sabit tutulan ve değişimlenen koşullar aşağıda açıklanmıştır.

Sabit Koşullar

Veri Üretmede Kullanılan Model. Bu çalışmada belirlenen yetenek dağılımına ilişkin yanıt örüntüleri, çok kategorili puanlama modellerinden Sınıflamalı Tepki Modeli (STM [Nominal Response Model - NRM]) kullanılarak elde edilmiştir. STM, yetenek kestirimi sırasında seçeneklerin sıralı olmasını gerektirmeyen maddelerde, madde tepkilerinin betimlenmesinde kullanılabilen bir MTK modelidir. STM'nin temel özelliği, çoktan seçmeli test maddelerinde çeldiricilerin madde karakteristik eğrilerinin çizilmesine izin vermesidir (Embretson ve Reise, 2000). Bu modelin amacı, yalnızca maddelerin doğru yanıtlanıp yanıtlanmadığına bağlı olarak değil, test alan bireylerin yanıtlarında yer alan tüm bilgileri

kullanarak yetenek kestiriminin hassasiyetini en üst düzeye çıkarmaktır (Hambleton, Swaminathan ve Rogers, 1991).

Örneklem Dağılımı. Yetenek dağılımları ortalaması 0, standart sapması 1 olan $N(0,1)$ standart normal dağılıma uygun şekilde üretilmiştir.

Hileli ve İyicil Silme Yapılan Test Alan Sayısının Oranı. İyicil silme yapılan test alan oranı Wollack ve diğerlerine (2015) benzer şekilde dizi sonu silmeleri için %0.4, kaydırma silmeleri için %0.4 ve tesadüfi silmeler için %63 olarak belirlenmiştir. Sabit silmelerin her koşulu (%10, %20 ve %30) ve değişken silme için, tüm test alanların içerisinde tahrifata uğrayan test alan sayısının oranı %2 olmak üzere toplamda %8 olarak belirlenmiştir. Örneklem sayısının az olduğu durumlarda her yetenek düzeyinden aynı sayıda olacak şekilde silme yapılmıştır. Silmelerin simülasyonu veri üretimi gerektirdiği için hileli ve iyicil silme yapılan test alan sayısının oranlarına “Verilerin Silinme Simülasyonu” başlığı altında ayrıntılı olarak yer verilmiştir.

Manipüle Edilen Koşullar

Madde Sayısı. Madde sayısı koşulu 50 ve 160 olmak üzere iki düzey olarak belirlenmiştir. 50 madde sayısı, indeksi geliştiren Wollack ve diğerlerinin (2015) yürüttüğü çalışmada kullanılan madde sayısı olduğu ve araştırma bulgularını karşılaştırmaya olanak sağlaması için seçilmiştir. Ayrıca, Türkiye’deki merkezi sınavlarda tek oturumda kullanılan testlerin, tek optik formda uygulanan madde sayıları incelenmiştir. MEB tarafından uygulanan Temel Eğitimden Ortaöğretime Geçiş (TEOG) sınavlarında 20, Sınavla Öğrenci Alacak Ortaöğretim Kurumlarına İlişkin Merkezî Sınavda sayısal alanda 40, sözel alanda 50; ÖSYM tarafından uygulanan Lisans Yerleştirme Sınavı’nda (LYS-1) ve Yabancı Dil Sınavı’nda (YDS) 80; KPSS Genel Yetenek-Genel Kültür Sınavı’nda, Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı’nda (ALES) ve Dikey Geçiş Sınavı’nda (DGS) 120; Yükseköğretime Geçiş Sınavı’nda (YGS) ise 160 madde tek oturumda ve tek optik formda uygulanmaktadır. Madde sayısı değiştiğinde, indeksin istatistiksel gücü ve I. Tip hatasında oluşacak farkın net olarak görülebilmesi için Türkiye’de merkezi sınavlarda kullanılan en yüksek madde sayısı olan 160 bir diğer düzey olarak seçilmiştir.

Örneklem Büyüklüğü. Örneklem büyüklüğü koşulu 500, 5000, 50000 ve 99990 örneklem büyüklükleri olmak üzere dört düzey olarak belirlenmiştir. Bu örneklem büyüklükleri, örneklem büyüklüklerine bağlı olarak SBI'nin etkililiğini gözlemlemeye olanak sağlayacak şekilde belirlenmeye çalışılmıştır. Wollack ve Cohen (1998), SBI ile benzer yapıdaki yanıt kopyalama indeksi ω 'nın gücünün, örneklem küçük olduğunda (örneğin, N=100) az miktarda düştüğünü, ancak örneklem büyüklüğü 500 civarında olduğunda hemen hemen hiç değişmediğini belirtmiştir. Eyalet çapında bir silme analizinin örneklem büyüklüğünün her zaman 500'den büyük olacağını kabul ederek I. Tip hata oranını ve gücü etkilememesini beklediklerini ifade etmişlerdir. Benzer şekilde Wollack ve diğerleri (2015), eyaletteki tüm okullar düşünüldüğünde, düşük oranda okulda, düşük oranda öğretmenin düşük oranda öğrenci için az miktarda yanıtı tahrifata uğratmış olabileceklerini belirterek, tahrifat boyutunun az olmasını beklediklerini ifade etmişlerdir. Ancak yapılan bu çalışmada farklı örneklem büyüklüklerinde indeksin etkililiğinin belirlenmesi de amaçlanmaktadır. Bu nedenle 500 ve 5000 örneklem büyüklükleri küçük ve büyük örneklemli okullar düzeyinde yapılacak olan çalışmalar için; 50000 örneklem büyüklüğü ise il düzeyinde yapılacak olan çalışmalar için ortalama örneklem büyüklükleri olarak düşünülmüştür.

Geniş ölçekli uygulamalarda gerçek durumu temsil etmesi için 1000000 örneklem büyüklüğü üzerinde çalışılması planlanmıştır; ancak MULTILOG programının ulaşılan sürümü ticari sürüm olmadığından analizlerde örneklem büyüklüğü sınırı 99999 olarak belirlenmiştir. Akademik sürüme erişilememesi nedeniyle en büyük örneklem, beş yetenek düzeyine eşit sayıda bölünebilecek olan 99990 olarak belirlenmiştir ve bu durum araştırmanın bir sınırlılığıdır. Daha küçük örneklemelerde de testin gücü ve I. Tip hatanın incelenmesi amacıyla 100 örneklem büyüklüğü üzerinde de çalışılmak istenmiş ancak 100 örneklem için üretilen verilerde MTK'nın temel varsayımlarından olan yerel bağımsızlık varsayımı sağlanamadığı için en küçük örneklem büyüklüğü olarak 500 kabul edilmiştir.

Tahrifat Türü ve Oranı. Sabit ve değişken tahrifat yapılmıştır. Sabit tahrifat kendi içerisinde madde sayısının %10'u, %20'si ve %30'u olacak şekilde değişimlenmiştir. Değişken tahrifat ise yapısı gereği test alanlarda farklı sayılarda silme ile sonuçlandığı için kesme puanı sabit tutularak değişimlenmemiştir. Silmelerin simülasyonu veri üretimi gerektirdiği için tahrifat oranlarına "Verilerin Silinme Simülasyonu" başlığı altında ayrıntılı olarak yer verilmiştir.

Yetenek Düzeyi. Yetenek düzeyi yüksek olan gruplarda gözlenen doğru yanıt sayısı yüksek olacağından hedeflenen oranda hileli silme işlemi yapılamaması durumu söz konusudur. Toplam test alanlar üzerinden genel bir I. Tip hata oranı ve istatistiksel güç yorumu yapmanın bulguların doğru değerlendirilememesine neden olabileceği düşünüldüğünden, bulgular yorumlanırken yetenek düzeylerine göre ayırım yapılması uygun görülmüştür. Bu nedenle test alanlar kestirilen yetenek düzeylerine göre sıralandıktan sonra her grupta eşit sayıda test alan olmak üzere beş eşit parçaya (quintile) ayrılarak yetenek düzeyleri oluşturulmuştur. Analizler, test alanlar yetenek düzeylerine göre ayrılmadan ve ayrılarak gerçekleştirilmiştir.

Alfa Seviyesi. SBİ'nin istatistiksel gücü ve I. Tip hata oranı, SBİ'nin hangi hata oranlarında güvenle kullanılabilmesinin belirlenmesi amacıyla .000001, .0001, .0005, .001, .005, .01 ve .05 olmak üzere yedi farklı α seviyesinde incelenmiştir.

Süreklilik Düzeltmesi. SBİ ile tahrifatlı test alanların belirlenmesinde süreklilik düzeltmesi için bir sabit ($C = -1/2$) kullanılarak ve kullanılmayarak analizler ayrı ayrı gerçekleştirilmiştir.

Tablo 2'de bu araştırmada kullanılan simülasyon deseni için manipüle edilen koşulların özeti yer almaktadır.

Tablo 2

Simülasyon Deseni

Madde Sayısı	Örneklem Büyüklüğü	Tahrifat Yapılan Madde Yüzdesi	Yetenek Düzeyi	Alfa Seviyesi	SBİ Türü
50	500, 5000, 50000, 99990	10, 20, 30, Değişken	1, 2, 3, 4, 5, Toplam	.05, .01, .005, .001, .0005, .0001, .00001	Süreklilik düzeltmesi kullanılarak, Süreklilik düzeltmesi kullanılmadan
160	500, 5000, 50000, 99990	10, 20, 30, Değişken	1, 2, 3, 4, 5, Toplam	.05, .01, .005, .001, .0005, .0001, .00001	Süreklilik düzeltmesi kullanılarak, Süreklilik düzeltmesi kullanılmadan

Tablo 2’de görüldüğü gibi sonuç olarak istatistiksel güç hesaplaması için 2(*madde sayısı*) X 4(*örneklem büyüklüğü*) X 4(*tahrifat yapılan madde yüzdesi*) X 6(*yetenek düzeyi*) X 7(*alfa seviyesi*) X 2(*SBI türü*) olmak üzere toplam 2688 simülasyon koşulu; I. Tip hata oranı hesaplanması için 2(*madde sayısı*) X 4(*örneklem büyüklüğü*) X 6(*yetenek düzeyi*) X 7(*alfa seviyesi*) X 2(*SBI türü*) olmak üzere toplam 672 simülasyon koşulu ortaya çıkmıştır. 2(*madde sayısı*) X 4(*örneklem büyüklüğü*) olmak üzere 8 temel veri dosyası üretilmiştir. 8 temel veri dosyası üzerinden silme simülasyonları ve analizler 25 kez tekrar edilerek toplam 200 farklı veri dosyası elde edilmiştir. 25 tekrardan elde edilen analiz bulgularının aritmetik ortalamaları raporlaştırılmıştır.

Verilerin Üretilmesi

Çalışmada kullanılan veriler, madde sayısı ve örneklem büyüklüğü olmak üzere iki koşula ve bu koşulların düzeylerine göre R programında kod yazılarak ve “mcIRT” paketi (Reif, 2015) kullanılarak üretilmiştir. Simülasyonda bireylerin yetenek parametreleri (θ), ortalaması 0, standart sapması 1 olan standart normal dağılımdan $N(0,1)$ üretilmiştir. Verilerin normal dağılım koşuluna ilişkin istatistikler “Verilerin Çözümlemesi” başlığı altında raporlaştırılmıştır.

STM’de madde parametreleri, ikili puanlanan maddelerden farklı olarak, her bir kategori için lambda (eğim [slope], λ) ve zeta (kesişim [intercept], ζ) olmak üzere ayrı ayrı üretilmektedir. STM’ye göre θ yetenek düzeyindeki bir bireyin maddenin k seçeneğini yanıtlama olasılığı Eşitlik 1’de verilmiştir (de Ayala, 2009):

$$p_j(x = (k|\theta, \alpha, \gamma) = \frac{e^{\gamma_{jk} + \alpha_{jk}\theta}}{\sum_{h=1}^{m_j} e^{\gamma_{jh} + \alpha_{jh}\theta}} \quad (1)$$

Eşitlik 1’de θ bireyin yetenek parametresi iken, k seçeneği için α_{jk} lambda (λ) ve γ_{jk} zeta (ζ) parametreleridir. Baker ve Kim (2004) zeta parametresini yer parametresi olarak tanımlamakta ve doğrudan b parametresi olarak yorumlanamayacağını ifade etmektedir. Eşitlik 2’de yer alan doğrusal dönüşümle zeta parametresinin b parametresine dönüştürülebileceğini belirtmektedir:

$$b = -\frac{\zeta}{\lambda} \quad (2)$$

de Ayala (2009) ile de Ayala ve Sava-Bolesta (1999) ise Baker tarafından önerilen dönüşümün gerçek bir değere karşılık gelmediğini belirterek zeta parametresine Eşitlik 3'teki dönüşüm uygulanarak b parametresi elde edilebileceğini ifade etmektedir:

$$b_{k^*,k} = \frac{\zeta_{k^*} - \zeta_k}{\lambda_k - \lambda_{k^*}} \quad (3)$$

Eşitlik 3'te k seçeneği göstermek üzere, kategori sayısı iki ve daha fazla olan maddeler için $k^* < k$ ve $\alpha_{k^*} \neq \alpha_k$ olarak tanımlanmıştır.

Lambda parametresi ise eğim olarak nitelendirilmekte ve a parametresi gibi madde ayırt ediciliği olarak yorumlanmaktadır (Baker ve Kim, 2004; Ostini ve Nering, 2006). Bu parametreye uygulanan dönüşüm Eşitlik 4'te yer almaktadır:

$$\lambda = a \quad (4)$$

STM'de lambda ve zeta parametrelerinin kestiriminde iki sınırlama getirilmektedir (Bock, 1972):

1. α_{jk} 'lar toplamı sifira eşittir ($m_j \sum \alpha_{jh} = 0$).
2. γ_{jk} 'lar toplamı sifira eşittir ($m_j \sum \gamma_{jh} = 0$).

Simülasyonda madde parametreleri belirlenirken her bir madde için eğim ve kesişim parametreleri $\alpha \sim U(0.5, 2.0)$ ve $\gamma \sim N(0, 1)$ olarak belirlenmiştir. Daha sonra her bir maddede α_{jk} 'lar ve γ_{jk} 'ların toplamları ayrı ayrı sifira eşit olacak şekilde diğer seçeneklere ilişkin lambda ve zeta parametreleri R'da yazılan kodlar ile üretilmiştir. Eğim parametreleri belirlenirken uniform (düzgün) bir dağılım kullanılmış olup (Gündeğer ve Doğan, 2016; Ogasawara, 2001; Paek ve Young, 2005), iyi hazırlanmış bir madde havuzunda karşılaşılan değer aralıkları olan 0.5-2.0 aralığıyla sınırlandırılmıştır (Baker, 2016). Sabit parametreleri normal dağılımdan türetilmiş olup ortalaması 0, standart sapması 1 olarak belirlenmiştir (Kim ve Lee 2004; Ogasawara, 2001). 50 ve 160 madde için madde parametreleri belirlendikten sonra 250, 2500, 25000 ve 49995 test alan için beş seçenekli çoktan seçmeli simülatif yanıt örüntüleri R programında mcIRT paketiyle üretilmiştir. Her bir örneklem için üretilen veriler MULTOLOG programında analiz edilerek birey ve madde parametreleri kestirilmiştir. Üretilen yetenek parametrelerinin dağılımı ile kestirilen yetenek parametrelerinin dağılımı arasında uyumsuzluk olduğu, tüm verilerin normal dağılıma uygun üretilmediği belirlenmiştir. Bu nedenle her bir örneklem için birçok veri üretildikten sonra, kestirilen yetenek parametrelerinde normal dağılım

koşulunu sağlayan veriler analizlerde kullanılmak üzere belirlenmiştir. 49995 test alanın 50 ve 160 maddelik veri setlerine ilişkin MULTİLOG ile kestirilen lambda ve zeta parametreleri Ek 1’de yer almaktadır.

İlk veriler üretildikten sonra iyicil silmeler yapılmadan önceki başlangıç yanıtlarını içeren veri setleri ile iyicil silmeleri içeren veri seti üzerinde hileli silmeleri içerecek veri setleri elde edilmiştir. İyicil silmelerin oranlarına ve hileli silmelerin koşullarına ilişkin detaylı bilgi “Verilerin Silinme Simülasyonu” başlığı altında yer almaktadır.

Verilerin Silinme Simülasyonu

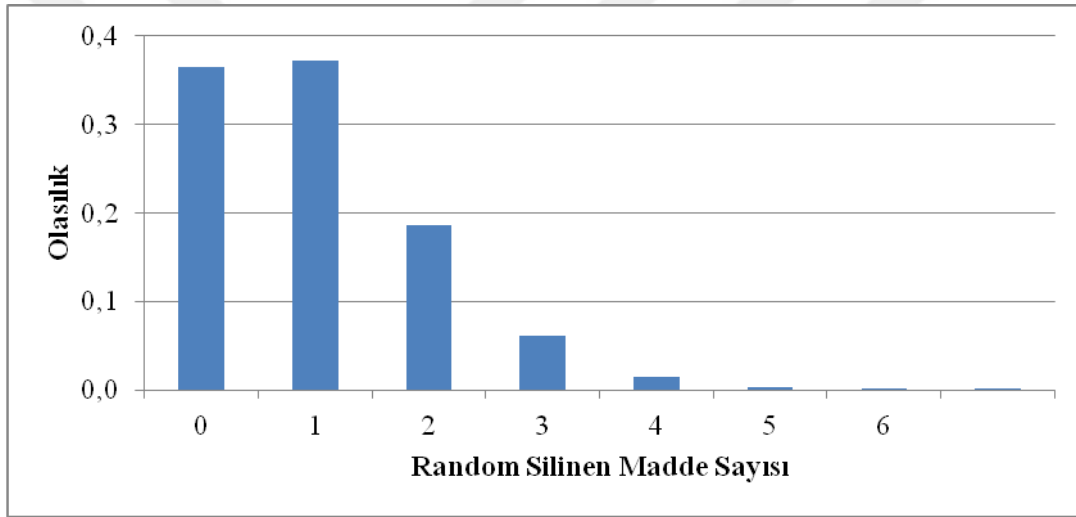
Veri setleri, silmelerin iki farklı türü olan iyicil silmeler ve hileli silmelerin her ikisini de içerecek şekilde, Wollack ve diğerlerinin (2015) simülasyonu model alınarak üretilmiştir. Başlangıç yanıtlarının (silindiği düşünülen ilk yanıtlar) tahmini ve hileli silmelerin üretilmesi için R programında kod yazılmıştır. Kod yazımının ve kontrolünün kolay yapılabilmesi için öncelikle küçük örneklemler ile silme simülasyonları yapılmaya başlanmış, kod yazımı ile kontroller tamamlandıktan sonra tüm veri setlerinde kullanılmıştır.

Madde sayısı ve örneklem büyüklüğü koşullarına göre silme işlemleri her veri seti için tekrarlanmıştır. Örnek oluşturması açısından 250 test alanına ilişkin veri seti için yapılan silme oranları aşağıda açıklanmıştır.

İyicil silmelerin tesadüfi, kaydırma ve dizi sonu silmeler olmak üzere üç türü bulunmaktadır. Verilerin kaydırma ve dizi sonu silmelerinin her ikisi de nadir olacağından (Wollack vd., 2015), silmelerin bu türlerinin her biri için 10’ar kişi atanmıştır. Farklı yetenek düzeylerindeki test alanlar, θ ’ya dayalı beş farklı yetenek düzeyi gruplarına ayrılarak, her durum için, kaydırma silmelerini üretmek üzere her yetenek düzeyine 2 kişi tesadüfi olarak atanmış ve başka 2 kişi de dizi sonu silmeleri için aynı şekilde atanmıştır. Kalan 230 kişi tesadüfi silmeler üretmek üzere kullanılmıştır.

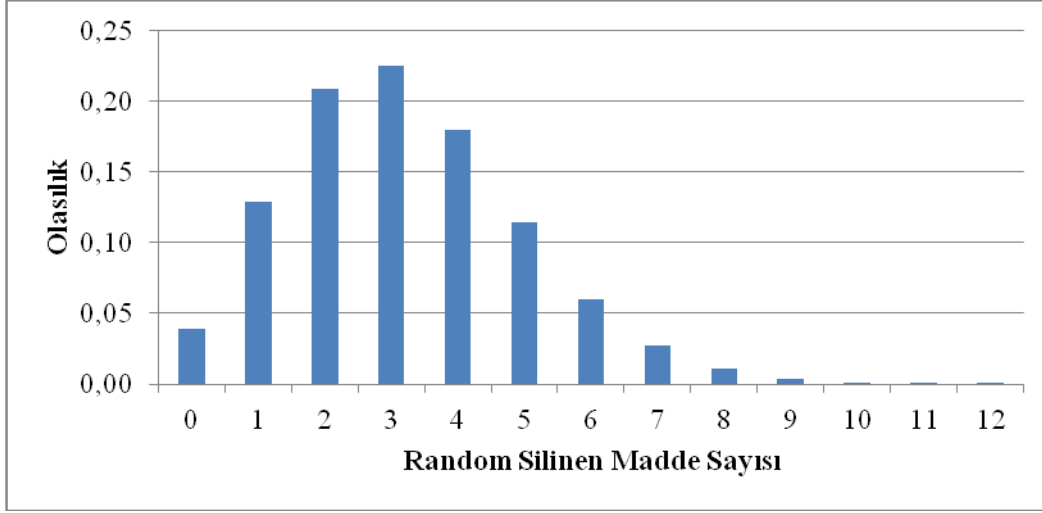
Tesadüfi silmelerin simülasyonunda Y_j (iyicil silme yapan birey), 50 maddelik veri seti için $p = .02$ ve $N = 50$ (madde sayısı), 160 maddelik veri seti için $p = .02$ ve $N = 160$ olan binominal dağılım kullanılmıştır. Test alan bireyler için tesadüfi silme sayıları belirlendikten sonra test alanların (Y_j) silme yaptıkları maddelerin belirlenmesi için, yerine koyma yapılmadan, 50 maddelik veri seti için $U[1,50]$ ve 160 maddelik veri seti için $U[1,160]$ rassal sayılar tablosu oluşturulmaktadır. Birey j ’nin her bir silinmiş maddesi için mevcut yanıtından farklı olma koşulu altında ikinci bir yanıt (silinmiş olan ilk yanıt) üretilmiştir. Yeni üretilen yanıtlar, test alanın ilk önce işaretlediği ve daha sonra silerek değiştirdiği başlangıç yanıtı olarak

kabul edilmiştir. İlk olarak üretilen yanıtlar (test alanın bütün yanıt seçeneklerinden birini seçmesine izin veren test alanın en sonda ulaştığı varsayılan yanıt) son yanıt olarak kabul edilmiştir. Wollack ve diğerleri (2015) tesadüfi silmelerde başlangıç yanıtını üretirken STM'yi kullanmışlardır. Sinharay ve Johnson (2016) ise tesadüfi silmelerin, test alanın yanlışlıkla bir yanıt seçtiğini, yanıtını gözden geçirirken bunu başka bir yanıtla değiştirdiğini ifade ederek başlangıç yanıtını tesadüfi olarak belirlemiştir. Bu çalışmada da tesadüfi silmelerin nadir olduğu ve analizlerde tesadüfi başlangıç yanıtı üretmenin STM ile üretmekten daha kolay olduğu göz önünde bulundurularak, analizlerde kolaylık sağlamak adına başlangıç yanıtı tesadüfi olarak tahmin edilmiştir. Veri setindeki madde sayısına göre tesadüfi silme sayılarının olasılık grafikleri Şekil 2 ve Şekil 3'te yer almaktadır.



Şekil 2. $N=50$ ve $p=.02$ Olan Binominal Dağılımdan Elde Edilen Tesadüfi Silme Sayılarının Olasılık Grafiği

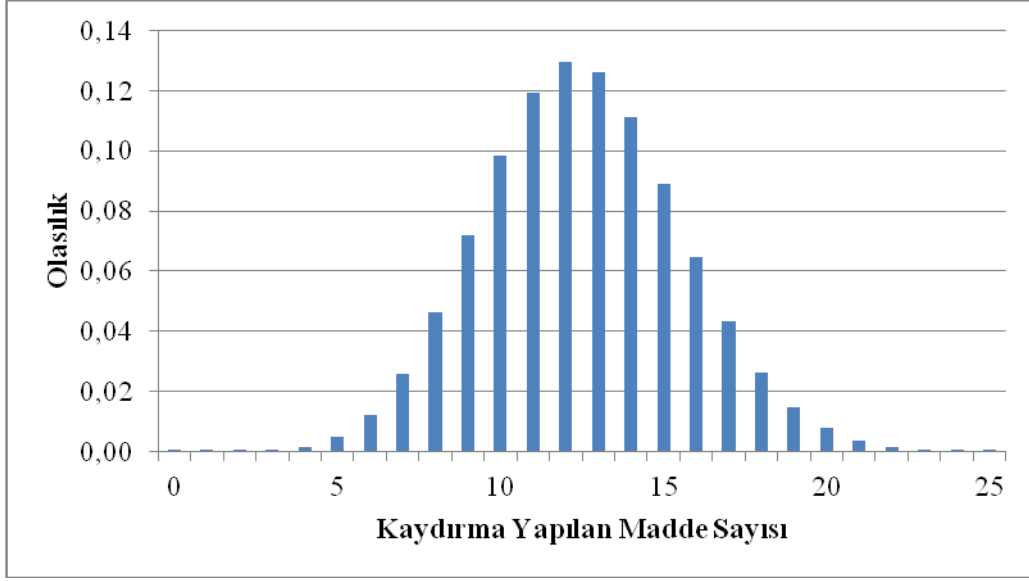
Şekil 2'ye göre 50 maddelik veri setinde, test alanlar ortalama bir tesadüfi silme ve test alanların yaklaşık %75'i sıfır ya da bir tesadüfi silmeye sahip olacak şekilde simüle edilmiştir.



Şekil 3. $N=160$ ve $p=.02$ Olan Binominal Dağılımdan Elde Edilen Tesadüfi Silme Sayılarının Olasılık Grafiği

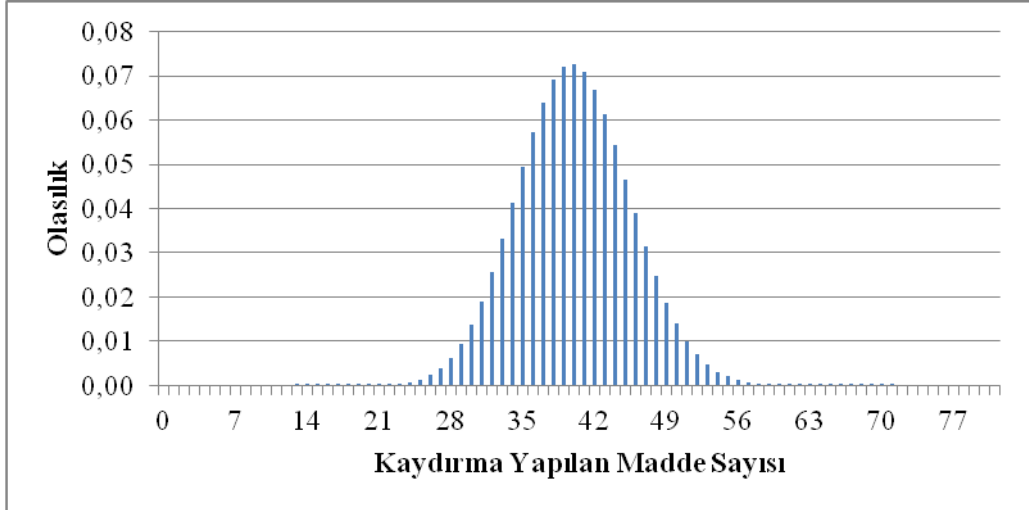
Şekil 3'e göre 160 maddelik veri setinde, test alanlar yaklaşık olarak ortalama üç tesadüfi silme ve test alanların yaklaşık %60'ı en fazla üç tesadüfi silmeye sahip olacak şekilde simüle edilmiştir.

Kayırdırma silmelerinde ise, test alan birey j için kaydırma maddelerinin sayısını (M_j) tanımlamak üzere; 50 maddelik veri seti için $N= 50$ ve $p= .25$, 160 maddelik veri seti için $N= 160$ ve $p= .25$ olan binominal dağılım kullanılmıştır. Veri setindeki madde sayısına göre kaydırma yapılan madde sayılarının olasılık grafikleri Şekil 4 ve Şekil 5'te yer almaktadır. Tüm kaydırma hataları silme ile sonuçlanmayacağından, burada M_j , Y_j için bir üst sınır sağlamaktadır. İlk kaydırma maddesinin yerini belirlemek üzere, 50 maddelik veri seti için tesadüfi olarak $U[1, 50 - M_j + 1]$ 'den ve 160 maddelik veri seti için tesadüfi olarak $U[1, 160 - M_j + 1]$ 'den bir sayı üretilmiştir. Her kaydırma maddesi için i ve $i-1$ konumundaki yanıtlar karşılaştırılmıştır. Eğer farklılarsa, $i - 1$ 'deki yanıt i maddesi için silinmiş yanıt olarak alınmıştır. İki maddedeki yanıtlar aynı ise, kaydırmayla yapılan işaretleme gerçek işaretleme ile aynı olacağı için, bu durum silme olarak kabul edilmemiştir.



Şekil 4. $N=50$ ve $p=.25$ Olan Binominal Dağılımdan Elde Edilen Kaydırma Yapılan Madde Sayılarının Olasılık Grafiği

Şekil 4'e göre 50 maddelik veri setinde, test alanların yaklaşık olarak ortalamada 12.5 maddede kaydırma yapması beklenmektedir. Test alanların yaklaşık %80'i 9 ile 16 madde arasında kaydırma yapacak şekilde simüle edilmiştir.



Şekil 5. $N=160$ ve $p=.25$ Olan Binominal Dağılımdan Elde Edilen Kaydırma Yapılan Madde Sayılarının Olasılık Grafiği

Şekil 5'e göre 160 maddelik veri setinde, test alanların yaklaşık olarak ortalamada 40 maddede kaydırma yapması beklenmektedir. Test alanların yaklaşık %80'i 33 ile 47 madde arasında kaydırma yapacak şekilde simüle edilmiştir.

Dizi sonu silmelerinde, 50 maddelik veri setinde testteki son 12 madde (39-50) için ve 160 maddelik veri setinde testteki son 38 madde (123-160) için tahmini başlangıç yanıtı tesadüfi ve her seçeneği işaretleme olasılığı eşit ve $p= 0.2$ olacak şekilde veriler üretilmiştir. Tesadüfi olarak seçilen başlangıç yanıtının nihai yanıt ile eşleşmesi durumunda, bu durum silme olarak kabul edilmemiştir.

Hileli silmelerde sabit tahrifat için madde sayısının %10'u, %20'si ve %30'u oranında madde; değişken tahrifat için ise her test alan için değişen sayıda madde tahrifata uğratılmıştır. Hileli silmeler, iyicil silmeleri içeren veri setleri üzerinden üretilmiştir, böylece hileli silmeleri olan test alanlar aynı zamanda iyicil silmelere de sahip olabilmektedir.

Sabit tahrifat durumlarında, sabit sayıda madde silinerek doğrusu ile değiştirilmektedir. Doğru yanıtlanmış maddeler silinemeyeceğinden (eğer silinseydi aynı seçenek tekrar işaretlenirdi ve ikincil silme izi gerçekleşmezdi), sadece başlangıçta yanlış yanıtlanan maddeler silinip değiştirilmeye aday olmaktadır. Testi alan her birey için, yanlış yanıtlanmış maddelerin hepsinin silinip doğruyla değiştirilme olasılığı eşittir. Sabit tahrifat, madde sayısının %10, %20 ve %30'u olacak şekilde üç koşul olarak belirlenmiştir. Buna göre sabit tahrifat durumlarında 50 maddelik veri seti için 5, 10 ve 15; 160 maddelik veri seti için ise 16, 32 ve 48 madde tahrifata uğratılmıştır.

Değişken tahrifat durumları ise sınavda yetkili kişinin, belirlenen kesme puanını aşmayan öğrencilerin bu eşiği aşmasını sağlayacak kadar maddeyi değiştirmesi durumudur. Bazı öğrenciler kesme noktasına diğer öğrencilerden daha uzak olduğundan, bu durum her birey için farklı sayıda silme ile sonuçlanmaktadır (Wollack vd., 2015). Değişken tahrifat için her bir yetenek düzeyi grubunda belirlenen sayıdaki kişi, tesadüfi olarak tahrifata uğratılmak üzere seçilmektedir. Sadece belirlenen kesme puanının altında kalan kişiler seçime uygun olmaktadır. Tahrifat yapılacak bireylerin her biri için, kesme puanına erişebilmelerini sağlayacak sayıda silme yapılarak değişken tahrifat tamamlanmaktadır. Madde sayısı değişimlendiğinde indeksin istatistiksel gücü ve I. Tip hatasında oluşacak farkın karşılaştırmasının doğru yapılabilmesi için, kesme puanının belirlenmesinde her iki madde sayısı için de normal yetenek dağılımı altındaki varsayıma göre yetenek düzeyi sıfıra karşılık gelen gerçek puan temel alınmıştır. Madde sayısı ve örneklem büyüklüğü koşullarına göre her bir örneklem için en küçük ve en büyük gerçek puan ile yetenek düzeyi sıfıra karşılık gelen kesme puanları Tablo 3'te yer almaktadır.

Tablo 3

Madde Sayısı ve Örneklem Büyüklüğü Koşullarına Göre Değişken Silme İçin Yetenek Düzeyi Sıfıra Karşılık Gelen Gerçek Puanlar ve Kullanılan Kesme Puanları

Örneklem Büyüklüğü	Madde Sayısı							
	50 Madde				160 Madde			
	250	2500	25000	49995	250	2500	25000	49995
Gerçek Puan _{min}	1.46	0.59	0.35	0.31	7.31	5.63	5.19	4.65
Gerçek Puan _{mak}	38.44	41.08	41.99	41.89	103.65	110.84	112.84	113.84
Gerçek Puan ₀₌₀	11.77	11.22	11.33	11.55	39.95	40.85	40.95	40.58
Kesme Puanı	12.00	12.00	12.00	12.00	40.00	41.00	41.00	41.00

Değişken tahrifat durumunda, tahrifat yapan kişinin bireylerin yetenek düzeylerine ilişkin kesin bir bilgiye sahip olmadığı ancak tahmin yürüttüğü varsayılmaktadır. Bu nedenle kesme puanlarının, Wollack ve diğerlerine (2015) benzer şekilde, belirlenen gerçek puanların üste yuvarlanmasıyla belirlenmesinin işlevsel olacağı düşünülmektedir. Veri setlerinin yetenek düzeylerinin dağılımı birbirine benzer olduğundan 50 madde için tüm kesme puanlarının aynı olduğu, 160 madde için ise bir veri seti dışında kesme puanlarının aynı olduğu Tablo 3'te görülmektedir. 160 madde için 250 örnekleme kesme puanı diğerlerinden yalnızca bir puan fazladır. Yetenek düzeyi yüksek olan gruplarda kesme puanı koşulunu sağlayan birey olmaması durumunda değişken tahrifat yapılmamaktadır.

Örneklem büyüklüğü ve silme türlerine göre silme yapılması beklenen tahmini birey sayısı Tablo 4'te yer almaktadır.

Tablo 4

Örneklem Büyüklüğü ve Silme Türlerine Göre Silme Yapılması Beklenen Tahmini Birey Sayısı

Silme Türü		Madde Sayısı								
		50 Madde				160 Madde				
		Örneklem Büyüklüğü				Örneklem Büyüklüğü				
		250	2500	25000	49995	250	2500	25000	49995	
İyicil Silme	Tesadüfi	156**	1577	15769	31534	236**	2382	23821	47638	
	Kaydırma	2*	10	100	200	2*	10	100	200	
	Dizi Sonu	2*	10	100	200	2*	10	100	200	
Hileli Silme	Sabit	%10	20	100	1000	2000	20	100	1000	2000
		%20	20	100	1000	2000	20	100	1000	2000
		%30	20	100	1000	2000	20	100	1000	2000
	Değişken	20	100	1000	2000	20	100	1000	2000	

*250 test alan için kaydırma ve dizi sonu silmelerinde belirlenen yüzdeye karşılık gelen test alan sayısı 2 olduğundan ve 250 test alan içerisinde 2 kişide silme yapılması tüm yetenek düzeylerinden silme yapılamayacağı anlamına geldiğinden her yetenek düzeyinden 2 olmak üzere silme türüne göre toplamda 10'ar test alan seçilmiştir.

**Tesadüfi silmeler, kaydırma ve dizi sonu silmeleri yapıldıktan sonra kalan test alanlar üzerinden yapıldığı için, tesadüfi silme yapan test alan sayısının revize edilmesi gerekmektedir.

Örneklem büyüklüğü ve silme türlerine göre silme yapılan tahmini birey sayısının revize edilmiş hali Tablo 5'te yer almaktadır.

Tablo 5

Örneklem Büyüklüğü ve Silme Türlerine Göre Silme Yapılan Tahmini Birey Sayısı (Revize)

Silme Türü		Madde Sayısı								
		50 Madde				160 Madde				
		Örneklem Büyüklüğü				Örneklem Büyüklüğü				
		250	2500	25000	49995	250	2500	25000	49995	
İyicil Silme	Tesadüfi	146	1577	15768	31534	221	2382	23821	47638	
	Kaydırma	10	10	100	200	10	10	100	200	
	Dizi Sonu	10	10	100	200	10	10	100	200	
Hileli Silme	Sabit	%10	20	100	1000	2000	20	100	1000	2000
		%20	20	100	1000	2000	20	100	1000	2000
		%30	20	100	1000	2000	20	100	1000	2000
	Değişken	20	100	1000	2000	20	100	1000	2000	

İyicil silmeleri içeren 250 kişilik veri setinde her bir yetenek düzeyinde eşit sayıda ve 4 kişi olmak üzere, %10 oranında sabit silme için toplamda 20 kişi, %20 oranında sabit silme için toplamda 20 kişi ve %30 oranında sabit silme için toplamda 20 kişi sabit tahrifatın mağduru olarak tesadüfi olarak seçilmiştir. Sadece iyicil silmeleri içeren 250 kişilik aynı veri seti için belirlenen kesme puanına göre her bir yetenek düzeyinde eşit sayıda ve 4 kişi olmak üzere toplamda 20 kişi değişken tahrifat mağduru olarak seçilmiştir. Sabit ve değişken tahrifata uğrayan test alanların veri setleri birleştirilerek iyicil silmeleri ve hileli silmelerin her iki türünü de içeren 500 test alandan oluşan nihai veri seti oluşturulmuştur.

İyicil silmeleri içeren 2500 kişilik veri setinde her bir yetenek düzeyinde eşit sayıda ve 20 kişi olmak üzere, %10 oranında sabit silme için toplamda 100 kişi, %20 oranında sabit silme için toplamda 100 kişi ve %30 oranında sabit silme için toplamda 100 kişi sabit tahrifatın mağduru olarak tesadüfi olarak seçilmiştir. Sadece iyicil silmeleri içeren 2500 kişilik aynı veri seti için belirlenen kesme puanına göre her bir yetenek düzeyinde eşit sayıda ve 20 kişi olmak üzere toplamda 100 kişi değişken tahrifat mağduru olarak seçilmiştir. Sabit ve değişken tahrifata uğrayan test alanların veri setleri birleştirilerek iyicil silmeleri ve hileli silmelerin her iki türünü de içeren 5000 test alandan oluşan nihai veri seti oluşturulmuştur.

İyicil silmeleri içeren 25000 kişilik veri setinde her bir yetenek düzeyinde eşit sayıda ve 200 kişi olmak üzere, %10 oranında sabit silme için toplamda 1000 kişi, %20 oranında sabit silme için toplamda 1000 kişi ve %30 oranında sabit silme için toplamda 1000 kişi sabit tahrifatın mağduru olarak tesadüfi olarak seçilmiştir. Sadece iyicil silmeleri içeren 25000 kişilik aynı veri seti için belirlenen kesme puanına göre her bir yetenek düzeyinde eşit sayıda ve 200 kişi olmak üzere toplamda 1000 kişi değişken tahrifat mağduru olarak seçilmiştir. Sabit ve değişken tahrifata uğrayan test alanların veri setleri birleştirilerek iyicil silmeleri ve hileli silmelerin her iki türünü de içeren 50000 test alandan oluşan nihai veri seti oluşturulmuştur.

İyicil silmeleri içeren 49995 kişilik veri setinde her bir yetenek düzeyinde eşit sayıda ve 400 kişi olmak üzere, %10 oranında sabit silme için toplamda 2000 kişi, %20 oranında sabit silme için toplamda 2000 kişi ve %30 oranında sabit silme için toplamda 2000 kişi sabit tahrifatın mağduru olarak tesadüfi olarak seçilmiştir. Sadece iyicil silmeleri içeren 49995 kişilik aynı veri seti için belirlenen kesme puanına göre her bir yetenek düzeyinde grupta eşit sayıda ve 400 kişi olmak üzere toplamda 2000 kişi değişken tahrifat mağduru olarak seçilmiştir. Sabit ve değişken tahrifata uğrayan test alanların veri setleri birleştirilerek iyicil silmeleri ve hileli silmelerin her iki türünü de içeren 99990 test alandan oluşan nihai veri seti oluşturulmuştur.

Tüm veri setlerinde yetenek düzeyi yüksek gruplarda kesme puanı koşulunu sağlayan birey olmaması durumunda değişken tahrifat yapılmamıştır. İyicil silmeleri içeren sabit ve

değişken silmeli veri setleri birleştirildiğinde analize hazır hale gelen 500, 5000, 50000, 99990 test alanın veri setlerine ilişkin, örneklem büyüklüğü ve silme türlerine göre silme yapılan tahmini birey sayısı Tablo 6’da, tahrifat türüne göre tahrifata uğrayan birey yüzdeleri ise Tablo 7’de yer almaktadır.

Tablo 6

Sabit ve Değişken Silmeli Veri Setleri Birleştirildiğinde Örneklem Büyüklüğü ve Silme Türlerine Göre Silme Yapılan Tahmini Birey Sayısı

Silme Türü		Madde Sayısı								
		50 Madde				160 Madde				
		Örneklem Büyüklüğü				Örneklem Büyüklüğü				
		500	5000	50000	99990	500	5000	50000	99990	
İyicil Silme	Tesadüfi	292	3154	31536	63068	442	4764	47642	95276	
	Kaydırma	20	20	200	400	20	20	200	400	
	Dizi Sonu	20	20	200	400	20	20	200	400	
Hileli Silme	Sabit	%10	20	100	1000	2000	20	100	1000	2000
		%20	20	100	1000	2000	20	100	1000	2000
		%30	20	100	1000	2000	20	100	1000	2000
	Değişken		20	100	1000	2000	20	100	1000	2000
			20	100	1000	2000	20	100	1000	2000

Tablo 6’da 500 örneklem büyüklüğü için kaydırma ve dizi sonu silme sayılarının 5000 örneklem büyüklüğü ile aynı olduğu görülmektedir.

Tablo 7

Örneklem Büyüklüğü ve Tahrifat Türüne Göre Tahrifatlı Birey Yüzdesi

Tahrifat Türü	Tahrifatlı Madde Yüzdesi	Örneklem Büyüklüğüne Göre Tahrifatlı Birey Yüzdesi (%)			
		500	5000	50000	99990
Sabit	%10	4	2	2	2
	%20	4	2	2	2
	%30	4	2	2	2
Değişken		4	2	2	2
Toplam		16	8	8	8

Tablo 7’de tahrifatlı birey yüzdelерinin, 500 örneklem büyüklüğü dışında, sabit kaldığı görülmektedir. 500 örneklem büyüklüğü için tahrifat yüzdelерinin eşit olmamasının gerekçeleri yukarıda açıklanmıştır.

R programında yazılan kodların algoritma örneği 50 madde ve 50000 örneklem için Ek 2’de yer almaktadır. Algoritma tüm madde ve örneklem koşullarına uygun olarak tekrarlanmaktadır.

Verilerin Çözümlemesi

Simülasyon çalışmalarında veri üretim sürecinde, verilerin istenilen parametrelere uygun şekilde üretilip üretilmediğinin test edilmesi önerilmektedir (Davey, Nering ve Thompson, 1997; Morris, White ve Crowther, 2019; Soltana, Sabetzadeh ve Briand, 2017). Bu nedenle verilerin çözümlenmesine geçmeden önce üretilen her bir veri seti için yetenek düzeylerinin normal dağılımı, MTK’nın temel varsayımları olan tek boyutluluk ve yerel bağımsızlık varsayımlarına ilişkin analizlere yer verilmiştir. Daha sonra silinmemiş maddeler üzerinden yetenek kestirimi ($\hat{\theta}_{j|i \in I_{E,j}}$), silinmemiş maddeler üzerinden yapılan yetenek kestirimine ilişkin yanlılık ve RMSE analizleri ile SBİ, istatistiksel güç ve I. Tip hata analizlerine ilişkin bilgiler sunulmuştur.

Üretilen Verilerden Elde Edilen Yetenek Parametrelerinin Normal Dağılım Varsayımının Kontrolü

Madde parametreleri ve yetenek parametrelerinin üretilmesinin ardından test alanlara ilişkin yanıtlayıcı matrisleri üretilmiştir. Üretilen yanıtlar kullanılarak yetenek düzeylerinin MULTILOG ile kestirilmesinden sonra normal dağılım gösterip göstermediğinin kontrolü için her bir veri setine ilişkin betimsel istatistiklerden aritmetik ortalama, mod, medyan, standart sapma, çarpıklık ve basıklık katsayıları hesaplanmıştır. Üretilen verilere ait betimsel istatistikler yerel bağımsızlık ve tek boyutluluk varsayımlarının sonuçları ile birlikte Tablo 9’da yer almaktadır.

Üretilen Verilerden Elde Edilen Yetenek Parametrelerinin Tek Boyutluluk

Varsayımının Kontrolü

Tek boyutluluk, bir testte yer alan maddelerin yalnız bir özelliği ölçmesi olarak tanımlanmaktadır. Bu varsayım her zaman tamamen karşılanamayabilmektedir. Başat faktörün varlığının belirlenmesi tek boyutluluk varsayımının karşılanması için yeterlidir (Hambleton vd., 1991). Parametrik bir teknik olan faktör analizinin kullanımı için sağlanması gereken örneklem büyüklüğü koşulu (Çokluk, Şekercioğlu ve Büyüköztürk, 2010) özellikle 160 maddelik veri setinde örneklem büyüklüğü 250 olduğu durumda sağlanamadığından nonparametrik bir teknik kullanılması uygun görülmüştür. Bu nedenle araştırmada tek boyutluluk varsayımının incelenmesinde Zhang (2007) tarafından önerilen çok kategorili maddelerde kullanılabilen DETECT analizinden yararlanılmıştır. DETECT analizi, koşullu kovaryans matrislerine dayanan parametrik olmayan bir tekniktir. DETECT analizi boyutluluğun belirlenmesinde, özellikle madde sayısı 20 ve üzerinde olduğunda (Roussos ve Ozbek, 2006), etkili bir yöntem olarak ifade edilmektedir (Zhang ve Stout, 1999). DETECT analizinin açımlayıcı ve doğrulayıcı olmak üzere iki türü bulunmakla birlikte, boyutluluğun belirlenmesinde açımlayıcı DETECT analizine kıyasla daha çok doğrulayıcı DETECT analizi önerilmektedir (Koğar, 2018). Doğrulayıcı DETECT analizi için R yazılımının “sirt” paketinden (Robitzsch, 2019) yararlanılmıştır. Tek boyutluluk analiz sonuçları, betimsel istatistikler ve yerel bağımsızlık analiz sonuçları ile birlikte Tablo 9’da yer almaktadır. Tek boyutluluk analiz sonuçlarının değerlendirilmesinde Tablo 8’de yer alan kriterler referans alınmıştır (Jang ve Roussos, 2007; Zhang, 2007).

Tablo 8

DETECT Analizi Sonuçları İçin Referans Değerler

Referans Değer	Yorum	
DETECT > 1.00	Güçlü çok boyutluluk	
0.40 < DETECT < 1.00	Orta dereceli çok boyutluluk	
0.20 < DETECT < 0.40	Zayıf çok boyutluluk	
DETECT < 0.20	Gerçek tek boyutluluk	
ASSI = 1.00	RATIO = 1.00	Basit yapı altında maksimum değer
ASSI > 0.25	RATIO > 0.36	Tek boyutluluktan gerçek sapma
ASSI < 0.25	RATIO < 0.36	Gerçek tek boyutluluk

Tablo 8'e göre gerçek tek boyutluluğun sağlanabilmesi için DETECT değerinin 0.20'nin altında, ASSI değerinin 0.25'in altında, RATIO değerinin ise 0.36'nın altında olması gerekmektedir. Negatif değerler de tek boyutluluğa işaret etmektedir.

Üretilen Verilerde Madde Yanıtlarının Yerel Bağımsızlık Varsayımının Kontrolü

Yerel bağımsızlık, test performansını etkileyen yetenek düzeyi sabit tutulduğunda, test alanların herhangi bir maddeye verdiği yanıtların istatistiksel olarak bağımsız olduğu anlamına gelmektedir. Başka bir deyişle, sınava girenlerin yetenekleri dikkate alındıktan sonra, sınava girenlerin farklı maddelere verdikleri yanıtlar arasında bir ilişki yoktur. Kısaca, test alanların test maddelerine yanıtlarını etkileyen tek faktörün yetenekleri olduğu durum yerel bağımsızlık olarak ifade edilmektedir (Hambleton vd., 1991). Yen (1984) tarafından geliştirilen Q_3 istatistiği yerel bağımsızlık varsayımının test edilmesinde sıklıkla kullanılan bir yöntemdir. Q_3 istatistiği madde çiftleri arasında hesaplanan, artık değerlere ilişkin bir korelasyon matrisi üretmektedir (DeMars, 2016). Maddelerin yerel bağımsız olması için Yen (1993) Q_3 istatistiğinin 0.20'den küçük olmasını önermiştir.

Alanyazında yerel bağımsızlık varsayımının sağlanması için Q_3 istatistiğinde kritik değer olarak ± 0.20 kullanılması önerilmekte (Chen ve Thissen, 1997) ve sıklıkla bu değer kesme puanı olarak kullanılmaktadır (Lee, 2004; Monseur, Baye, Lafontaine ve Quittre, 2011; Reeve vd., 2007). Buna karşın alanyazında kesme puanı olarak ± 0.36 'nın kullanıldığı da görülmektedir (Flens vd., 2017; Smits, Cuijpers ve van Straten 2011; Tan, Cai, Li, Zhang ve Tu, 2018). Yerel bağımsızlık varsayımının test edilmesi amacıyla Yen'in Q_3 istatistiğinin hesaplanması için R yazılımının "mirt" paketinden (Chalmers vd., 2018) yararlanılmıştır. Yerel bağımsızlık analiz sonuçları, betimsel istatistikler ve tek boyutluluk analiz sonuçlarıyla birlikte Tablo 9'da yer almaktadır.

Tablo 9

Üretilen Veriler İçin Hesaplanan Betimsel İstatistik, Tek Boyutluluk ve Yerel Bağımsızlık Analizlerine Ait Sonuçlar

Madde Sayısı	Örneklem Büyüklüğü	Yetenek Dağılımı Betimsel İstatistikleri						Tek Boyutluluk Varsayımı Analizleri			Yerel Bağımsızlık Varsayımı Analizleri	
		Aritmetik Ortalama	Standart Sapma	Medyan	Mod	Çarpıklık Katsayısı	Basıklık Katsayısı	DETECT	ASSI	RATIO	Q ₃ _min	Q ₃ _max
50	250	0.06	0.90	0.00	0.00	0.21	-0.52	-0.51	-0.06	-0.07	-0.19	0.21
	2500	0.02	0.99	0.00	0.00	0.25	-0.53	-0.90	-0.20	-0.35	-0.09	0.07
	25000	0.05	1.04	0.00	0.00	0.27	-0.71	-0.89	-0.45	-0.69	-0.04	0.03
	49995	0.06	1.03	0.00	0.00	0.24	-0.75	-0.93	-0.56	-0.79	-0.04	0.04
160	250	0.10	0.90	0.08	0.06	0.12	-0.54	0.08	0.01	0.01	-0.25	0.22
	2500	0.05	0.96	0.05	0.03	0.00	-0.75	-0.17	-0.04	-0.07	-0.08	0.12
	25000	0.10	0.99	0.08	0.03	0.06	-0.87	-0.24	-0.17	-0.30	-0.03	0.13
	49995	0.12	0.99	0.09	0.19	0.08	-0.85	-0.24	-0.24	-0.40	-0.07	0.14

Tablo 9’da yer alan sonuçlara göre veri setlerinden kestirilen yetenek düzeylerinin aritmetik ortalama deęerleri 50 madde için 0.02 ile 0.06 deęerleri arasında; 160 madde için 0.05 ile 0.10 deęerleri arasında deęişmektedir. Standart sapma deęerleri incelendięinde 50 madde için 0.90 ile 1.04 deęerleri arasında; 160 madde için 0.90 ile 0.99 deęerleri arasında deęişkenlik gösterdięi görülmektedir. Medyan deęerleri incelendięinde 50 madde için tüm veri setlerinde 0.00, 160 madde için 0.05 ile 0.09 deęerleri arasında deęişkenlik gösterdięi görülmektedir. Tüm veri setleri için çarpıklık ve basıklık katsayılarının ± 1 aralıęında yer aldığı görülmektedir. Tüm veri setleri için medyan ve mod deęerlerinin aritmetik ortalamadan çok farklılaşmadıęından; çarpıklık ve basıklık katsayıları da normal daęılım sınırları ierisinde yer aldığından veri setlerine ait yetenek parametreleri normal daęılım göstermektedir. DETECT analizine göre tüm veri setleri için DETECT deęerleri 0.20’nin altında; ASSI deęerleri 0.25’in altında ve RATIO deęerleri 0.36’nın altında olduęundan tüm veri setlerinde yetenek parametreleri tek boyutludur. Yen’in Q_3 istatistięine göre 50 ve 160 maddelik veri setlerinde 250 örneklem büyüklüęü dıřında, tüm madde çiftlerinin artık deęerlerinin korelasyonu ± 0.20 aralıęı ierisindedir. 250 örneklem büyüklüęü için ise küçük örneklemelerde analizlere devam edilebilmesi amacıyla, alanyazında ± 0.36 aralıęının kullanılmasına iliřkin örneklere dayanarak, tüm veri setleri için yerel baęımsızlık varsayımının karşılandığı kabul edilmektedir.

Yanıtları Silinmemiř Maddelere Dayalı Yetenek Kestirimi

Veriler ve silmeler üretildikten sonra, MULTILOG kullanılarak her bir test alan için yetenek düzeyi, $\hat{\theta}_{j|i \in I_{E,j}}$, yanıtları silinmemiř maddelere dayanarak kestirilmiřtir.

Bireysel Yetenek Kestiriminde Yanlılık (Bias) ve Hataların Ortalama Karekökü (RMSE) Deęerlerinin Hesaplanması

SBİ ile maddelerin sadece bir alt testi kullanılarak θ kestirildięinden, öncelikle bu yaklaşımın θ kestirimlerinde yanlılığa (bias) neden olmadıęının ve θ ’nın yeterli hassasiyette kestirilebildięinin gösterilmesi gerekmektedir. Tahrifatın simüle edilmedięi tüm test alanları kullanarak, θ üzerinde yanlılığı ve hataların ortalama karekökü (RMSE) Eřitlik 5 ve Eřitlik 6’ya göre hesaplanmaktadır (Wollack vd., 2015):

$$Yanluluk = \frac{\sum_{j=1}^N \left[\left(\hat{\theta}_{j[i \notin I_{E,j}]} \right)^{-\theta_j} \cdot (1-I(T_j)) \right]}{\sum_{j=1}^N [1-I(T_j)]} \quad (5)$$

$$RMSE = \frac{\sum_{j=1}^N \left[\left(\hat{e}_{j[i \notin I_{E,j}]} \right)^2 \cdot (1-I(T_j)) \right]}{\sum_{j=1}^N [1-I(T_j)]} \quad (6)$$

Eşitlik 5'te $\hat{\theta}_{j[i \notin I_{E,j}]}$, silme yapılmamış maddeler üzerinden kestirilen yetenek düzeyini; θ_j ise tüm maddeler üzerinden kestirilen yetenek düzeyini göstermektedir. Eşitlik 6'da $\hat{e}_{j[i \notin I_{E,j}]}$, silme yapılmamış maddeler üzerinden kestirilen yetenek düzeyinin hata değerini göstermektedir. $I(T_j)$, test alan j'nin yanıtları tahrifata uğramış ise 1'e, tahrifata uğramamış ise 0'a eşit olan gösterge fonksiyonudur. Tüm maddeler üzerinden yanlılık ve RMSE değerleri Eşitlik 5 ve Eşitlik 6'daki $\hat{\theta}_{j[i \notin I_{E,j}]}$ değeri $\hat{\theta}_j$ ile değiştirilerek hesaplanmaktadır.

Tahrifatlı test alanlarda, kestirimden uzaklaştırılan maddeler büyük çoğunlukla test alan tarafından yanlış cevaplanırken arda kalan maddeler ise büyük çoğunlukla doğru cevaplanmıştır. Bu nedenle θ kestiriminin arda kalan maddeler ile yapılmasının kestirimlerin pozitif yanlı olması (olduklarından daha büyük çıkması) ile sonuçlanabileceği düşünülmektedir. Ancak θ kestiriminde, tahrifatlı yanıtların da dâhil edilmesinden kaynaklı pozitif yanlı sonuçlanma miktarının, silinmemiş maddeler ile yapılan θ kestiriminden daha büyük yanlılık içermesi, θ 'nın sadece silinmemiş maddeler ile kestirilmesine gerekçe oluşturması beklenmektedir (Wollack vd., 2015).

Tüm veri setlerine ilişkin yanlılık ve RMSE analiz sonuçları Ek 3'te yer almaktadır. 50 ve 160 maddelik veri setlerinin tüm örneklem büyüklüklerinde yanlılık ve RMSE bulguları genel olarak benzerlik göstermektedir. Ek 3'te (Tablo 28, Tablo 30, Tablo 32 ve Tablo 34) yer alan yetenek düzeylerine göre ayrılmamış analizlerde her iki veri seti için de hiç silme yapılmamış test alanların yanlılık ve RMSE değerleri, silme yapılmış test alanların yanlılık ve RMSE değerleri ile kıyaslama yapılması amacıyla yer almaktadır.

Tahrifatsız test alanlar için 50 madde üzerinden yapılan analizlerde, iyicil silme sayısının 15 ve üzerinde; 160 madde üzerinden yapılan analizlerde ise 38 silme ve üzerinde olduğu durumlarda dahi yanlılığın önemsiz olduğu görülmektedir. Genel olarak 50 maddelik veri setleri için silme sayısı arttıkça yanlılığın az miktarda arttığı görülmektedir. 160 maddelik veri setleri için ise silme sayısı ve yanlılık arasında doğrudan bir ilişki görülmemekle birlikte silme sayısının az olduğu durumda çoğunlukla yanlılığın da az olduğu görülmektedir. Silme

sayısı arttıkça θ daha az madde ile kestirildiğinden standart hatanın artması ve buna bağlı olarak da RMSE değerlerinin artması beklenmektedir ancak 50 ve 160 maddelik veri setlerinde silme sayısı ile RMSE'ler arasında doğrudan bir ilişki görülmemektedir. Tahrifatsız test alanlar için yanlılık değerleri Wollack ve diğerleri (2015) ile benzerdir ancak RMSE değerlerinin yüksek olduğu görülmektedir.

Tahrifatlı test alanlar için ise, daha önce de belirtildiği üzere tahrifatlı test alanların θ kestirimlerinin, tahrifat yapılan maddelerin test alan tarafından yanlış yanıtlanan maddeler olması nedeniyle, silinmemiş maddeler üzerinden yetenek kestirimi yapılması durumunun pozitif yanlılığa neden olması beklenmektedir. Tüm maddeler üzerinden yapılan θ kestirimlerinin de benzer şekilde pozitif yanlılığa neden olması beklenmektedir. Her iki veri seti için de tahrifatlı test alanlarda beklendiği üzere pozitif yanlılık görülmektedir. Silmelerin sayısı arttıkça hem yanlılık hem de RMSE değerleri çoğunlukla artma eğilimindedir. Silinmiş maddelerin kestirimden uzaklaştırılması özellikle silme sayısı arttıkça daha az yanlılığa neden olmakta ancak RMSE değerlerinde belirgin bir farka neden olmamaktadır.

Yanlılık ve RMSE bulguları genel olarak 50 ve 160 madde üzerinden yapılan analizlerde benzerlik göstermektedir. 50 ve 160 madde üzerinden yapılan analizlerde aynı silme sayısına sahip test alanlar karşılaştırıldığında 50 maddelik veri setinin yanlılık ve RMSE değerlerinin 160 maddelik veri setine göre her zaman daha yüksek olduğu görülmektedir. 160 maddelik veri setinde kestirim yapılan madde sayısının daha fazla olması nedeniyle beklendik bir durumdur. Bulgular genel olarak Wollack ve diğerlerinin (2015) bulgularına benzerlik göstermekle birlikte Ek 3'te yer alan tahrifatsız test alanların RMSE değerleri daha yüksektir.

Farklı yetenek düzeylerindeki test alıcılar için silme yapılmamış maddeler üzerinden yetenek kestirimi yapılmasının etkisini incelemek amacıyla, yetenek düzeylerine göre hesaplanan yanlılık ve RMSE değerleri de Ek 3'te (Tablo 29, Tablo 31, Tablo 33 ve Tablo 35) yer almaktadır. Silme sayılarına göre yetenek düzeylerinde örneklem sayısının az olmasından dolayı silme sayıları gruplanarak raporlaştırılmıştır. Her iki veri setinde de yanlılık ve RMSE değerlerinin yetenek düzeyi ve silme sayısına bağlı olarak değişimi benzerlik göstermektedir.

Tahrifatsız test alanlarda aynı silme sayısı grubu içerisinde düşük yetenek düzeyindeki gruplar ile yüksek yetenek düzeyindeki grupların yanlılık düzeyleri arasında belirgin bir fark olmadığı ve yanlılığın önemsiz olduğu görülmektedir. Wollack ve diğerlerine (2015) benzer olarak yetenek düzeylerine göre silme sayısı arttıkça yanlılığın da genellikle artma eğiliminde olduğu, ayrıca yanlılığın silme sayısı gruplarının tamamında en düşük değerini orta yetenek düzeyinde aldığı, en üst ve en alt yetenek düzeylerinde ise yanlılığın en büyük değerlerini aldığı görülmektedir. Wollack ve diğerleri (2015) bu durumu Thissen ve Orlando'nun (2001) da

belirttiği gibi, MULTILOG ile yetenek kestirimlerinde ortalama sifıra doğru çekildikçe yanlılıkta daralma eğilimi beklenmesine dayandırmaktadır. RMSE'nin de silme sayısı gruplarının tamamında Wollack ve diğerleriyle (2015) uyumlu olarak yanlılığa benzer bir değişim gösterdiği görülmektedir. Yetenek düzeylerine göre silme sayısı ile RMSE değerleri arasında bir ilişki görülmemektedir.

Tahrifatlı test alanlar için Wollack ve diğerlerinde (2015) de görüldüğü gibi genel olarak yetenek düzeyi ve silme sayısı arttıkça yanlılık artmaktadır. RMSE değerlerindeki değişim ise tahrifatsız test alanlardakine benzerlik göstererek orta yetenek düzeyinde en düşük değerlerini almaktadır. Tahrifatlı test alanların RMSE değerleri aynı silme sayısı grubu içerisinde tahrifatsız test alanların RMSE değerlerinden büyük farklılık göstermemektedir.

Tahrifatlı ve tahrifatsız test alanlar için yapılan yanlılık ve RMSE analizleri birlikte değerlendirildiğinde tüm veri setleri için iyicil silmelerin θ kestiriminden uzaklaştırılmasının tahrifatsız test alanlarda negatif yanlılığa neden olduğu; tahrifatlı test alanlarda ise iyicil ve hileli silmelerin θ kestiriminden uzaklaştırılmasının pozitif yanlılığı düşürdüğü görülmektedir. θ 'nın kestiriminde silme yapılan maddeleri dışarda tutmak, yetenek kestiriminde tahrifatın pozitif yanlılık etkisini kısmen gidermektedir ve silmelerin sayısının fazla olduğu durumlarda dahi tahrifatsız test alıcılar üzerindeki etkisi önemsizdir. RMSE değerlerinin ise tahrifatsız test alanların RMSE değerlerinden büyük farklılık göstermediği görülmektedir. Tüm bulgular göz önünde bulundurulduğunda araştırma sürecinde Wollack ve diğerlerine (2015) benzer şekilde θ kestiriminde silme yapılmamış maddelerin kullanılması uygundur.

SBI Hesaplanması

Wollack ve diğerleri (2015) tarafından geliştirilen SBI hesaplamaları aşağıda yer almaktadır. Test tahrifatı durumlarındaki varsayımın, yanıtların yanlıştan doğruya değiştirilmesi olmasından dolayı, doğru puan sayısı olduğundan yüksek belirlenmektedir. Burada amaç, silinmiş maddelerdeki gözlenen doğru sayısı ile bu maddeler için beklenen doğru sayısı (silme yapılmamış olan maddelerdeki doğru sayısından kestirilen) arasında karşılaştırma yapmaktır.

Silinmiş maddelerdeki gözlenen değer, $X_{j,I_{E,j}}$, Eşitlik 7'deki gibi hesaplanmaktadır:

$$X_{j,I_{E,j}} = \sum_{i \in I_{E,j}} x_{ij} \quad (7)$$

Burada x_{ij} madde i 'de test alan j için doğru/yanlış puandır. $X_{j,I_{E,j}}$ normal binominal olarak modellenmektedir. Yine, silme yapılmamış maddelere dayalı olarak kestirilen yetenek $\hat{\theta}_{j[i \notin I_{E,j}]}$ ve madde parametre kestirimleri kullanılarak, silme yapılmış maddeler için beklenen doğru/yanlış puanı $E(X_{j,I_{E,j}})$, Eşitlik 8'deki gibi hesaplanmaktadır:

$$E(X_{j,I_{E,j}}) = \sum_{i \in I_{E,j}} P(x_{ij} = 1) \quad (8)$$

Dataya uygun olan madde tepki modeli, $P(x_{ij} = 1)$ kestirmek için kullanılabilir. Bu araştırmada STM kullanılmaktadır.

STM'ye göre θ yetenek düzeyindeki bir bireyin maddenin k seçeneğini yanıtlama olasılığı Eşitlik 9'da verilmiştir (de Ayala, 2009):

$$P_j(x = k | \theta, \alpha, \gamma) = \frac{e^{\gamma_{jk} + \alpha_{jk}\theta}}{\sum_{h=1}^{m_j} e^{\gamma_{jh} + \alpha_{jh}\theta}} \quad (9)$$

Eşitlik 9'da k seçeneği için α_{jk} eğim (lambda) ve γ_{jk} kesişim (zeta) parametreleridir. θ ise bireyin yetenek düzeyidir. k , her bir madde için doğru yanıtı, θ ise bireyin silme yapılmamış maddelerinden kestirilen yetenek düzeyini göstermek üzere, silme yapılmış her bir test alan için $E(X_{j,I_{E,j}})$, Eşitlik 8'den yararlanarak hesaplanmaktadır.

$X_{j,I_{E,j}}$ için standart hata Eşitlik 10'da verilmiştir:

$$SE(X_{j,I_{E,j}}) = \sqrt{\sum_{i \in I_{E,j}} P(x_{ij} = 1)[1 - P(x_{ij} = 1)]} \quad (10)$$

Yanıştan doğruya silmelerin sayısının, beklenen şans değerini aşıp aşmadığını belirlemek için kullanılan SBİ Eşitlik 11'de verilmiştir:

$$SBİ = \frac{X_{j,I,E,j} - E(X_{j,I,E,j}) + C}{SE(X_{j,I,E,j})} \quad (11)$$

Wollack (1997) benzer yapıdaki indekslerin (örneğin, yanıt kopyalama belirleme için ω istatistiği) test kısa olduğunda testi alan düşük yetenekli bireyler için şişirilmiş yalancı pozitif oranlara sahip olduğunu ifade etmektedir (Akt. van der Linden ve Sotaridona, 2006). Bu nedenle süreklilik için düzeltmenin uygulanmasının tüm koşullar için yalancı pozitif hata oranının daha iyi kontrol edilmesine yardımcı olacağı düşünülmüştür. Eşitlik 11’de yer alan C değeri, süreklilik için bir düzeltmedir. Süreklilik düzeltmesi yapılmadığında C değeri sıfır, süreklilik düzeltmesinin yapılması durumunda ise C değeri $-1/2$ ’ye eşit alınmaktadır (Wollack vd., 2015). SBİ değerleri her bir koşul için silme yapılan tüm test alanlar üzerinden hesaplanmıştır.

Testin İstatistiksel Gücünün ve I. Tip Hata Oranının Hesaplanması

İstatistiksel bir testin gücü, testin istatistiksel olarak anlamlı sonuç verebilme olasılığı (Cohen, 1988), alternatif hipotezin doğru olduğu durumlarda yokluk hipotezinin reddedilme olasılığı olarak da tanımlanmaktadır (Tabachnick ve Fidell, 2007). Güç, olasılık değeri olduğundan 0.00 ile 1.00 arasında değer almaktadır. Güç ne kadar yüksek olursa, bağımsız değişkenin gerçek etkisi o kadar hassas belirlenmektedir (Pagano, 2013). Cohen (1988), sağlıklı veriler elde edebilmek için gücün 0.80 civarında olmasını önerirken, Olejnik (1984) 0.70 ve 0.85 arasında değişen istatistiksel gücün kabul edilebilir olduğunu ifade etmektedir. Pagano (2013) ise gücün 0.80 veya daha yüksek bir değere sahip olmasının istendiğini ancak davranış bilimlerinde yaygın olarak 0.40 ile 0.60 arasında değerler görüldüğünü belirtmektedir. Bu çalışmada kabul edilebilir istatistiksel güç oranı 0.70 olarak belirlenmiştir.

Bu çalışma kapsamında istatistiksel güç oranı, tahrifatlı bireylerin doğru olarak tespit edilme oranı olarak tanımlanmıştır. SBİ için testin gücü, her bir kategoride, gerçekte tahrifatlı olan ve SBİ ile tahrifatlı olarak belirlenen test alan sayısı, tüm tahrifata uğrayan test alan sayısına bölünerek hesaplanmaktadır.

I. Tip hata oranı ise, tahrifata uğramadığı halde istatistiksel olarak anlamlı SBİ değerleri üreten test alanların sayısının, tahrifata uğramayan test alanların toplam sayısına bölünmesiyle elde edilmektedir. I. Tip hata oranı belirlenen manidarlık düzeyine göre (nominal düzey) yorumlanmaktadır. I. Tip hata oranlarının belirlenen nominal düzeyin altında olması beklenmektedir. Serlin (2000), I. Tip hata oranı nominal düzeyi aşarsa, elde edilen sonuçlara

yönelik önyargının artacağını ancak I. Tip hata oranının istenenden düşük olursa çalışmanın gerçek gücünün tehlikeye girebileceğini ifade etmektedir. I. Tip hatanın kabul edilebilir düzeyi için nominal düzeyin dışında alanyazında farklı sağlamlık kriterleri yer almaktadır (Bradley, 1978; Cochran, 1954). Sıklıkla kullanılan sağlamlık kriterlerinden biri olan Bradley'in (1978) esnek referans aralığına göre p değeri için $0.5\alpha \leq p \leq 1.5\alpha$ aralığı kabul edilebilir olarak tanımlanmaktadır.

Bu çalışmada testin istatistiksel gücü ve I. Tip hata oranı yedi farklı α seviyesinde (.000001, .0001, .0005, .001, .005, .01 ve .05) belirlenmektedir. I. Tip hata oranları öncelikle nominal düzeye göre, ardından Bradley'in esnek referans aralığına göre değerlendirilmektedir. I. Tip hata oranları kontrol edilen koşullar için istatistiksel güç yorumlanmaktadır.

Verilerin çözümlenmesinde, yetenek parametrelerinin kestirilmesi için MULTLOG programı; verilerin silme simülasyonunun yapılması, SBI'nin istatistiksel gücü ve I. Tip hatanın hesaplanması için R programında yazılan kodlar kullanılmıştır. Her bir analiz için 25 tekrar yapılmış ve ortalamaları raporlaştırılmıştır.

BÖLÜM 3

BULGULAR VE YORUMLAR

Bu bölümde, amaçlar bölümünde yer alan 1. ve 2. alt amaçtaki araştırma sorularına yönelik olarak, belirlenen örneklem ve madde sayıları için farklı α seviyelerinde, süreklilik düzeltilmesi uygulanan ve uygulanmayan SBİ'nin ürettiği istatistiksel güç ve I. Tip hata oranı bulguları ile yorumlarına yer verilmiştir.

İstatistiksel gücün doğru değerlendirilmesi için öncelikle I. Tip hata oranlarının bilinmesi gerekmektedir. İstatistiksel güç, ancak I. Tip hata oranının yüksek olmadığı koşullar için değerlendirilmelidir. Yetenek düzeyi yüksek olan gruplarda gözlenen doğru yanıt sayısı yüksek olacağından hedeflenen sayıda hileli silme işlemi yapılamaması durumu söz konusudur. Bu nedenle genel bir istatistiksel güç yorumu yapmak yerine tüm koşulların birlikte değerlendirilmesi gücün doğru yorumlanmasına olanak sağlamaktadır.

Bulgular örneklem büyüklüklerine göre ayrı başlıklar altında verilmiştir. Her bir örneklem büyüklüğü için öncelikle I. Tip hata oranlarına ilişkin bulgular daha sonra istatistiksel güç oranlarına ilişkin bulgular raporlaştırılmıştır.

Her bir örneklem büyüklüğünde; süreklilik düzeltilmesi uygulanmayan SBİ için I. Tip hata oranları tüm alfa düzeylerini içerecek şekilde madde sayılarına göre karşılaştırmalı olarak tabloleştirilmiştir. Tablolarda I. Tip hata oranları yetenek düzeylerine göre ayrılarak verilmiştir. Süreklilik düzeltilmesi uygulanan SBİ için de I. Tip hata oranları benzer şekilde tabloleştirilmiştir. I. Tip hata oranlarına ilişkin bulguların yorumlanmasında aşağıdaki yol izlenmiştir:

I. Tip hata oranlarının incelenmesinde öncelikle nominal düzey dikkate alınmış daha sonra Bradley'in esnek referansına göre de genel bir yorum yapılmıştır. Buna göre I. Tip hata oranları tüm alfa düzeylerinde;

- Süreklilik düzeltilmesi uygulanmaksızın kullanılan SBİ için
- Süreklilik düzeltilmesi uygulanarak kullanılan SBİ için
- Madde sayılarına göre süreklilik düzeltilmesi kullanılmayan ve kullanılan durumlarda SBİ için
- Örneklem büyüklüklerine göre süreklilik düzeltilmesi kullanılmayan ve kullanılan durumlarda SBİ için

tüm koşullarda karşılaştırmalı olarak incelenmiştir.

Her bir örneklem büyüklüğünde; süreklilik düzeltmesi uygulanmayan SBİ için istatistiksel güç oranları tüm alfa düzeylerini içerecek şekilde madde sayılarına göre karşılaştırmalı olarak tablolaştırılmıştır. Tablolarda istatistiksel güç oranları tahrifat oranlarına bağlı olarak yetenek düzeylerine göre ayrılarak verilmiştir. Süreklilik düzeltmesi uygulanan SBİ için de istatistiksel güç oranları benzer şekilde tablolaştırılmıştır. İstatistiksel güç oranlarına ilişkin bulguların yorumlanmasında aşağıdaki yol izlenmiştir:

İstatistiksel gücün incelenmesinde öncelikle nominal düzey dikkate alınmış daha sonra Bradley'in esnek referansına göre de genel bir yorum yapılmıştır. Buna göre istatistiksel güç oranları tüm alfa düzeylerinde;

- Süreklilik düzeltmesi uygulanmaksızın kullanılan SBİ için:
 - Tahrifat oranlarına göre ayrılmaksızın, tüm örnekleme ve yetenek düzeylerine göre,
 - Tahrifat oranlarına göre ayrılarak tüm örnekleme ve yetenek düzeylerine göre,
- Süreklilik düzeltmesi uygulanarak kullanılan SBİ için:
 - Tahrifat oranlarına göre ayrılmaksızın, tüm örnekleme ve yetenek düzeylerine göre,
 - Tahrifat oranlarına göre ayrılarak tüm örneklem ve yetenek düzeylerine göre,
- Madde sayılarına göre süreklilik düzeltmesi kullanılmayan ve kullanılan SBİ için
- Örneklem büyüklüklerine göre süreklilik düzeltmesi kullanılmayan ve kullanılan SBİ için

tüm koşullarda karşılaştırmalı olarak incelenmiştir.

Bu bölümde yer alan I. Tip hata oranlarına ilişkin tablolarda; I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik yazılarak belirtilmiştir. SBİ'nin gücünün sadece I. Tip hata oranının kontrol edildiği koşullarda dikkate alınması gerekmektedir. Bu nedenle istatistiksel güce ilişkin tablolarda ilgili koşullar, I. Tip hata oranlarına ilişkin tablolara benzer şekilde belirtilmiştir.

500 Örneklem Büyüklüğü İçin SBİ'nin İstatistiksel Güç ve I. Tip Hata Bulguları

500 örneklem büyüklüğü için I. Tip hata oranlarına ilişkin bulgular Tablo 10 ve Tablo 11'de, istatistiksel güce ilişkin bulgular ise Tablo 12 ve Tablo 13'te yer almaktadır.

Tablo 10

Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi Yapılmaksızın SBI'nin I. Tip Hataları (n= 500)

50 Madde								160 Madde							
Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
1	<u>0.040</u>	0.028	0.0200	0.0124	0.00905	0.00476	0.004762	1	0.134	0.088	0.0733	0.0490	0.04429	0.02905	0.013333
2	0.099	0.057	0.0443	0.0262	0.01667	0.01000	0.005238	2	0.251	0.158	0.1267	0.0867	0.07333	0.05429	0.035714
3	0.132	0.056	0.0381	0.0186	0.01476	0.00762	0.002381	3	0.277	0.138	0.1024	0.0619	0.05238	0.03000	0.016190
4	0.112	0.034	0.0223	0.0095	0.00636	0.00091	<u>0.000000</u>	4	0.284	0.116	0.0827	0.0473	0.03591	0.02318	0.012273
5	<u>0.050</u>	<u>0.011</u>	<u>0.0073</u>	<u>0.0009</u>	0.00091	<u>0.00000</u>	<u>0.000000</u>	5	0.237	0.070	0.0509	0.0345	0.03182	0.01591	0.007273
Toplam	0.087	0.037	0.0262	0.0134	0.00944	0.00458	0.002430	Toplam	0.237	0.113	0.0868	0.0556	0.04729	0.03028	0.016822

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 10 incelendiğinde 50 maddelik veri setinde süreklilik düzeltilmesi yapılmaksızın I. Tip hata oranları, yetenek düzeyi ayrımı yapılmadan tüm test alanlar üzerinden incelendiğinde, tüm alfa seviyelerinde manidarlık düzeyini aştığı görülmektedir. Bu durum I. Tip hata oranlarının iyi kontrol edilemediğini dolayısıyla aslında tahrifatlı olmadığı halde tahrifatlı olarak belirlenen test alanların oranının alfa düzeyini aştığını göstermektedir. Benzer şekilde Qin (2016), $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ manidarlık düzeylerinde yürüttüğü çalışmada, tüm araştırma koşullarında I. Tip hatanın belirlenen alfa seviyelerini aştığını belirtmiştir. Wollack ve diğerleri (2015) ise yalnızca düşük alfa seviyelerinde hata oranlarının kontrol edilemediğini belirtmiştir.

I. Tip hata oranlarının yetenek düzeylerine göre değişimi incelendiğinde ise 1. yetenek düzeyinin en büyük alfa seviyesinde; 4. yetenek düzeyinde ise en küçük alfa seviyesinde I. Tip hata oranının manidarlık düzeyini aşmadığı, hatanın iyi kontrol edildiği görülmektedir. 5. yetenek düzeyinde ise $\alpha = .05$, $\alpha = .001$, $\alpha = .0001$ ve $\alpha = .00001$ düzeyleri için I. Tip hata oranlarının manidarlık düzeylerini aşmadığı görülmektedir. İlgili koşullar için tahrifatlı olmadığı halde tahrifatlı olarak belirlenen test alanların oranı alfa düzeyini aşmamaktadır. 2. ve 3. yetenek düzeyinin hiçbir alfa seviyesinde hata oranları kontrol edilememektedir. Wollack ve diğerleri (2015) de en düşük yetenek düzeyinde I. Tip hatanın yalnızca $\alpha = .05$ düzeyinde kontrol edilebildiğini ancak, Tablo 10'un aksine, diğer tüm yetenek düzeylerinde I. Tip hatanın iyi kontrol edilebildiğini belirtmiştir. Sinharay ve Johnson (2016) araştırmasında, tahrifatlı test alan yüzdesini %50 olarak belirleyip, I. Tip hata oranlarını dizi sonu ve tesadüfi silmelere göre ayrı ayrı hesaplamıştır. Dizi sonu silmeleri için Tablo 10 bulgularına benzer şekilde orta ve düşük yetenek düzeylerinde hiçbir alfa düzeyinde hatanın kontrol edilemediğini belirtmiştir. Tesadüfi silmeler için ise Tablo 10 bulgularının aksine, 1. yetenek düzeyinin en yüksek alfa seviyesi dışındaki tüm seviyelerde I. Tip hata oranlarının nominal seviyeyi aşmadığı görülmektedir. Sinharay ve Johnson (2016) tarafından iyicil silme türlerine göre ayrılarak hesaplanan I. Tip hata oranlarının birbirinden oldukça farklılaştığı göz önünde bulundurulduğunda, Sinharay ve Johnson'ın bulgularını iyicil silme türlerine göre ayrılmadan hesaplanan I. Tip hata oranı bulgularıyla doğrudan karşılaştırmanın yanıltıcı olabileceği düşünülmektedir. Ayrıca Sinharay ve Johnson'ın tahrifatlı birey yüzdesinin (%50), araştırmadaki tahrifatlı birey yüzdesinden (yaklaşık %14) oldukça yüksek olmasının I. Tip hata oranlarını etkileyebileceği düşünülmektedir. Tablo 10 incelendiğinde I. Tip hata oranlarının alfa seviyelerinin çoğunu aştığı görülmektedir. Uygulamada test tahrifatının çoğunlukla düşük

yetenekli bireylerde gerçekleşmesi beklenmektedir. I. Tip hatanın düşük yetenek düzeylerinde iyi kontrol edilememesi Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) da belirttiği üzere istenmeyen bir durumdur.

160 Maddelik Veri Seti

160 maddelik veri seti üzerinden hesaplanan I. Tip hata oranlarının ise tüm yetenek düzeyleri için tüm alfa seviyelerinde manidarlık düzeylerini aştığı ve I. Tip hata oranlarının iyi kontrol edilemediği görülmektedir. 50 maddelik veri setinde, 160 maddelik veri setine göre I. Tip hata oranlarının daha iyi kontrol edildiği görülmektedir.

I. Tip hatanın değerlendirilmesi için Bradley'in (1978) esnek referansı kullanıldığında 50 maddelik veri setinde, nominal seviyeye göre kabul edilebilir olan düşük alfa seviyelerindeki I. Tip hata oranlarının referans aralığının dışında kaldığı görülmektedir. Bununla birlikte 5. yetenek düzeyinde $\alpha = .01$ ve $\alpha = .005$ seviyelerinde I. Tip hatanın kontrol edildiği görülmektedir. I. Tip hatanın değerlendirilmesinde Bradley'in esnek referans aralığının kullanılması 5. yetenek düzeyi için, en küçük alfa seviyelerinde hata oranlarının kabul edilebilirken daha büyük alfa seviyelerinde hata oranlarının kontrol edilemediği durumları ortadan kaldırarak, büyük alfa seviyelerinde hata kontrolü sağlanırken, küçük alfa seviyelerinde hata kontrolü sağlanamaz olarak yorumlanmaktadır. Qin (2016) süreklilik düzeltmesi uygulanmadığı durumda hata oranlarının Bradley'in esnek referansına göre değerlendirildiğinde hiçbir koşulda I. Tip hatanın kontrol edilemediğini ifade etmiştir. SBI'nin birden fazla yanlıştan doğruya silmesi olanlar için kullanılan versiyonu olarak tanımladığı EDI_WTR kullanıldığında ise $\alpha = .05$ düzeyinde hiçbir koşulda I. Tip hatanın kontrol edilemediğini ancak $\alpha = .01$ düzeyinde ise tüm koşullarda I. Tip hatanın kontrol edildiğini belirtmiştir. 160 maddelik veri setinde ise nominal seviye referans alındığında elde edilen bulgulara benzer şekilde, I. Tip hata oranlarının tüm koşullarda referans aralığının dışında olduğu görülmektedir.

Aynı veri setlerine süreklilik düzeltmesi yapılarak hesaplanan I. Tip hata oranları Tablo 11'de yer almaktadır.

Tablo 11

Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi İle SBI'nin I. Tip Hataları (n= 500)

50 Madde								160 Madde							
Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
1	<u>0.008</u>	<u>0.006</u>	<u>0.0057</u>	0.0024	0.00238	0.00143	0.001429	1	0.055	0.021	0.0152	0.0090	0.00810	0.00238	0.001429
2	<u>0.028</u>	<u>0.010</u>	<u>0.0062</u>	0.0043	0.00333	0.00190	0.000952	2	0.104	0.054	0.0443	0.0324	0.02667	0.01571	0.010000
3	<u>0.039</u>	<u>0.011</u>	<u>0.0033</u>	0.0024	0.00095	<u>0.00000</u>	<u>0.000000</u>	3	0.098	0.052	0.0462	0.0281	0.02429	0.01810	0.009048
4	<u>0.030</u>	<u>0.012</u>	0.0082	0.0018	<u>0.00000</u>	<u>0.00000</u>	<u>0.000000</u>	4	0.098	0.050	0.0427	0.0305	0.02409	0.01500	0.003636
5	<u>0.013</u>	<u>0.006</u>	<u>0.0009</u>	<u>0.0005</u>	<u>0.00000</u>	<u>0.00000</u>	<u>0.000000</u>	5	0.088	0.044	0.0350	0.0295	0.02091	0.01182	0.002727
Toplam	<u>0.023</u>	<u>0.009</u>	<u>0.0049</u>	0.0022	0.00131	0.00065	0.000467	Toplam	0.089	0.044	0.0367	0.0260	0.02084	0.01262	0.005327

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 11'e göre süreklilik düzeltmesinin, bazı koşullarda I. Tip hata oranının iyi kontrol edilmesini sağladığı görülmektedir. 50 maddelik veri setinde süreklilik düzeltmesi yapılarak hesaplanan I. Tip hata oranları, yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden incelendiğinde daha büyük alfa seviyelerinde manidarlık düzeyini aşmadığı ve I. Tip hatanın iyi kontrol edildiği görülmektedir. Alfa seviyesi azaldıkça manidarlık düzeyini aşmakta ve I. Tip hata kontrol edilememektedir. Wollack ve diğerleri (2015) tüm test alanlar üzerinden süreklilik düzeltmesi ile hesaplanan I. Tip hata oranlarının tüm alfa seviyelerinde kontrol edildiğini; Sinharay ve diğerleri (2017) de Tablo 11 bulgularına çoğunlukla benzer şekilde $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ seviyelerinde hata oranlarının kontrol edildiğini belirtmiştir. Qin (2016) süreklilik düzeltmesi uygulanan SBI'nin I. Tip hata oranlarını raporlaştırmamış olup süreklilik düzeltmesi ile uygulanan EDI_WTR'nin I. Tip hata oranlarını raporlaştırmıştır. Qin, EDI_WTR'nin hiçbir koşulda ve hiçbir alfa seviyesinde I. Tip hatayı kontrol edemediğini belirtmiştir ancak EDI_WTR, SBI'nin özel bir kullanımı olduğu için bulgularının doğrudan karşılaştırılmasının yanıltıcı olabileceği düşünülmektedir.

I. Tip hata oranlarının yetenek düzeylerine göre değişimi incelendiğinde ise $\alpha = .05$ için tüm yetenek düzeylerinde I. Tip hatanın iyi kontrol edildiği görülmektedir. Ancak düşük yetenek düzeylerinde, 1. ve 2. yetenek düzeyinde, büyük alfa seviyelerinde I. Tip hata oranları manidarlık düzeyini aşmazken alfa seviyesi azaldıkça I. Tip hatanın iyi kontrol edilemediği görülmektedir. 3. ve 4. yetenek düzeyinde en düşük alfa seviyelerinde I. Tip hata oranlarının iyi kontrol edildiği görülmektedir. 5. yetenek düzeyinde ise tüm alfa seviyelerinde I. Tip hata oranlarının iyi kontrol edildiği görülmektedir. I. Tip hatanın düşük yetenek düzeylerinde düşük alfa seviyelerinde iyi kontrol edilememesi problemleri bir durum olmakla birlikte süreklilik düzeltmesi ile iki alfa seviyesinde I. Tip hatanın kontrol edilmesi önemlidir. Wollack ve diğerlerinin (2015) yetenek düzeylerine göre I. Tip hata kontrolü incelendiğinde, en düşük yetenek düzeyinin en düşük alfa seviyesi dışında tüm yetenek düzeyleri ve alfa seviyelerinde hata oranlarının iyi kontrol edildiği görülmektedir. Sinharay ve Johnson'ın (2016) bulgularında dizi sonu silmeleri için, $\alpha = .05$, $\alpha = .01$, $\alpha = .001$ düzeylerinde tüm yetenek düzeylerinde, $\alpha = .0001$ düzeyinde ise 3. yetenek düzeyi dışında tüm yetenek düzeylerinde; tesadüfi silmeler için ise tüm yetenek düzeylerinde ve tüm alfa seviyelerinde I. Tip hatanın kontrol edildiği görülmektedir. Sinharay ve Johnson (2016) tarafından süreklilik düzeltmesi ile iyicil silme türlerine göre ayrılarak hesaplanan I. Tip hata oranlarının birbirine oldukça benzer olduğu, yalnızca $\alpha = .0001$ seviyesinde farklılaştığı göz önünde bulundurulduğunda, araştırma

bulgularının Sinharay ve Johnson'ın bulgularıyla karşılaştırılmasının uygun olduğu düşünülmektedir. Tablo 11'de yer alan bulgular, Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) bulgularından en büyük alfa seviyeleri dışında farklılaşmaktadır.

160 Maddelik Veri Seti

160 maddelik veri seti üzerinden hesaplanan I. Tip hata oranlarının süreklilik düzeltmesi yapıldığı halde tüm yetenek düzeyleri için tüm manidarlık düzeylerinde, alfa seviyesini aştığı ve I. Tip hata oranlarının iyi kontrol edilemediği görülmektedir. Sinharay ve diğerlerinin (2017) 100 madde üzerinden yürüttükleri çalışmalarında I. Tip hata oranlarının $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ düzeylerinde kontrol edildiği görülmektedir. Madde sayısının yüksek olması dışında, Sinharay ve diğerlerinin (2017) simülasyon deseninin örneklem büyüklüğü ve tahrifata uğrayan birey oranı açısından, yürütülen araştırmanın simülasyon deseninden farklı olmasının araştırma bulgularındaki farklılıkla ilişkilendirilebileceği düşünülmektedir. Süreklilik düzeltmesi yapılarak hesaplanan I. Tip hata oranlarında da 50 maddelik veri setinde, 160 maddelik veri setine göre hata oranları daha iyi kontrol edilmektedir. Süreklilik düzeltmesi, beklendiği gibi, tüm koşullarda I. Tip hata oranının düşmesine neden olmaktadır.

I. Tip hatanın değerlendirilmesinde Bradley'in (1978) esnek referansı kullanıldığında 50 maddelik veri setinde, nominal seviyeye göre kabul edilebilir olan düşük alfa seviyelerindeki ($\alpha = .0005$, $\alpha = .0001$ ve $\alpha = .00001$) I. Tip hata oranlarının kontrol edilemediği olduğu görülmektedir. Bununla birlikte $\alpha = .01$ ve $\alpha = .005$ seviyelerinde kabul edilebilir hata oranlarının sayısı artmaktadır. I. Tip hatanın değerlendirilmesinde Brandley'in esnek referans aralığının kullanılması, en düşük alfa seviyelerinde hata oranlarının kabul edilebilirken daha yüksek alfa seviyelerinde hata oranlarının kontrol edilememesi olması gibi durumları çoğunlukla ortadan kaldırarak, alfa seviyesi düştükçe hata oranlarının kontrol edilemediği şeklinde yorumlanmasına neden olmuştur. Bradley'in esnek referansı kullanıldığında süreklilik düzeltmesi uygulanan EDI_WTR'nin I. Tip hatayı yalnızca $\alpha = .05$ düzeyinde kontrol ettiği belirtilmiştir (Qin, 2016). Süreklilik düzeltmesi uygulandığı halde I. Tip hata oranları kontrol edilemeyen 160 maddelik veri setinin Bradley'in esnek referansına göre değerlendirilmesi sonucunda ise I. Tip hata oranlarının yalnızca en düşük yetenek düzeyinin en yüksek alfa seviyesinde kabul edilebilir referans aralığı içerisinde olduğu görülmektedir.

500 test alan için SBİ'nin süreklilik düzeltmesi yapılmaksızın güç oranları Tablo 12'de yer almaktadır.

Tablo 12

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeyine Göre İstatistiksel Güç Oranları (n= 500)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
%10 (5 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%10 (16 Silme)	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	1.000	1.000	1.000	0.980	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	1.000	0.980	0.980	0.880	0.690	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	1.000	0.970	0.930	0.720	0.630	0.380	<u>0.140</u>	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.980</u>	<u>0.640</u>	<u>0.450</u>	<u>0.180</u>	0.120	<u>0.050</u>	<u>0.010</u>	5	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.830
Toplam		0.996	0.922	0.876	0.776	0.746	0.662	0.564	Toplam		1.000	1.000	1.000	1.000	1.000	0.998	0.966
%20 (10 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%20 (32 Silme)	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	1.000	1.000	1.000	1.000	0.980	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	1.000	1.000	0.990	0.970	0.930	0.790	<u>0.530</u>	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>1.000</u>	<u>0.950</u>	<u>0.850</u>	<u>0.490</u>	0.390	<u>0.140</u>	<u>0.040</u>	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Toplam		1.000	0.990	0.968	0.892	0.864	0.786	0.710	Toplam		1.000	1.000	1.000	1.000	1.000	1.000	1.000
%30 (15 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%30 (48 Silme)	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	1.000	1.000	1.000	1.000	1.000	0.960	<u>0.800</u>	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>1.000</u>	<u>1.000</u>	<u>0.990</u>	<u>0.680</u>	0.480	<u>0.130</u>	<u>0.030</u>	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Toplam		1.000	1.000	0.998	0.936	0.896	0.818	0.766	Toplam		1.000	1.000	1.000	1.000	1.000	1.000	1.000
Değişken Silme	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	Değişken Silme	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	0.990	0.990	0.950	0.900	2	1.000	1.000	1.000	1.000	1.000	0.990	0.980	
	3	0.910	0.800	0.700	0.560	0.440	0.330	0.250	3	0.980	0.890	0.850	0.720	0.660	0.570	0.420	
	4	—	—	—	—	—	—	—	4	—	—	—	—	—	—	—	
	5	—	—	—	—	—	—	—	5	—	—	—	—	—	—	—	
Toplam		0.970	0.933	0.900	0.850	0.810	0.760	0.717	Toplam		0.993	0.963	0.950	0.907	0.887	0.853	0.800

(Devam ediyor)

Tablo 12 (Devam)

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeyine Göre İstatistiksel Güç Oranları (n= 500)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
Toplam	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	Toplam	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	0.998	0.998	0.988	0.970		2	1.000	1.000	1.000	1.000	1.000	0.998	0.995
	3	0.978	0.950	0.925	0.885	0.855	0.803	0.730		3	0.995	0.973	0.963	0.930	0.915	0.893	0.855
	4	1.000	0.990	0.973	0.897	0.853	0.710	<u>0.490</u>		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.993</u>	<u>0.863</u>	<u>0.763</u>	<u>0.450</u>	0.330	<u>0.107</u>	<u>0.027</u>		5	1.000	1.000	1.000	1.000	1.000	0.997	0.943
	Toplam	0.994	0.964	0.939	0.865	0.831	0.756	0.686		Toplam	0.999	0.994	0.992	0.984	0.981	0.975	0.957

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

İstatistiksel güç tahrifat miktarına göre ayrılmaksızın toplam test alanlar üzerinden incelendiğinde, 1. yetenek düzeyinde $\alpha = .05$ düzeyinde tahrifatlı bireylerin mükemmel düzeyde, 5. yetenek düzeyinde ise mükemmel yakın düzeyde belirlendiği görülmektedir. 5. yetenek düzeyi için alfa seviyesi küçüldükçe istatistiksel gücün kabul edilemez düzeyde düştüğü görülmektedir. 4. yetenek düzeyinde ise I. Tip hatanın en düşük alfa seviyesinde kontrol edilebildiği ancak bu düzeyde de istatistiksel gücün kabul edilebilir düzeyde olmadığı görülmektedir.

İstatistiksel güç tahrifat miktarına göre incelendiğinde tahrifat oranı arttıkça her yetenek düzeyinde gücün arttığı görülmektedir. %10 oranında tahrifat (5 silme) yapıldığı durumda $\alpha = .05$ düzeyinde 1. yetenek düzeyinde tahrifatlı bireyler mükemmel düzeyde, 5. yetenek düzeyinde ise mükemmel yakın düzeyde belirlenmektedir. %20 oranında (10 silme) ve %30 oranında tahrifat (15 silme) yapıldığı durumlarda $\alpha = .05$ düzeyinde 1. yetenek düzeyinde istatistiksel gücün mükemmel olduğu görülmektedir. %20 oranında tahrifat (10 silme) yapıldığı durumda diğer alfa seviyelerinde istatistiksel gücün kabul edilebilir düzeyde olmadığı görülürken; %30 oranında tahrifat (15 silme) yapıldığı durumda ise 4. yetenek düzeyinde en düşük alfa seviyesinde istatistiksel gücün tatmin edici düzeyde olduğu, 5. yetenek düzeyinde en büyük alfa seviyesi dışında kabul edilebilir istatistiksel güç elde edilemediği görülmektedir. Değişken sayıda silme yapıldığı durumda da 1. yetenek düzeyinde $\alpha = .05$ düzeyinde istatistiksel güç mükemmeldir. Yüksek yetenek düzeylerinde ise tahrifata uğrayan test alan olmadığı için istatistiksel güç hesaplanmamıştır. Benzer araştırmalarda da olduğu gibi, silme oranı arttıkça istatistiksel gücün arttığı ancak alfa seviyesi azaldıkça istatistiksel gücün düştüğü görülmektedir (Qin, 2016; Sinharay vd., 2017; Wollack vd., 2015). Ayrıca Wollack ve diğerlerinin (2015) de belirttiği üzere yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir.

160 Maddelik Veri Seti

160 maddelik veri seti için yetenek düzeylerinin hiçbirinde, hiçbir alfa seviyesinde I. Tip hata kontrol edilemediği için istatistiksel güç dikkate alınmamaktadır.

Tablo 12’de yer alan istatistiksel güç değerleri Bradley’in esnek referans aralığına göre incelendiğinde ise; en düşük alfa seviyelerindeki istatistiksel güç değerlerinin I. Tip hata

oranlarının kontrol edilememesi nedeniyle deęerlendirme dıřında kaldığı grlmektedir. En dřk alfa seviyelerinde oęunlukla istatistiksel gcn olduka dřk olduęu gz nnde bulundurulduęunda esnek referans aralıęına gre istatistiksel gcn deęerlendirilmesinin %30 oranında tahrifat (15 silme) yapıldığı durumda 4. yetenek dzeyinin en dřk alfa seviyesi dıřında, kabul edilebilir istatistiksel g saęlayan kořul sayısını olumsuz etkilemedięi grlmektedir. Aksine %10 oranında tahrifat (5 silme) kořulu dıřında, 5. yetenek dzeyi iin kabul edilebilir istatistiksel gcn $\alpha = .01$ ve $\alpha = .005$ dzeylerinde, saęlandıęı belirlenmiřtir.

500 test alan iin sreklilik dzeltmesi yapılarak uygulanan SBI'nin istatistiksel gc Tablo 13'te yer almaktadır.



Tablo 13

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBI'nin Yetenek Düzeyine Göre İstatistiksel Güç Oranları (n= 500)

		50 Madde							160 Madde										
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$		
%10 (5 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%10 (16 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000		
	2	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	0.980	0.880		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	3	<u>1.000</u>	<u>0.990</u>	<u>0.980</u>	0.910	0.860	<u>0.710</u>	<u>0.430</u>		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	4	<u>0.970</u>	<u>0.850</u>	<u>0.730</u>	0.500	<u>0.330</u>	<u>0.140</u>	<u>0.060</u>		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.770</u>	<u>0.310</u>	<u>0.200</u>	<u>0.060</u>	<u>0.050</u>	<u>0.030</u>	<u>0.000</u>		5	1.000	1.000	1.000	1.000	1.000	1.000	0.970	0.720	
	Toplam	<u>0.948</u>	<u>0.830</u>	<u>0.782</u>	0.694	0.648	0.572	0.474		Toplam	1.000	1.000	1.000	1.000	1.000	0.994	0.944		
%20 (10 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%20 (32 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000		
	2	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	3	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	<u>1.000</u>	<u>0.940</u>		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	4	<u>1.000</u>	<u>0.990</u>	<u>0.990</u>	0.910	<u>0.860</u>	<u>0.600</u>	<u>0.340</u>		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.990</u>	<u>0.780</u>	<u>0.640</u>	<u>0.290</u>	<u>0.180</u>	<u>0.060</u>	<u>0.020</u>		5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Toplam	<u>0.998</u>	<u>0.954</u>	<u>0.926</u>	0.840	0.808	0.732	0.660		Toplam	1.000	1.000	1.000	1.000	1.000	1.000	1.000		
%30 (15 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%30 (48 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000		
	2	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	3	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	<u>1.000</u>	<u>1.000</u>		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	4	<u>1.000</u>	<u>1.000</u>	1.000	0.990	<u>0.970</u>	<u>0.890</u>	<u>0.690</u>		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>1.000</u>	<u>0.960</u>	<u>0.860</u>	<u>0.450</u>	<u>0.240</u>	<u>0.080</u>	<u>0.010</u>		5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Toplam	<u>1.000</u>	<u>0.992</u>	<u>0.972</u>	0.888	0.842	0.794	0.740		Toplam	1.000	1.000	1.000	1.000	1.000	1.000	1.000		
Değişken Silme	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	Değişken Silme	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000		
	2	<u>1.000</u>	<u>0.990</u>	<u>0.990</u>	0.950	0.920	0.880	0.830		2	1.000	1.000	1.000	0.990	0.990	0.980	0.960		
	3	<u>0.620</u>	<u>0.440</u>	<u>0.350</u>	0.280	0.260	<u>0.190</u>	<u>0.120</u>		3	0.830	0.750	0.650	0.540	0.470	0.390	0.290		
	4	—	—	—	—	—	—	—		4	—	—	—	—	—	—	—	—	
	5	—	—	—	—	—	—	—		5	—	—	—	—	—	—	—	—	
	Toplam	<u>0.873</u>	<u>0.810</u>	<u>0.780</u>	0.743	0.727	0.690	0.650		Toplam	0.943	0.917	0.883	0.843	0.820	0.790	0.750		

(Devam ediyor)

Tablo 13 (Devam)

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBİ'nin Yetenek Düzeyine Göre İstatistiksel Güç Oranları (n= 500)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
Toplam	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	Toplam	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000
	2	<u>1.000</u>	<u>0.998</u>	<u>0.998</u>	0.988	0.980	0.965	0.928		2	1.000	1.000	1.000	0.998	0.998	0.995	0.990
	3	<u>0.905</u>	<u>0.858</u>	<u>0.833</u>	0.798	0.780	<u>0.725</u>	<u>0.623</u>		3	0.958	0.938	0.913	0.885	0.868	0.848	0.823
	4	<u>0.990</u>	<u>0.947</u>	0.907	0.800	<u>0.720</u>	<u>0.543</u>	<u>0.363</u>		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.920</u>	<u>0.683</u>	<u>0.567</u>	<u>0.267</u>	<u>0.157</u>	<u>0.057</u>	<u>0.010</u>		5	1.000	1.000	1.000	1.000	1.000	0.990	0.907
	Toplam	<u>0.964</u>	<u>0.906</u>	<u>0.874</u>	0.797	0.759	0.698	0.629		Toplam	0.991	0.986	0.981	0.974	0.970	0.963	0.943

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

İstatistiksel güç tahrifat miktarına ve yetenek düzeylerine göre ayrılmaksızın toplam test alanlar üzerinden incelendiğinde, alfa seviyesi düştükçe istatistiksel güç azalmakla birlikte, en büyük üç alfa seviyesinde istatistiksel gücün oldukça yüksek olduğu görülmektedir. İstatistiksel güç tahrifat miktarına göre ayrılmaksızın toplam test alanlar üzerinden incelendiğinde, 1. ve 2. yetenek düzeyinde $\alpha = .05$ ve $\alpha = .01$ düzeylerinde tahrifatlı bireylerin mükemmel düzeyde belirlendiği görülmektedir ancak en düşük yetenek düzeylerindeki bu iki grup için daha düşük alfa seviyelerinde hatalar iyi kontrol edilemediği için istatistiksel güç değerlendirilememiştir. 3. yetenek düzeyinde daha büyük alfa seviyelerinde istatistiksel gücün oldukça yüksek olduğu görülürken, $\alpha = .0001$ gibi katı alfa seviyesinde bile gücün kabul edilebilir seviyede olduğu belirlenmiştir. 4. yetenek düzeyinde $\alpha = .05$ seviyesinde gücün neredeyse mükemmel olduğu, alfa seviyesi katılaştıkça istatistiksel gücün düştüğü ancak $\alpha = .0005$ düzeyinde bile gücün kabul edilebilir seviyede olduğu belirlenmiştir. 5. yetenek düzeyinde $\alpha = .05$ seviyesinde istatistiksel güç yüksek olmakla birlikte, tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edildiği ancak alfa seviyesi katılaştıkça diğer alfa seviyelerinde istatistiksel gücün kabul edilemez düzeyde düştüğü görülmektedir.

İstatistiksel güç tahrifat miktarına göre incelendiğinde ise Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) da belirttiği üzere tahrifat oranı arttıkça her yetenek düzeyinde istatistiksel gücün arttığı görülmektedir.

%10 oranında tahrifat (5 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde; istatistiksel gücün $\alpha = .05$ seviyesinde oldukça yüksek olduğu, $\alpha = .005$ seviyesinde ise kabul edilebilir düzeyde olduğu belirlenmiştir. Sinharay ve diğerlerinin (2017) 4 silme yaparak hesapladıkları istatistiksel güç $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ düzeylerinin hiç birinde kabul edilebilir değildir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde; $\alpha = .05$ ve $\alpha = .01$ düzeylerinde 1. ve 2. yetenek düzeylerinde tahrifatlı bireylerin mükemmel düzeyde belirlendiği görülmektedir. 3. yetenek düzeyinde büyük alfa seviyelerinde istatistiksel gücün oldukça yüksek olduğu görülürken, $\alpha = .0001$ gibi düşük alfa seviyesinde bile gücün kabul edilebilir düzeyde olduğu belirlenmiştir. 4. yetenek düzeyinde en büyük alfa seviyesinde istatistiksel güç çok yüksek iken alfa seviyesi katılaştıkça istatistiksel gücün kabul edilemez düzeyde düştüğü görülmektedir. 5. yetenek düzeyinde $\alpha = .05$ seviyesinde kabul edilebilir düzeydedir ancak diğer tüm alfa seviyelerinde I. Tip hata iyi kontrol edilmesine rağmen istatistiksel gücün kabul edilemez olduğu görülmektedir. Wollack ve diğerleri (2015) tarafından 5 silme yapıldığı durumda yalnızca 1. yetenek düzeyinde ve $\alpha = .05$

düzeyinde çok yüksek istatistiksel güç; 1. yetenek düzeyi için $\alpha = .01$ ve 2. yetenek düzeyi için $\alpha = .05$ düzeylerinde kabul edilebilir istatistiksel güç rapor edilmiştir. Araştırma bulguları, üst yetenek düzeylerinde büyük alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilmesi ile Wollack ve diğerlerinden (2015) farklılık göstermektedir. Sinharay ve Johnson'ın (2016) ise 5 silme yapıldığı durumda hiçbir yetenek düzeyinde ve alfa seviyesinde istatistiksel gücün kabul edilebilir olmadığı görülmektedir.

%20 oranında tahrifat (10 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel güç en büyük üç alfa seviyesinde de oldukça yüksektir. Sinharay ve diğerlerinin (2017) ise 8 silme yaparak hesapladıkları istatistiksel gücün yalnızca $\alpha = .05$ düzeyinde kabul edilebilir olduğu görülmektedir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde $\alpha = .05$ ve $\alpha = .01$ düzeylerinde 1. ve 2. yetenek düzeylerinde tahrifatlı bireyler mükemmel düzeyde belirlenmektedir. 3. yetenek düzeyinde en büyük alfa seviyesinde tahrifatlı bireyler mükemmel şekilde belirlenirken en düşük alfa seviyesinde bile istatistiksel gücün yüksek olduğu görülmektedir. 4. yetenek düzeyinde $\alpha = .05$ seviyesinde istatistiksel güç mükemmel iken alfa seviyesi düşüğe istatistiksel gücün azaldığı ancak $\alpha = .0005$ seviyesinde bile istatistiksel gücün yüksek olduğu görülmektedir. 5. yetenek düzeyinde $\alpha = .05$ seviyesinde mükemmel yakın düzeyde olmak üzere, tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edildiği ancak alfa seviyesi katılaştıkça istatistiksel gücün düştüğü ve $\alpha = .005$ seviyesinden itibaren gücün kabul edilebilir düzeyde olmadığı belirlenmiştir. Wollack ve diğerlerinin (2015) çalışmasında 10 silme yapıldığı durumda en büyük alfa seviyesinde 3. yetenek düzeyine kadar istatistiksel gücün mükemmel yakın olduğu, 3. yetenek düzeyinde çok yüksek olduğu ve 4. yetenek düzeyinde ise kabul edilebilir düzeyde olduğu görülmektedir. 1. yetenek düzeyinde neredeyse en düşük alfa seviyelerine kadar istatistiksel gücün oldukça yüksek olduğu; 2. yetenek düzeyinde istatistiksel gücün büyük alfa seviyelerinde kabul edilebilir olduğu görülmektedir. Araştırma bulguları, en üst yetenek düzeyinde büyük alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilmesi ve düşük yetenek düzeylerinin düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden (2015) farklılaşmaktadır. Sinharay ve Johnson'ın (2016) ise 10 silme yapıldığı durumda kabul edilebilir istatistiksel güç değerlerini $\alpha = .05$ düzeyinde 1. ve 2. yetenek düzeyinde; $\alpha = .01$ düzeyinde ise yalnızca 1. yetenek düzeyinde elde ettiği görülmektedir.

%30 oranında tahrifat (15 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün en büyük üç alfa seviyesinde de çok yüksek olduğu görülmektedir. İstatistiksel güç yetenek düzeylerine göre

incelendiğinde 1, 2 ve 3. yetenek düzeylerinde I. Tip hatanın iyi kontrol edildiği tüm alfa seviyelerinde istatistiksel gücün mükemmel olduğu görülmektedir. Özellikle 3. yetenek düzeyinde en düşük alfa seviyesinde bile tahrifatlı bireylerin tamamı tespit edilebilmektedir. 4. yetenek düzeyinde $\alpha = .05$ seviyesinde istatistiksel güç mükemmel iken alfa seviyesi katılaştıkça istatistiksel gücün bir miktar düştüğü ancak $\alpha = .0001$ seviyesinde bile istatistiksel gücün çok yüksek olduğu görülmektedir. 5. yetenek düzeyinde $\alpha = .05$ seviyesinde mükemmel düzeyde olmak üzere, tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edildiği ancak alfa seviyesi düştükçe istatistiksel gücün düştüğü, en son kabul edilebilir istatistiksel gücün $\alpha = .005$ seviyesinde elde edildiği görülmektedir. Wollack ve diğerlerinin (2015) çalışmasında 15 silme yapıldığı durumda en büyük alfa seviyesinde en düşük yetenek düzeylerinde istatistiksel gücün neredeyse mükemmel olduğu, 3. yetenek düzeyinde ise çok yüksek olduğu görülmektedir. 1. yetenek düzeyinde neredeyse en düşük alfa seviyesinde bile istatistiksel güç tatmin edicidir. 2. yetenek düzeyinde istatistiksel gücün $\alpha = .001$ seviyesinde dahi kabul edilebilir olduğu görülmektedir. Araştırma bulguları, en üst yetenek düzeyinde büyük alfa seviyelerinde yüksek istatistiksel güç elde edilmesi ve düşük yetenek düzeylerinin düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden (2015) farklılaşmaktadır.

Değişken sayıda silme yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün en büyük iki alfa seviyesinde tatmin edici, $\alpha = .005$ seviyesinde ise kabul edilebilir düzeyde olduğu görülmektedir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde 1. yetenek düzeyinde $\alpha = .05$ ve $\alpha = .01$ düzeylerinde güç mükemmeldir. 2. yetenek düzeyinde en büyük alfa seviyesinde istatistiksel gücün mükemmel, $\alpha = .01$ seviyesinde ise mükemmel yakın olduğu belirlenmiştir. 3. yetenek düzeyinde ise hiçbir alfa seviyesinde kabul edilebilir düzeyde istatistiksel güç elde edilememiştir. Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) çalışmalarında da görüldüğü gibi yüksek yetenek düzeylerinde tahrifata uğrayan test alan olmadığı için istatistiksel güç hesaplanmamıştır. Wollack ve diğerlerinin (2015) çalışmasında değişken sayıda silme yapıldığı durumda, en düşük alfa seviyesi dışında 1. yetenek düzeyinde tüm alfa seviyelerinde istatistiksel gücün kabul edilebilir olduğu ancak 2. ve 3. yetenek düzeylerinde kabul edilebilir istatistiksel gücün olmadığı görülmektedir. Araştırma bulguları, 2. yetenek düzeyinde büyük alfa seviyelerinde yüksek istatistiksel güç elde edilmesi ve 1. yetenek düzeyinde düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden (2015) farklılaşmaktadır. Sinharay ve Johnson (2016) tarafından ise değişken silme yapıldığı durumda kabul edilebilir istatistiksel gücün yalnızca $\alpha = .05$ düzeyinde 1. yetenek düzeyinde elde edildiği raporlaştırılmıştır. Genel olarak Sinharay ve Johnson'ın (2016)

istatistiksel güç değerlerinin oldukça düşük olduğu görülmekte; tahrifatlı birey yüzdesinin (%50), araştırmadaki tahrifatlı birey yüzdesinden (yaklaşık %16) oldukça yüksek olmasının istatistiksel gücü etkileyebileceği düşünülmektedir. Qin (2016) tarafından EDI_WTR kullanılarak yürütülen araştırmanın bazı koşullarında tahrifat yüzdesinin artmasının istatistiksel gücü bir miktar düşürdüğü görülmektedir.

Tüm tahrifat miktarı koşulları için yetenek düzeyi arttıkça kontrol edilen I. Tip hata oranı sayısının arttığı ancak istatistiksel gücün düştüğü görülmektedir. En düşük yetenek düzeyleri dışında diğer yetenek düzeyleri için tahrifat oranı arttıkça istatistiksel gücün arttığı görülmektedir. 1. ve 2. yetenek düzeyinde ise istatistiksel güç en düşük tahrifat oranı koşulunda bile mükemmel olduğu için bir artış gözlenmemiştir. Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) da belirttiği üzere tahrifat oranı arttıkça istatistiksel gücün arttığı ancak alfa seviyesi azaldıkça ve yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir. Wollack ve diğerleri (2015) yüksek yetenek düzeylerinde istatistiksel gücün düşük çıkmasını SBİ'nin hesaplanmasındaki doğru yanıt olasılık ortalamasına bağlı olarak açıklamaktadır. Buna göre, düşük yetenek düzeyindeki test alanlar için, silinmiş maddeler kolay olmadıkça, test alanın beklenen doğru yanıt olasılık ortalamasına sahip olarak maddeyi doğru yanıtlaması beklenmemektedir. Bu durum ilgili maddeyi doğru yanıtlamış test alanlar için SBİ'nin manidar çıkararak test alanı tahrifatlı olarak işaretlemesine neden olmaktadır. Yüksek yetenek düzeyindeki test alanlarda ise, silinmiş maddelerin doğru yanıtlanma olasılığının, maddeler çok zor olmadıkça, doğru yanıt olasılık ortalamasından yüksek olması ve dolayısıyla maddenin doğru yanıtlanması beklenmektedir. Bu durum, madde tahrifat sonucunda doğru yanıtlanmış olsa bile SBİ'nin manidar çıkmamasına ve test alanın tahrifatlı olarak işaretlenmemesine neden olmaktadır. Sonuç olarak tahrifatlı test alanların yüksek yetenek düzeylerinde SBİ ile belirlenmesi zorlaşmaktadır.

Süreklilik düzeltmesi yapılmadığı durumda I. Tip hatanın belirlenen alfa seviyesini aşmadığı koşullar için süreklilik düzeltmesi yapılmasının istatistiksel gücü ne kadar düşürdüğü incelenmiştir. Buna göre tahrifat oranı dikkate alınmadığında istatistiksel gücün, 1. yetenek düzeyinde değişmediği, 4. yetenek düzeyinde 0.13; 5. yetenek düzeyinde ise 0.02 ile 0.18 arasında azaldığı görülmektedir. Silme koşullarının tümünde 1. yetenek düzeyi için süreklilik düzeltmesi yapılmasının istatistiksel gücü değiştirmedeği görülmektedir. 4. yetenek düzeyi için süreklilik düzeltmesinin yapılması istatistiksel gücü 5 silme koşulunda 0.08; 10 silme koşulunda 0.19; 15 silme koşulunda 0.11 azaltmıştır. 5. yetenek düzeyi için ise süreklilik düzeltmesinin yapılması istatistiksel gücü 5 silme koşulunda 0.01 ile 0.21 arasında; 10 silme koşulunda 0.01 ile 0.20 arasında; 15 silme koşulunda 0.00 ile 0.23 arasında azaltmıştır.

Uygulamada tahrifata uğrayan bireylerin daha çok düşük yetenek düzeylerinde olması beklenmektedir (Wollack vd., 2015). SBİ ile tahrifatlı bireylerin belirlenmesinde süreklilik düzeltmesi uygulanmasının I. Tip hatanın daha iyi kontrol edilmesini sağlayarak elde edilen kabul edilebilir istatistiksel güç koşulunu artırdığı ve bu düzeltmenin düşük yetenek düzeylerinde istatistiksel gücü neredeyse hiç etkilemediği belirlenmiştir. Ayrıca tahrifat oranı arttıkça istatistiksel gücün değişim miktarının da azaldığı görülmektedir.

160 Maddelik Veri Seti

160 maddelik veri seti için hiçbir durumda I. Tip hata kontrol edilemediği için istatistiksel güç dikkate alınmamaktadır.

Tablo 13'te yer alan 50 maddelik veri seti için istatistiksel güç değerleri Bradley'in esnek referans aralığına göre incelendiğinde; çoğunlukla orta ve üzerindeki yetenek düzeylerinde düşük alfa seviyelerindeki nominal düzeye göre kabul edilebilir istatistiksel gücün değerlendirme dışında kaldığı; düşük yetenek düzeylerinde ise $\alpha = .005$ düzeyinde istatistiksel gücün değerlendirmeye alındığı ve bu düzeyde istatistiksel gücün mükemmel olduğu görülmektedir. Uygulamada tahrifata uğrayan bireylerin daha çok düşük yetenek düzeylerinde olması beklendiğinden düşük yetenek düzeylerinde daha düşük alfa seviyelerinde istatistiksel gücün değerlendirilmesine olanak sağladığı için; yüksek yetenek düzeylerindeki düşük alfa seviyelerindeki istatistiksel güç kaybının göz ardı edilebileceği düşünülmektedir. 160 maddelik veri setinin Bradley'in esnek referansına göre değerlendirilmesi sonucunda ise I. Tip hata oranının kontrol edilebildiği tek koşul olan en düşük yetenek düzeyinin en büyük alfa seviyesinde istatistiksel gücün mükemmel olduğu görülmektedir.

5000 Örneklem Büyüklüğü İçin SBİ'nin İstatistiksel Güç ve I. Tip Hata Bulguları

5000 örneklem büyüklüğü için I. Tip hata oranlarına ilişkin bulgular Tablo 14 ve Tablo 15'te, istatistiksel güce ilişkin bulgular ise Tablo 16 ve Tablo 17'de yer almaktadır.

Tablo 14

Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi Yapılmaksızın SBI'nin I. Tip Hataları (n= 5000)

Yetenek Düzeyi	50 Madde							Yetenek Düzeyi	160 Madde						
	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$		$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
1	<u>0.029</u>	0.023	0.0196	0.0152	0.01357	0.01087	0.008391	1	0.082	0.049	0.0432	0.0317	0.02774	0.02017	0.014435
2	0.079	0.052	0.0423	0.0274	0.02287	0.01691	0.012522	2	0.197	0.119	0.0897	0.0521	0.04113	0.02674	0.015000
3	0.105	0.044	0.0321	0.0197	0.01609	0.01091	0.007000	3	0.256	0.104	0.0694	0.0287	0.02313	0.01183	0.005565
4	<i>0.065</i>	0.017	0.0111	0.0051	0.00426	0.00149	0.001064	4	0.242	0.068	0.0389	0.0114	0.00804	0.00353	0.001532
5	<u>0.021</u>	<u>0.004</u>	<u>0.0018</u>	<u>0.0006</u>	<u>0.00043</u>	0.00021	0.000043	5	0.184	0.027	0.0131	0.0041	0.00306	0.00136	0.000383
Toplam	<i>0.060</i>	0.028	0.0213	0.0135	0.01136	0.00802	0.005759	Toplam	0.192	0.073	0.0507	0.0254	0.02049	0.01264	0.007328

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 14'e göre 50 maddelik veri setinde süreklilik düzeltilmesi yapılmaksızın I. Tip hata oranlarına ilişkin bulgular, Tablo 10'da yer alan 500 test alan için toplam test alanlar ile 2. ve 3. yetenek düzeyinde I. Tip hatanın hiçbir alfa seviyesinde kontrol edilememesi bulgularıyla benzerlik göstermektedir. Hata oranları yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden incelendiğinde, tüm alfa seviyelerinde nominal düzeyin aşıldığı ve hatanın iyi kontrol edilemediği görülmektedir. Bu durum tahrifatlı olmadığı halde tahrifatlı olarak belirlenen test alanlarının oranının alfa düzeyini aştığını göstermektedir. Araştırma bulguları bu yönüyle Qin'in (2016) bulgularına benzerlik göstermektedir. Wollack ve diğerleri (2015) ise yalnızca düşük alfa seviyelerinde hata oranlarının kontrol edilemediğini belirtmiştir.

I. Tip hata oranlarının yetenek düzeylerine göre değişimi incelendiğinde, 1. yetenek düzeyinin yalnızca en yüksek alfa seviyesinde hatanın iyi kontrol edildiği görülmektedir. 5. yetenek düzeyinde ise en düşük iki alfa seviyesi dışında I. Tip hata oranlarının nominal düzeyi aşmadığı görülmektedir. 2, 3 ve 4. yetenek düzeyinin tüm alfa seviyelerinde, I. Tip hata oranlarının kontrol edilemediği görülmektedir. Wollack ve diğerlerinde (2015) en düşük yetenek düzeyinde I. Tip hatanın yalnızca $\alpha=.05$ düzeyinde kontrol edilebildiği ancak diğer tüm yetenek düzeylerinde hata oranlarının nominal seviyeyi aşmadığı görülmektedir. Sinharay ve Johnson (2016) test alanların %50'sine tahrifat uygulayıp, I. Tip hata oranlarını iki farklı iyicil silme türüne göre ayrı ayrı hesaplamıştır. Dizi sonu silmeleri için, 1. yetenek düzeyinin en büyük alfa seviyesi dışında Tablo 14 bulgularına benzer şekilde orta ve düşük yetenek düzeylerinde hiçbir alfa seviyesinde hatanın kontrol edilemediğini belirtmiştir. Tesadüfi silmeler için ise Tablo 14 bulgularının aksine, 1. yetenek düzeyinin en büyük alfa seviyesi dışındaki tüm seviyelerde I. Tip hata oranlarının nominal seviyeyi aşmadığı görülmektedir. Yürütülen araştırmada tüm iyicil silme türlerini içerecek şekilde hesaplanan I. Tip hata oranlarına ilişkin bulguların, Sinharay ve Johnson'ın (2016) iyicil silme türlerine göre ayrı ayrı hesapladığı I. Tip hata oranları bulgularıyla doğrudan karşılaştırmanın yanıltıcı olabileceği düşünülmektedir. Ayrıca Sinharay ve Johnson'ın tahrifatlı birey yüzdesinin yürütülen araştırmanın tahrifatlı birey yüzdesinden (yaklaşık %8) oldukça yüksek olmasının I. Tip hata oranlarını etkileyebileceği düşünülmektedir. Genel olarak Tablo 14 incelendiğinde I. Tip hata oranlarının iyi kontrol edilemediği görülmektedir. Uygulamada test tahrifatının çoğunlukla düşük yetenekli bireylerde gerçekleşmesi beklendiğinden, 500 test alan için olduğu gibi 5000 test alan için de I. Tip hatanın düşük yetenek düzeylerinde iyi kontrol edilememesi, Wollack ve diğerlerinin (2015) de belirttiği gibi, bir problem oluşturmaktadır.

160 Maddelik Veri Seti

160 maddelik veri seti üzerinden hesaplanan I. Tip hata oranlarının hiçbir koşulda ve hiçbir alfa seviyesinde kontrol edilemediği görülmektedir. Dolayısıyla 50 maddelik veri setinde, 160 maddelik veri setine göre I. Tip hata oranları daha iyi kontrol edilmektedir. 160 maddelik veri seti için tahrifatlı olmayan test alanların, 50 maddelik veri setine göre daha fazla hata oranı ile tahrifatlı olarak işaretlendiği görülmektedir. Tablo 14’te yer alan 160 maddelik veri seti için I. Tip hata oranları Tablo 10 bulgularına göre bir miktar düşüş göstermekle birlikte, hatanın kontrol edilememesi açısından tamamen benzerlik göstermektedir.

50 maddelik veri setinde en yüksek yetenek düzeyi için, nominal seviyeye göre kabul edilebilir olan büyük alfa seviyelerindeki I. Tip hata oranlarının Bradley’in (1978) esnek referansı kullanıldığında referans aralığının altında kalarak kontrol edilemediği görülmektedir. Bununla birlikte 4. yetenek düzeyinde $\alpha = .05$ seviyesinde I. Tip hatanın kontrol edildiği görülmektedir. Qin (2016) yürütülen araştırmanın bulgularından farklı olarak, Bradley’in (1978) esnek referansı kullanıldığında süreklilik düzeltilmesi uygulanmayan SBI’nin hiçbir koşulda I. Tip hatayı kontrol etmediğini; EDI_WTR’nin ise $\alpha = .05$ düzeyinde hiçbir koşulda I. Tip hatayı kontrol edemezken $\alpha = .01$ düzeyinde tüm koşullarda I. Tip hatayı kontrol ettiğini belirtmiştir. Değerlendirmede esnek referans aralığının kullanılmasının 160 maddelik veri setinde I. Tip hata oranlarının kontrolünde bir değişikliği yol açmadığı görülmektedir.

Aynı veri setlerine süreklilik düzeltilmesi yapılarak hesaplanan I. Tip hata oranları Tablo 15’te yer almaktadır.

Tablo 15

Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi İle SBI'nin I. Tip Hataları (n= 5000)

Yetenek Düzeyi	50 Madde							Yetenek Düzeyi	160 Madde						
	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$		$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
1	<u>0.012</u>	<u>0.006</u>	<u>0.0051</u>	0.0025	0.00230	0.00191	0.001478	1	<u>0.028</u>	<u>0.013</u>	0.0092	0.0058	0.00509	0.00357	0.002261
2	<u>0.022</u>	<u>0.011</u>	<u>0.0077</u>	0.0049	0.00426	0.00317	0.001826	2	<u>0.058</u>	<u>0.023</u>	0.0167	0.0083	0.00617	0.00357	0.001522
3	<u>0.022</u>	<u>0.008</u>	<u>0.0059</u>	0.0027	0.00213	0.00143	0.000739	3	<u>0.063</u>	<u>0.017</u>	0.0114	0.0046	0.00339	0.00235	0.000826
4	<u>0.008</u>	<u>0.002</u>	<u>0.0011</u>	<u>0.0008</u>	<u>0.00051</u>	0.00021	0.000043	4	<u>0.054</u>	<u>0.012</u>	<u>0.0069</u>	0.0032	0.00264	0.00153	0.000766
5	<u>0.003</u>	<u>0.000</u>	<u>0.0002</u>	<u>0.0000</u>	<u>0.00000</u>	<u>0.00000</u>	<u>0.000000</u>	5	<u>0.036</u>	<u>0.006</u>	<u>0.0042</u>	0.0024	0.00166	0.00089	0.000085
Toplam	<u>0.013</u>	<u>0.005</u>	<u>0.0040</u>	0.0022	0.00183	0.00134	0.000810	Toplam	<u>0.048</u>	<u>0.014</u>	0.0096	0.0048	0.00378	0.00237	0.001086

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 15'e göre 50 maddelik veri setinde süreklilik düzeltilmesi yapılarak hesaplanan I. Tip hata oranlarına ilişkin bulgular 3. ve 4. yetenek düzeyi dışında genel olarak Tablo 11'de yer alan 500 test alan üzerinden elde edilen bulgulara benzemektedir. 3. ve 4. yetenek düzeyi için Tablo 11 bulgularının aksine düşük alfa seviyelerinde I. Tip hata kontrol edilememiştir.

Tablo 15'e göre süreklilik düzeltilmesinin, bazı koşullarda I. Tip hata oranının iyi kontrol edilmesini sağladığı görülmektedir. 50 maddelik veri setinde süreklilik düzeltilmesi yapılarak hesaplanan I. Tip hata oranları, yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden incelendiğinde daha büyük alfa seviyelerinde nominal düzeyi aşmadığı ve I. Tip hatanın iyi kontrol edildiği görülmektedir. Alfa seviyesi azaldıkça nominal düzeyi aşmakta ve I. Tip hata kontrol edilememektedir. Wollack ve diğerleri (2015) tüm test alanlar üzerinden süreklilik düzeltilmesi ile hesaplanan I. Tip hata oranlarının tüm alfa seviyelerinde kontrol edildiğini; Sinharay ve diğerleri (2017) de $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ seviyelerinde hata oranlarının kontrol edildiğini belirtmiştir. $\alpha = .001$ seviyesi dışında araştırma bulguları Sinharay ve diğerleriyle benzerlik göstermektedir. Qin (2016) süreklilik düzeltilmesi uygulanan EDI_WTR'nin hiçbir koşulda ve hiçbir alfa seviyesinde I. Tip hatayı kontrol edemediğini belirtmiştir ancak EDI_WTR'nin SBİ'nin özel bir kullanımı olduğu göz ardı edilmemelidir.

I. Tip hata oranlarının yetenek düzeylerine göre değişimi incelendiğinde ise $\alpha = .05$ için tüm yetenek düzeylerinde I. Tip hatanın iyi kontrol edildiği görülmektedir. Orta yetenek düzeyi ve altında, 1, 2 ve 3. yetenek düzeylerinde, düşük alfa seviyelerinde I. Tip hata oranlarının nominal düzeyi aştığı ve I. Tip hatanın iyi kontrol edilemediği, yalnızca en büyük alfa seviyelerinde hata oranlarının kontrol edildiği görülmektedir. 4. yetenek düzeyinde düşük alfa seviyeleri dışında; 5. yetenek düzeyinde ise tüm alfa seviyelerinde I. Tip hata oranlarının iyi kontrol edildiği görülmektedir. Süreklilik düzeltilmesi uygulamasına rağmen I. Tip hatanın düşük yetenek düzeylerinin düşük alfa seviyelerinde iyi kontrol edilememesi problemleri bir durum olmakla birlikte süreklilik düzeltilmesi ile iki alfa seviyesinde I. Tip hata kontrol edilebilir duruma gelmiştir. Wollack ve diğerlerinin (2015) I. Tip hata oranları yetenek düzeylerine göre incelendiğinde, en düşük yetenek düzeyinin en düşük alfa seviyesi dışında hata oranlarının iyi kontrol edildiği görülmektedir. Sinharay ve Johnson'ın (2016) dizi sonu silmeleri için, $\alpha = .05$, $\alpha = .01$, $\alpha = .001$ seviyelerinde tüm yetenek düzeylerinde, $\alpha = .0001$ seviyesinde ise orta yetenek düzeyi hariç tüm yetenek düzeylerinde I. Tip hatanın kontrol edildiği görülmektedir. Tesadüfi silmeler için ise tüm yetenek düzeylerinde ve tüm alfa seviyelerinde I. Tip hatanın kontrol edildiği görülmektedir. Sinharay ve Johnson (2016) tarafından süreklilik düzeltilmesi ile iyicil

silme türlerine göre ayrılarak hesaplanan I. Tip hata oranlarının yalnızca $\alpha = .0001$ seviyesinde farklılaştığı göz önünde bulundurulduğunda, araştırma bulgularının Sinharay ve Johnson'ın bulgularıyla karşılaştırılmasının uygun olduğu düşünülmektedir. Buna göre Tablo 15'te yer alan bulgular Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) bulgularıyla karşılaştırıldığında I. Tip hata oranlarının, yürütülen araştırma bulgularından daha iyi kontrol edildiği görülmektedir.

160 Maddelik Veri Seti

160 maddelik veri seti için süreklilik düzeltmesi ile hesaplanan I. Tip hata oranları, süreklilik düzeltmesi yapılmadan hesaplanan hata oranlarının yer aldığı Tablo 14 ile karşılaştırıldığında bazı koşullar için hata oranlarının kontrol edilebildiği görülmektedir. I. Tip hata oranları yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden incelendiğinde, yalnızca en büyük alfa seviyesinde hatanın kontrol edildiği görülmektedir.

I. Tip hata oranlarının yetenek düzeylerine göre değişimi incelendiğinde, 1. yetenek düzeyinin en büyük alfa seviyesinde hatanın iyi kontrol edildiği görülmektedir. 5. yetenek düzeyinde ise en büyük üç alfa seviyesinde I. Tip hata oranlarının nominal düzeyi aşmadığı; 2, 3 ve 4. yetenek düzeyinin alfa seviyelerinin tamamında ise I. Tip hata oranlarının nominal düzeye göre kontrol edilemediği görülmektedir. Sinharay ve diğerlerinin (2017) 100 madde üzerinden yürüttükleri çalışmada I. Tip hata oranlarının $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ düzeylerinde kontrol edildiği görülmektedir. Örneklem büyüklüğü arttıkça araştırma bulgularının, örneklem büyüklüğü ve tahrifat yapılan birey oranı açısından simülasyon deseni oldukça farklılık gösteren Sinharay ve diğerlerinin (2017) bulgularına kısmen de olsa benzerlik gösterdiği görülmektedir. 160 maddelik veri seti için Tablo 15 incelendiğinde süreklilik düzeltmesi uygulandığı halde I. Tip hata oranlarının alfa seviyelerinin çoğunu aştığı görülmektedir. Süreklilik düzeltmesi uygulanması koşulların çoğunda, tahrifatlı olmayan test alanların tahrifatlı olarak işaretlenme oranının alfa düzeyini aşmasını engelleyememektedir. Tablo 15 bulguları, 500 test alan için bulguların yer aldığı Tablo 11 ile karşılaştırıldığında genel olarak hata oranlarının düştüğü görülmekte ve az da olsa bazı koşullar için I. Tip hatanın kontrol edilmesi önemlidir. Bu durum 160 maddelik veri seti için örneklem büyüklüğünün I. Tip hata kontrolü üzerinde etkili olabileceğini düşündürmektedir. Ancak 160 maddelik veri seti için de I. Tip hatanın düşük yetenek düzeylerinde iyi kontrol edilememesi Wollack ve diğerlerinin (2015) de belirttiği gibi istenmeyen bir durumdur. Ayrıca süreklilik düzeltmesinin madde sayısının az olduğu durumlarda I. Tip hatanın şişme eğiliminde olmasından dolayı kullanımının

önerildiği göz önünde bulundurulduğunda, 160 madde sayısının oldukça yüksek olmasına rağmen I. Tip hatanın hala yeterince kontrol edilememesi dikkat çekicidir.

Süreklilik düzeltmesi yapılarak hesaplanan I. Tip hata oranlarında da 500 test alan için olduğu gibi 50 maddelik veri setinde, 160 maddelik veri setine göre hata oranları daha iyi kontrol edilmektedir. Süreklilik düzeltmesi, beklendiği gibi, tüm koşullarda I. Tip hata oranının düşmesine neden olmaktadır. Bu bakımdan bulgular 500 test alan bulguları ile benzerlik göstermektedir.

I. Tip hatanın değerlendirilmesinde Bradley'in (1978) esnek referansı kullanıldığında 50 maddelik veri seti için $\alpha = .05$ seviyesindeki tüm yetenek düzeylerinde, alfa seviyesi kriterine göre kabul edilebilir olan I. Tip hata oranlarının referans aralığının altında kalarak kontrol edilemediği görülmektedir. Qin (2016) tarafından Bradley'in (1978) esnek referansı kullanıldığında süreklilik düzeltmesi uygulanan EDI_WTR'nin yalnızca $\alpha = .05$ düzeyinde I. Tip hatayı kontrol ettiği belirtilmiştir. Araştırma bulguları Qin (2016) ile benzerlik göstermemekle birlikte Qin'in SBİ'nin özel hali olan EDI_WTR'yi kullandığının da göz önünde bulundurulması gerekmektedir. 5. yetenek düzeyinde nominal düzeye göre kabul edilebilir olan hiçbir alfa seviyesinde I. Tip hata oranları kontrol edilememektedir. 4. yetenek düzeyinde de düşük alfa seviyelerindeki hata oranlarının referans aralığının altında kaldığı ancak $\alpha = .0005$ düzeyinde I. Tip hata oranlarının kontrol edildiği görülmektedir. I. Tip hata oranlarının değerlendirilmesinde esnek referans aralığının kullanılması 500 test alanına ilişkin Tablo 11 bulgulardan farklı olarak, kontrol altına alınan hata oranı sayısını azaltacak yönde etki göstermektedir. 160 maddelik veri seti için esnek referans aralığı kullanıldığında ise genel olarak kontrol edilen hata oranlarının sayısı artmaktadır. Yetenek düzeyi ayrımı yapılmaksızın I. Tip hata oranları incelendiğinde $\alpha = .05$ ve $\alpha = .01$ düzeylerinde hata oranlarının iyi kontrol edildiği görülmektedir. Tüm yetenek düzeylerinde $\alpha = .05$ seviyesindeki I. Tip hata oranlarının kontrol edildiği; en düşük yetenek düzeyi için $\alpha = .01$ seviyesinde de hata oranlarının referans aralığı aştığı görülmektedir. 4. yetenek düzeyinde ise en büyük üç alfa seviyesinde de hata oranları iyi kontrol edilmektedir.

5000 test alan için SBİ'nin süreklilik düzeltmesi yapılmaksızın güç oranları Tablo 16'da yer almaktadır.

Tablo 16

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 5000)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
%10 (5 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%10	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	1.000	1.000	0.998	0.990	(16 Silme)	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	1.000	0.988	0.970	0.924	0.774	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>1.000</u>	0.956	0.900	0.672	0.554	0.352	0.168	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.928</u>	<u>0.460</u>	<u>0.308</u>	<u>0.082</u>	<u>0.052</u>	0.016	0.000	5	1.000	1.000	1.000	1.000	1.000	1.000	0.956	0.764
	Toplam	<u>0.986</u>	0.883	0.842	0.748	0.715	0.658	0.586		Toplam	1.000	1.000	1.000	1.000	1.000	0.991	0.953
%20 (10 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%20	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	(32 Silme)	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	1.000	1.000	1.000	1.000	0.996	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>1.000</u>	1.000	0.998	0.964	0.934	0.798	0.560	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.992</u>	<u>0.812</u>	<u>0.622</u>	<u>0.268</u>	<u>0.158</u>	0.046	0.004	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
	Toplam	<u>0.998</u>	0.962	0.924	0.846	0.818	0.769	0.712		Toplam	1.000	1.000	1.000	1.000	1.000	1.000	1.000
%30 (15 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%30	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	(48 Silme)	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>1.000</u>	1.000	1.000	0.998	0.990	0.896	0.710	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.992</u>	<u>0.856</u>	<u>0.734</u>	<u>0.380</u>	<u>0.258</u>	0.092	0.006	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Toplam	<u>0.998</u>	0.971	0.947	0.876	0.850	0.798	0.743		Toplam	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Değişken Silme	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	Değişken Silme	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	0.998	0.998	0.996	0.992	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	0.946	0.832	0.784	0.640	0.598	0.520	0.412	3	0.974	0.916	0.882	0.812	0.776	0.728	0.626	
	4	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	4	—	—	—	—	—	—	—	—
	5	—	—	—	—	—	—	—	5	—	—	—	—	—	—	—	—
	Toplam	<u>0.982</u>	0.944	0.928	0.879	0.865	0.839	0.801		Toplam	0.991	0.972	0.961	0.937	0.925	0.909	0.875

(Devam ediyor)

Tablo 16 (Devam)

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 5000)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
Toplam	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	Toplam	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	1.000	1.000	0.999	0.996		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	0.987	0.958	0.946	0.907	0.892	0.861	0.796		3	0.994	0.979	0.971	0.953	0.944	0.932	0.907
	4	1.000	0.985	0.966	0.878	0.826	0.682	0.480		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.971</u>	<u>0.709</u>	<u>0.555</u>	<u>0.243</u>	<u>0.156</u>	0.051	0.003		5	1.000	1.000	1.000	1.000	1.000	0.985	0.921
	Toplam	0.992	0.940	0.908	0.833	0.806	0.758	0.701		Toplam	0.999	0.995	0.993	0.990	0.988	0.982	0.966

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

İstatistiksel güç tahrifat miktarına ve yetenek düzeylerine göre ayrılmaksızın incelendiğinde hata oranlarının tüm alfa seviyelerinde nominal düzeyi aşmasından dolayı istatistiksel güç değerlendirilememektedir. İstatistiksel güç tahrifat oranlarına göre ayrılmaksızın toplam test alanlar üzerinden yetenek düzeylerine göre incelendiğinde, 1. yetenek düzeyinde tahrifatlı bireylerin $\alpha = .05$ seviyesinde mükemmel düzeyde, 5. yetenek düzeyinde ise çok yüksek düzeyde doğru belirlendiği görülmektedir. 5. yetenek düzeyi için $\alpha = .01$ düzeyinde istatistiksel gücün kabul edilebilir olduğu ancak alfa seviyesi düştükçe istatistiksel gücün kabul edilemez düzeyde düştüğü görülmektedir. İstatistiksel güç tahrifat oranlarına göre incelendiğinde ise tahrifat oranı arttıkça yetenek düzeylerinde istatistiksel gücün arttığı 5. yetenek düzeyi için gözlenmektedir. 1. yetenek düzeyinde en düşük tahrifat oranında dahi tahrifatlı bireyler mükemmel şekilde belirlendiği için tahrifat oranı arttıkça gücün değişimi gözlenmemektedir.

%10 oranında tahrifat (5 silme) yapıldığı durumda $\alpha = .05$ düzeyinde 1. yetenek düzeyinde tahrifatlı bireyler mükemmel düzeyde, 5. yetenek düzeyinde ise yüksek düzeyde belirlenmektedir. 5. yetenek düzeyi için diğer alfa seviyelerinde istatistiksel gücün kabul edilebilir düzeyde olmadığı görülmektedir. %20 oranında tahrifat (10 silme) yapıldığı durumda 1. yetenek düzeyinde $\alpha = .05$ düzeyinde istatistiksel gücün mükemmel olduğu; 5. yetenek düzeyinde ise $\alpha = .05$ seviyesinde mükemmel yakınken, $\alpha = .01$ seviyesinde tatmin edici düzeyde olduğu ancak daha düşük alfa seviyelerinde istatistiksel gücün kabul edilebilir düzeyde olmadığı görülmektedir. %30 oranında tahrifat (15 silme) yapıldığı durumda %20 oranında tahrifat (10 silme) yapıldığı duruma benzer olarak, 1. yetenek düzeyinde $\alpha = .05$ seviyesinde istatistiksel gücün mükemmel olduğu; 5. yetenek düzeyinde ise $\alpha = .05$ seviyesinde mükemmel yakınken, $\alpha = .01$ seviyesinde tatmin edici düzeyde, $\alpha = .005$ seviyesinde ise kabul edilebilir düzeyde olduğu ancak daha düşük alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilemediği görülmektedir. Değişken sayıda silme yapıldığı durumda da 1. yetenek düzeyinde $\alpha = .05$ seviyesinde istatistiksel güç mükemmeldir. En yüksek yetenek düzeyinde ise tahrifata uğrayan test alan olmadığı için istatistiksel güç hesaplanmamıştır.

160 Maddelik Veri Seti

160 maddelik veri seti için hiçbir yetenek düzeyinin hiçbir alfa seviyesinde I. Tip hata kontrol edilemediği için istatistiksel güç dikkate alınmamaktadır.

Tablo 16 bulguları, 500 test alan için bulguların yer aldığı Tablo 12 ile karşılaştırıldığında kabul edilebilir istatistiksel güç değerlerinin sayısının biraz daha fazla olması önemlidir. Tablo 16 genel olarak incelendiğinde benzer araştırmalarda da olduğu gibi, tahrifat oranı arttıkça istatistiksel gücün arttığı ancak alfa seviyesi azaldıkça istatistiksel gücün düştüğü görülmektedir (Qin, 2016; Sinharay vd., 2017; Wollack vd., 2015). Wollack ve diğerlerinin (2015) de belirttiği üzere yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir.

Tablo 16'da yer alan istatistiksel güç değerleri Bradley'in esnek referans aralığına göre incelendiğinde ise 5. yetenek düzeyinde en büyük alfa seviyelerindeki istatistiksel güç değerlerinin I. Tip hata oranlarının kontrol edilememesi nedeniyle değerlendirme dışında kaldığı görülmektedir. Ancak 4. yetenek düzeyi için $\alpha = .05$ seviyesinde hata oranlarının referans aralığında kalması nedeniyle değerlendirilebilir olduğu ve bu durumda istatistiksel gücün mükemmel düzeyde olduğu görülmektedir.

5000 test alan için süreklilik düzeltmesi yapılarak uygulanan SBI'nin istatistiksel gücü Tablo 17'de yer almaktadır.

Tablo 17

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 5000)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
%10 (5 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%10	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	2	<u>1.000</u>	<u>1.000</u>	1.000	1.000	0.996	0.990	0.948	(16 Silme)	2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>1.000</u>	<u>1.000</u>	<u>0.986</u>	<u>0.942</u>	<u>0.896</u>	<u>0.780</u>	<u>0.548</u>	3	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>0.980</u>	<u>0.820</u>	<u>0.694</u>	<u>0.396</u>	<u>0.324</u>	<u>0.176</u>	<u>0.062</u>	4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.996
	5	<u>0.610</u>	<u>0.182</u>	<u>0.090</u>	<u>0.024</u>	<u>0.012</u>	<u>0.000</u>	<u>0.000</u>	5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	0.980	0.888	0.606	
	Toplam		<u>0.918</u>	<u>0.800</u>	<u>0.754</u>	0.672	0.646	0.589	0.512	Toplam		<u>1.000</u>	<u>1.000</u>	1.000	1.000	0.996	0.978
%20 (10 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%20	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	2	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	(32 Silme)	2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	0.986	3	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>1.000</u>	<u>0.996</u>	<u>0.982</u>	<u>0.898</u>	<u>0.826</u>	<u>0.668</u>	<u>0.400</u>	4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	5	<u>0.924</u>	<u>0.542</u>	<u>0.364</u>	<u>0.112</u>	<u>0.064</u>	<u>0.008</u>	<u>0.000</u>	5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.994
	Toplam		<u>0.985</u>	<u>0.908</u>	<u>0.869</u>	0.802	0.778	0.735	0.677	Toplam		<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000
%30 (15 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%30	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	2	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	(48 Silme)	2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	3	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>0.980</u>	<u>0.948</u>	<u>0.816</u>	<u>0.604</u>	4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	5	<u>0.960</u>	<u>0.678</u>	<u>0.506</u>	<u>0.218</u>	<u>0.126</u>	<u>0.026</u>	<u>0.000</u>	5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.996
	Toplam		<u>0.992</u>	<u>0.936</u>	<u>0.901</u>	0.840	0.815	0.768	0.721	Toplam		<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000
Değişken Silme	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	Değişken Silme	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	2	<u>1.000</u>	<u>0.998</u>	<u>0.998</u>	<u>0.996</u>	<u>0.996</u>	<u>0.992</u>	<u>0.972</u>	2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>0.744</u>	<u>0.576</u>	<u>0.538</u>	<u>0.448</u>	<u>0.406</u>	<u>0.328</u>	<u>0.252</u>	3	<u>0.896</u>	<u>0.796</u>	<u>0.772</u>	<u>0.704</u>	<u>0.668</u>	<u>0.594</u>	<u>0.466</u>	
	4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>0.000</u>	4	—	—	—	—	—	—	—	—
	5	—	—	—	—	—	—	—	5	—	—	—	—	—	—	—	—
	Toplam		<u>0.915</u>	<u>0.858</u>	<u>0.845</u>	0.815	0.801	0.773	0.741	Toplam		<u>0.965</u>	<u>0.932</u>	<u>0.924</u>	<u>0.901</u>	<u>0.889</u>	<u>0.865</u>

(Devam ediyor)

Tablo 17 (Devam)

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 5000)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
Toplam	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	Toplam	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	2	<u>1.000</u>	<u>1.000</u>	1.000	0.999	0.998	0.996	0.980		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>0.936</u>	<u>0.894</u>	<u>0.881</u>	0.848	0.826	0.777	0.697		3	<u>0.974</u>	0.949	0.943	0.926	0.917	0.899	0.867
	4	<u>0.993</u>	<u>0.939</u>	<u>0.892</u>	<u>0.758</u>	<u>0.699</u>	0.554	0.355		4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	0.999
	5	<u>0.831</u>	<u>0.467</u>	<u>0.320</u>	<u>0.118</u>	<u>0.067</u>	<u>0.011</u>	<u>0.000</u>		5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	0.993	0.963	0.865
	Toplam	<u>0.957</u>	<u>0.877</u>	<u>0.842</u>	0.779	0.755	0.710	0.654		Toplam	<u>0.994</u>	<u>0.989</u>	<u>0.987</u>	<u>0.984</u>	0.980	0.971	0.948

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

İstatistiksel güç tahrifat miktarına ve yetenek düzeylerine göre ayrılmaksızın toplam test alanlar üzerinden incelendiğinde, alfa seviyesi düştükçe istatistiksel güç azalmakla birlikte, en büyük üç alfa seviyesinde istatistiksel gücün oldukça yüksek olduğu görülmektedir. İstatistiksel güç tahrifat miktarına göre ayrılmaksızın toplam test alanlar üzerinden yetenek düzeylerine göre incelendiğinde, genel olarak yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir. 1. yetenek düzeyinde $\alpha = .05$ ve $\alpha = .01$ düzeylerinde tahrifatlı bireyler mükemmel düzeyde belirlenmektedir ancak daha düşük alfa seviyelerinde hatalar iyi kontrol edilemediği için istatistiksel güç değerlendirilememiştir. 2. yetenek düzeyinde yalnızca $\alpha = .05$ seviyesinde istatistiksel gücün değerlendirilebildiği ve bu seviyede tahrifatlı bireylerin mükemmel düzeyde belirlendiği görülmektedir. 3. yetenek düzeyinde değerlendirilebilen istatistiksel gücün yalnızca $\alpha = .05$ ve $\alpha = .01$ seviyelerinde olduğu ve bu seviyelerde istatistiksel gücün yüksek olduğu görülmektedir. 4. yetenek düzeyinde $\alpha = .05$ seviyesinde güç neredeyse mükemmel olmakla birlikte, alfa seviyesi küçüldükçe istatistiksel güç düşmektedir ancak $\alpha = .001$ seviyesinde bile güç kabul edilebilir düzeydedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hata oranlarının nominal düzeyi aşmadığı ancak yalnızca $\alpha = .05$ seviyesinde kabul edilebilir gücün elde edildiği görülmektedir.

İstatistiksel güç tahrifat miktarına göre incelendiğinde ise Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) da belirttiği ve 500 test alan üzerinden yapılan analizlerde de görüldüğü üzere tahrifat oranı arttıkça her yetenek düzeyinde istatistiksel gücün arttığı görülmektedir. %10 oranında tahrifat (5 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün $\alpha = .05$ seviyesinde yüksek olduğu, alfa seviyesi azaldıkça gücün düştüğü ancak $\alpha = .005$ seviyesinde hala kabul edilebilir düzeyde olduğu belirlenmiştir. Araştırma bulgularından farklı olarak Sinharay ve diğerlerinin (2017) 4 silme yaparak hesapladıkları istatistiksel gücün $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ düzeylerinin hiçbirinde kabul edilebilir olmadığı görülmektedir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde düşük yetenek düzeyleri ve orta yetenek düzeyinde büyük alfa seviyelerinde tahrifatlı bireylerin mükemmel düzeyde belirlendiği görülmektedir. 4. yetenek düzeyinde $\alpha = .05$ seviyesinde istatistiksel gücün çok yüksek, $\alpha = .01$ seviyesinde ise tatmin edici olduğu görülürken alfa seviyesi azaldıkça istatistiksel gücün kabul edilemez düzeyde düştüğü görülmektedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hata iyi kontrol edilmesine rağmen istatistiksel güç hiçbir alfa seviyesinde kabul edilebilir düzeyde değildir. Wollack ve diğerlerinin (2015) araştırmasında 5 silme yapıldığı durumda yalnızca 1.

yetenek düzeyinde ve $\alpha = .05$ düzeyinde çok yüksek istatistiksel güç; 1. yetenek düzeyi için $\alpha = .01$ ve 2. yetenek düzeyi için $\alpha = .05$ düzeylerinde kabul edilebilir istatistiksel güç elde edilmiştir. Araştırma bulguları, $\alpha = .05$ ve $\alpha = .01$ seviyelerinde en üst yetenek düzeyine kadar yüksek istatistiksel güç elde edilmesi ile Wollack ve diğerlerinden farklılaşmaktadır. Ancak en üst yetenek düzeyinde Wollack ve diğerlerinden daha yüksek istatistiksel güç elde edilmesine rağmen güç oranları kabul edilebilir değildir. Sinharay ve Johnson (2016) ise 5 silme yapıldığı durumda hiçbir yetenek düzeyinde ve alfa seviyesinde kabul edilebilir istatistiksel güç elde edilemediğini belirtmiştir.

%20 oranında tahrifat (10 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde en büyük üç alfa seviyesinde tahrifatlı bireylerin yüksek oranda doğru belirlendiği görülmektedir. Sinharay ve diğerlerinin (2017) 8 silme için hesapladıkları istatistiksel güç ise yalnızca $\alpha = .05$ seviyesinde kabul edilebilir düzeydedir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde en yüksek yetenek düzeyine kadar büyük iki alfa seviyesinin neredeyse tamamında tahrifatlı bireylerin mükemmel düzeyde belirlendiği görülmektedir. 4. yetenek düzeyinde $\alpha = .0005$ seviyesine kadar istatistiksel gücün yüksek olduğu görülmektedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edilmesine rağmen yalnızca en büyük alfa seviyesinde istatistiksel gücün yüksek olduğu ve diğer alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilemediği görülmektedir. Sinharay ve Johnson'ın (2016) 10 silme yapıldığı durumda kabul edilebilir istatistiksel güç değerlerini $\alpha = .05$ düzeyinde 1. ve 2. yetenek düzeyinde; $\alpha = .01$ düzeyinde ise yalnızca 1. yetenek düzeyinde elde ettiği görülmektedir. Araştırma bulguları, Sinharay ve Johnson'ın düşük yetenek düzeylerinde daha düşük alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilmemesi bulgularıyla benzerlik göstermektedir. Wollack ve diğerlerinin (2015) çalışmasında ise 10 silme yapıldığı durumda en büyük alfa seviyesinde 3. yetenek düzeyine kadar istatistiksel güç mükemmel yakinken, 3. yetenek düzeyinde güç çok yüksek ve 4. yetenek düzeyinde ise kabul edilebilir düzeydedir. 1. yetenek düzeyinde neredeyse en düşük alfa seviyelerine kadar istatistiksel gücün oldukça yüksek olduğu; 2. yetenek düzeyinde istatistiksel gücün büyük alfa seviyelerinde kabul edilebilir olduğu görülmektedir. Araştırma bulguları, en üst yetenek düzeyinin en büyük alfa seviyesinde yüksek istatistiksel güç elde edilmesi ve düşük yetenek düzeylerinin düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden farklılaşmaktadır.

%30 oranında tahrifat (15 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün en büyük üç alfa seviyesinde de oldukça yüksek olduğu görülmektedir. İstatistiksel güç yetenek düzeylerine göre

incelendiğinde 1, 2 ve 3. yetenek düzeylerinde I. Tip hatanın iyi kontrol edildiği en büyük alfa seviyelerinde istatistiksel gücün mükemmel olduğu görülmektedir. 4. yetenek düzeyinde en büyük alfa seviyelerinde istatistiksel güç mükemmel iken $\alpha = .001$ seviyesinde bile istatistiksel gücün çok yüksek olduğu görülmektedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edilmesine rağmen, 10 silme yapıldığı durumdakine benzer olarak, yalnızca en büyük alfa seviyesinde tahrifatlı bireylerin çok yüksek oranda doğru belirlendiği ve diğer alfa seviyelerinde gücün kabul edilebilir düzeyde olmadığı görülmektedir. Wollack ve diğerlerinin (2015) çalışmasında 15 silme yapıldığı durumda en büyük alfa seviyesinde en düşük yetenek düzeylerinde istatistiksel gücün neredeyse mükemmel olduğu, 3. yetenek düzeyinde ise gücün çok yüksek olduğu görülmektedir. 1. yetenek düzeyinde neredeyse en düşük alfa seviyesinde bile istatistiksel güç tatmin edicidir. 2. yetenek düzeyinde istatistiksel gücün $\alpha = .001$ seviyesinde dahi kabul edilebilir olduğu görülmektedir. Araştırma bulguları 10 silme yapıldığı durumdakine benzer olarak, en üst yetenek düzeyinin en büyük alfa seviyesinde yüksek istatistiksel güç elde edilmesi ve düşük yetenek düzeylerinin düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden (2015) farklılaşmaktadır.

Değişken sayıda silme yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden istatistiksel güç en büyük üç alfa seviyesinde de tatmin edici düzeydedir. İstatistiksel güç 1. ve 2. yetenek düzeylerinde büyük alfa seviyelerinde mükemmeldir. 3. yetenek düzeyinde yalnızca $\alpha = .05$ seviyesinde kabul edilebilir istatistiksel güç elde edilmiştir. 4. yetenek düzeyinde ise istatistiksel gücün tüm alfa seviyelerinde mükemmel olduğu görülmektedir. 4. yetenek düzeyinde tahrifatlı test alanların mükemmel düzeyde belirlenmesinin çok az sayıda test alanın tahrifatlı olmasından kaynaklanabileceği düşünülmektedir. Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) çalışmalarında da görüldüğü gibi en yüksek yetenek düzeyinde tahrifata uğrayan test alan olmadığı için istatistiksel güç hesaplanmamıştır. Wollack ve diğerlerinin (2015) çalışmasında değişken sayıda silme yapıldığı durumda, en düşük alfa seviyesi dışında 1. yetenek düzeyinde tüm alfa seviyelerinde istatistiksel gücün kabul edilebilir olduğu ancak 2. ve 3. yetenek düzeylerinde kabul edilebilir istatistiksel gücün olmadığı görülmektedir. Araştırma bulguları, diğer yetenek düzeylerinde de yüksek istatistiksel güç elde edilmesi ve 1. yetenek düzeyinde düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden (2015) farklılaşmaktadır. Sinharay ve Johnson (2016) tarafından ise değişken silme yapıldığı durumda kabul edilebilir istatistiksel gücün yalnızca $\alpha = .05$ düzeyinde 1. yetenek düzeyinde elde edildiği raporlaştırılmıştır. Qin (2016) EDI_WTR'ye ilişkin bulgularının bazı koşullarında istatistiksel gücün tahrifat yüzdesinin artması ile bir miktar düştüğü görülmektedir. Genel olarak Sinharay

ve Johnson'ın (2016) tüm koşullar için istatistiksel güç değerlerinin oldukça düşük olduğu göz önünde bulundurulduğunda, Sinharay ve Johnson'ın simülasyon desenindeki tahrifat yüzdesi gibi farklılıkların istatistiksel gücü etkileyebileceği düşünülmektedir.

Tüm tahrifat miktarı koşulları için yetenek düzeyi arttıkça nominal düzeyi aşmayan I. Tip hata oranı sayısının arttığı ancak istatistiksel gücün düştüğü görülmektedir. Düşük yetenek düzeylerinde en düşük tahrifat oranı koşulunda bile güç mükemmel olduğu için tahrifat oranına bağlı olarak bir değişim gözlenmemiştir. Yüksek yetenek düzeylerinde ise tahrifat oranı arttıkça istatistiksel gücün arttığı görülmektedir. Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) bulgularına benzer olarak tahrifat oranı arttıkça istatistiksel gücün arttığı ancak alfa seviyesi azaldıkça ve yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir. Wollack ve diğerlerine (2015) göre tahrifatlı bir madde için, düşük yetenek düzeyindeki test alanın doğru yanıt olasılık ortalamasına sahip olarak maddeyi doğru yanıtlama olasılığı düşük olduğundan kolaylıkla SBİ ile tahrifatlı olarak belirlenmektedir. Yüksek yetenek düzeylerindeki test alanların ise doğru yanıt olasılık ortalaması yüksek olduğundan maddeyi doğru yanıtlaması beklenmekte ve SBİ tarafından tahrifatlı olarak işaretlenmemektedir. Sonuç olarak tahrifatlı test alanların yüksek yetenek düzeylerinde SBİ ile belirlenmesi zorlaşmaktadır.

Süreklilik düzeltmesi yapılması, benzer araştırmalarda olduğu gibi, I. Tip hata oranının alfa seviyesini aşmadığı koşul sayısını artırmakta ancak istatistiksel gücün düşmesine neden olmaktadır (Sinharay ve Johnson, 2016; Wollack ve diğerleri, 2015). Süreklilik düzeltmesi yapılmadığı durumda I. Tip hatanın nominal düzeyi aşmadığı koşullar için süreklilik düzeltmesi yapılmasının istatistiksel gücü ne kadar düşürdüğü incelenmiştir. Buna göre tahrifat oranı dikkate alınmadığında istatistiksel gücün, 1. yetenek düzeyinde değişmediği, 5. yetenek düzeyinde ise 0.09 ile 0.24 arasında azaldığı görülmektedir. 500 test alan bulgularına benzer şekilde silme koşullarının tümünde 1. yetenek düzeyi için süreklilik düzeltmesi yapılmasının istatistiksel gücü değiştirmediği görülmektedir. 5. yetenek düzeyi için ise süreklilik düzeltmesinin yapılması istatistiksel gücü %10 oranında tahrifat (5 silme) koşulunda 0.04 ile 0.32 arasında; %20 oranında tahrifat (10 silme) koşulunda 0.07 ile 0.27 arasında; %30 oranında tahrifat (15 silme) koşulunda 0.03 ile 0.09 arasında azaltmıştır. Uygulamada çoğunlukla düşük yetenek düzeyindeki bireylerin tahrifata uğraması beklenmektedir (Wollack vd., 2015). SBİ ile tahrifatlı bireylerin belirlenmesinde süreklilik düzeltmesi uygulanmasının nominal düzeyi aşmayan I. Tip hata oranı sayısının artmasını sağlayarak kabul edilebilir istatistiksel güç koşulunu artırdığı ve bu durumun düşük yetenek düzeylerinde istatistiksel gücü neredeyse hiç etkilemediği görülmüştür. Ayrıca tahrifat oranı arttıkça istatistiksel gücün değişim miktarının da azaldığı görülmektedir.

Tablo 17 bulguları 500 test alana ilişkin bulguların yer aldığı Tablo 13 ile karşılaştırıldığında istatistiksel gücün genel olarak tüm silme koşulları için, düşük yetenek düzeylerinde önemli bir değişiklik göstermediği, üst yetenek düzeylerinde ise bir miktar düştüğü belirlenmiştir.

160 Maddelik Veri Seti

Tablo 17’de 160 maddelik veri seti için istatistiksel güç tahrifat miktarına ve yetenek düzeylerine göre ayrılmaksızın toplam test alanlar üzerinden incelendiğinde, I. Tip hatanın alfa seviyesini aşmadığı tek durum olan en büyük alfa seviyesinde istatistiksel gücün mükemmel yakın olduğu görülmektedir. İstatistiksel güç tahrifat oranlarına göre ayrılmaksızın toplam test alanlar üzerinden incelendiğinde, 1. yetenek düzeyinde $\alpha = .05$ düzeyinde, 5. yetenek düzeyinde ise en büyük üç alfa seviyesinde tahrifatlı bireylerin mükemmel şekilde belirlendiği görülmektedir.

160 maddelik veri seti için istatistiksel güç tahrifat oranlarına göre incelendiğinde ise tüm sabit silme koşullarında I. Tip hatanın kontrol edildiği durumlarda güç mükemmeldir. Değişken silme koşulunda ise yetenek düzeyi ayrımı yapılmaksızın tahrifatlı bireylerin çok yüksek oranda belirlendiği görülmektedir. En düşük yetenek düzeyinde ise I. Tip hatanın kontrol edilebildiği tek durum olan $\alpha = .05$ düzeyinde istatistiksel güç mükemmeldir. 50 maddelik veri seti ile karşılaştırıldığında daha az sayıda istatistiksel güç yorumlanabilmektedir. Buna karşılık, tüm test alanlar üzerinden ve yetenek düzeyleri için aynı tahrifat miktarı oranlarında karşılaştırma yapıldığında 5. yetenek düzeyinde istatistiksel gücün 160 maddelik veri seti için her durumda daha yüksek olduğu görülmektedir. 500 test alan üzerinden istatistiksel güce ait bulguların yer aldığı Tablo 13 ile kıyaslandığında ise az sayıda da olsa kabul edilebilir istatistiksel gücün elde edilmiş olması önemlidir.

Tablo 17’de yer alan 50 maddelik veri seti için istatistiksel güç değerleri Bradley’in (1978) esnek referans aralığına göre değerlendirildiğinde; $\alpha = .05$ seviyesinde tüm yetenek düzeylerinde istatistiksel güç değerleri I. Tip hata oranlarının referans aralığın altında kalması nedeniyle değerlendirme dışında kalmıştır. Esnek referans aralığının kullanılması 1. ve 3. yetenek düzeyinde $\alpha = .005$, 2. yetenek düzeyinde ise $\alpha = .01$ seviyelerinde istatistiksel gücün değerlendirilebilir hale gelmesini sağlamıştır. Belirtilen durumlar için istatistiksel güç düşük yetenek düzeylerinde mükemmel, orta yetenek düzeyinde ise çok yüksektir. 4. yetenek düzeyinde de düşük alfa seviyelerinde istatistiksel gücün hata oranlarının referans aralığın

altında kalması nedeniyle değerlendirilemediği ancak $\alpha = .0005$ düzeyinde istatistiksel gücün değerlendirilebilir hale geldiği görülmektedir. Toplam test alanlar üzerinden yapılan analizler ve %10 oranında tahrifat (5 silme) koşulu dışında 4. yetenek düzeyi için belirtilen alfa seviyesinde istatistiksel gücün yüksek olduğu görülmektedir. 5. yetenek düzeyinde tüm alfa seviyelerindeki istatistiksel güç değerlerinin I. Tip hata oranlarının kontrol edilememesi nedeniyle değerlendirme dışında kaldığı görülmektedir. Esnek referans aralığının kullanılması üst yetenek düzeylerinde istatistiksel gücün değerlendirilememesi ile sonuçlanmaktadır ancak tahrifatın çoğunlukla düşük yetenek düzeylerinde yapılmasının beklenmesi nedeniyle bu durumun uygulamada problem teşkil etmeyeceği düşünülmektedir. Ayrıca esnek referans aralığı düşük yetenek düzeylerinde büyük alfa seviyelerinde istatistiksel gücün değerlendirilememesine neden olmaktadır ancak daha düşük alfa seviyelerinde istatistiksel gücün değerlendirilmesini sağlaması ve bu düzeylerde istatistiksel gücün yüksek olması önemlidir.

160 maddelik veri seti için esnek referans aralığı kullanıldığında değerlendirilebilen istatistiksel güç sayısının arttığı görülmektedir. Yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden $\alpha = .05$ ve $\alpha = .01$ düzeylerinde tahrifatlı bireylerin mükemmel yakın düzeyde belirlendiği görülmektedir. Tüm silme koşullarında $\alpha = .05$ düzeyinde tüm yetenek düzeylerinde istatistiksel gücün değerlendirilebildiği ve neredeyse tüm koşullarda gücün mükemmel olduğu görülmektedir. Ayrıca 1. yetenek düzeyinde $\alpha = .01$ yetenek düzeyinde gücün değerlendirilebilir hale gelmesi ve bu düzeyde gücün mükemmel olması önemlidir. Üst yetenek düzeylerinde de daha düşük alfa seviyelerinde istatistiksel güç değerlendirilebilir hale gelmiş ve belirtilen düzeylerde tahrifatlı bireyler mükemmel düzeyde belirlenmiştir.

50000 Örneklem Büyüklüğü İçin SBİ'nin İstatistiksel Güç ve I. Tip Hata Bulguları

50000 örneklem büyüklüğü için I. Tip hata oranlarına ilişkin bulgular Tablo 18 ve Tablo 19'da, istatistiksel güce ilişkin bulgular ise Tablo 20 ve Tablo 21'de yer almaktadır.

Tablo 18

Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi Yapılmaksızın SBI'nin I. Tip Hataları (n= 50000)

50 Madde								160 Madde							
Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
1	<u>0.027</u>	0.021	0.0184	0.0139	0.01251	0.01003	0.008070	1	0.088	0.050	0.0444	0.0322	0.02790	0.02033	0.013748
2	0.077	0.050	0.0415	0.0265	0.02274	0.01617	0.010891	2	0.189	0.115	0.0882	0.0496	0.04056	0.02655	0.015409
3	0.099	0.040	0.0289	0.0167	0.01363	0.00851	0.005043	3	0.260	0.104	0.0708	0.0314	0.02292	0.01285	0.005748
4	<u>0.055</u>	0.014	0.0088	0.0036	0.00273	0.00139	0.000889	4	0.232	0.060	0.0332	0.0102	0.00645	0.00324	0.001111
5	<u>0.013</u>	<u>0.002</u>	<u>0.0012</u>	<u>0.0004</u>	<u>0.00029</u>	0.00019	0.000111	5	0.168	0.021	0.0096	0.0036	0.00259	0.00122	0.000277
Toplam	0.054	0.025	0.0196	0.0121	0.01030	0.00720	0.004962	Toplam	0.188	0.070	0.0490	0.0252	0.01995	0.01275	0.007202

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 18'e göre 50 maddelik veri setinde süreklilik düzeltilmesi yapılmaksızın I. Tip hata oranlarına ilişkin bulgular Tablo 14'te yer alan 5000 test alan üzerinden elde edilen bulgulara tamamen benzemektedir. Yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden hata oranları incelendiğinde, hiçbir alfa seviyesinde I. Tip hata oranlarının kontrol edilemediği, tahrifatlı olmayan test alanların tahrifatlı olarak belirlenme oranının alfa düzeyini aştığı görülmektedir. Qin (2016) de, $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ seviyelerinde I. Tip hata oranlarının kontrol edilemediğini belirtmiştir. Wollack ve diğerleri (2015) ise düşük alfa seviyeleri dışında hata oranlarının iyi kontrol edildiğini ifade etmiştir.

Yetenek düzeylerine göre I. Tip hata oranlarının değişimi incelendiğinde, 1. yetenek düzeyinin yalnızca en yüksek alfa seviyesinde nominal seviyenin aşılmadığı görülmektedir. 5. yetenek düzeyinde ise en düşük alfa seviyeleri olan $\alpha = .0001$ ve $\alpha = .00001$ düzeyleri dışındaki alfa seviyelerinde I. Tip hata oranlarının iyi kontrol edildiği görülmektedir. 2, 3 ve 4. yetenek düzeyinin alfa seviyelerinin tamamında, I. Tip hata oranlarının kabul edilemez olduğu görülmektedir. Wollack ve diğerlerinde (2015) de en düşük yetenek düzeyinde I. Tip hatanın yalnızca $\alpha = .05$ düzeyinde kontrol edilebildiği ancak Tablo 18'in aksine diğer tüm yetenek düzeylerinde I. Tip hata oranlarının nominal düzeyin altında kaldığı görülmektedir. Test alanların %50'sine tahrifat yapan Sinharay ve Johnson (2016), I. Tip hata oranlarını dizi sonu ve tesadüfi silmelere göre ayrı ayrı hesaplamıştır. Dizi sonu silmeleri için Tablo 18 bulgularına çoğunlukla benzer şekilde orta ve düşük yetenek düzeylerinin tüm alfa düzeylerinde hata oranlarının nominal düzeyi aştığı görülmektedir. Tesadüfi silmeler için ise Tablo 18 bulgularının aksine, 1. yetenek düzeyinin en büyük alfa seviyesi dışındaki tüm seviyelerde I. Tip hata oranlarının iyi kontrol edildiği görülmektedir. Sinharay ve Johnson (2016) tarafından iyicil silme türlerine göre ayrı ayrı hesaplanan I. Tip hata oranlarının birbirinden oldukça farklılaştığı göz önünde bulundurulduğunda, Sinharay ve Johnson'ın bulgularını iyicil silme türlerine göre ayrılmadan hesaplanan I. Tip hata oranı bulgularıyla doğrudan karşılaştırmanın yanıltıcı olabileceği düşünülmektedir. Ayrıca Sinharay ve Johnson'ın tahrifatlı birey yüzdesinin, araştırmadaki tahrifatlı birey yüzdesinden (yaklaşık %7) oldukça yüksek olmasının I. Tip hata oranlarını etkileyebileceği düşünülmektedir. Tablo 18 incelendiğinde genel olarak I. Tip hata oranlarının alfa seviyelerinin çoğunda nominal düzeyi aştığı görülmektedir. Tahrifatın çoğunlukla düşük yetenekli bireylerde gerçekleştirilmesi beklendiğinden 500 ve 5000 test alanında olduğu gibi I. Tip hatanın düşük yetenek düzeylerinde iyi kontrol edilememesi Wollack ve diğerlerinin (2015) de belirttiği üzere bir problem olarak görülmektedir.

160 Maddelik Veri Seti

160 maddelik veri seti üzerinden hesaplanan I. Tip hata oranlarının ise tüm yetenek düzeyleri için tüm alfa seviyelerinde, nominal düzeyi aştığı ve I. Tip hata oranlarının iyi kontrol edilemediği görülmektedir. Dolayısıyla 50 maddelik veri setinde, 160 maddelik veri setine göre I. Tip hata oranları daha iyi kontrol edilmektedir. Tahrifatlı olmayan test alanlar 160 maddelik veri setinde, 50 maddelik veri setine göre daha fazla oranda tahrifatlı olarak işaretlenmektedir.

50 maddelik veri setinde en yüksek yetenek düzeyi için, nominal düzeye göre kabul edilebilir olan büyük alfa seviyelerindeki I. Tip hata oranları Bradley'in (1978) esnek referans aralığına göre değerlendirildiğinde referans aralığın altında kalarak kabul edilemez olduğu görülmektedir. Bununla birlikte 4. yetenek düzeyinde $\alpha = .05$ seviyesinde ve 5. yetenek düzeyinde $\alpha = .0005$ seviyesinde I. Tip hatanın kontrol edildiği görülmektedir. Tablo 18'deki araştırma bulgularından farklı olarak, Bradley'in esnek referansı kullanıldığında süreklilik düzeltmesi uygulanmayan EDI'nin hiçbir koşulda I. Tip hatayı kontrol etmediği; EDI_WTR'nin ise $\alpha = .05$ düzeyinde hiçbir koşulda I. Tip hatayı kontrol edemezken $\alpha = .01$ düzeyinde tüm koşullarda I. Tip hatayı kontrol ettiği belirtilmiştir (Qin, 2016). 160 maddelik veri setinde hata oranlarının değerlendirilmesinde esnek referans aralığı kullanıldığında da 500 ve 5000 test alan için olduğu gibi hiçbir durumda I. Tip hata oranları kontrol edilememiştir.

Aynı veri setlerine süreklilik düzeltmesi yapılarak hesaplanan I. Tip hata oranları Tablo 19'da yer almaktadır.

Tablo 19

Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi İle SBI'nin I. Tip Hataları (n= 50000)

50 Madde								160 Madde							
Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
1	<u>0.011</u>	<u>0.006</u>	<u>0.0050</u>	0.0035	0.00318	0.00237	0.001765	1	<u>0.028</u>	<u>0.012</u>	0.0096	0.0053	0.00432	0.00287	0.001696
2	<u>0.021</u>	<u>0.010</u>	0.0078	0.0051	0.00441	0.00324	0.002061	2	<u>0.056</u>	0.022	0.0157	0.0078	0.00572	0.00320	0.001661
3	<u>0.018</u>	<u>0.006</u>	<u>0.0042</u>	0.0020	0.00140	0.00067	0.000417	3	<u>0.065</u>	0.019	0.0121	0.0052	0.00399	0.00213	0.001009
4	<u>0.007</u>	<u>0.002</u>	<u>0.0010</u>	<u>0.0003</u>	<u>0.00020</u>	<u>0.00006</u>	0.000017	4	<u>0.052</u>	<u>0.010</u>	<u>0.0065</u>	0.0033	0.00254	0.00140	0.000464
5	<u>0.001</u>	<u>0.000</u>	<u>0.0001</u>	<u>0.0000</u>	<u>0.00001</u>	<u>0.00000</u>	<u>0.000000</u>	5	<u>0.029</u>	<u>0.005</u>	<u>0.0039</u>	0.0022	0.00162	0.00071	0.000119
Toplam	<u>0.012</u>	<u>0.005</u>	<u>0.0036</u>	0.0022	0.00183	0.00126	0.000845	Toplam	<u>0.046</u>	<u>0.014</u>	0.0095	0.0047	0.00362	0.00205	0.000984

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 19'a göre 50 maddelik veri setinde süreklilik düzeltmesi yapılarak hesaplanan I. Tip hata oranlarına ilişkin bulgular genel olarak Tablo 15'te yer alan 5000 test alan üzerinden elde edilen bulgulara benzerlik göstermekle birlikte daha fazla sayıda I. Tip hata oranının kontrol edildiği görülmektedir.

Tablo 19 incelendiğinde süreklilik düzeltmesi uygulanmasının, birçok koşulda I. Tip hata oranının iyi kontrol edilmesini sağladığı görülmektedir. 50 maddelik veri setinde süreklilik düzeltmesi yapılarak hesaplanan I. Tip hata oranları, yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden incelendiğinde büyük alfa seviyelerinde I. Tip hatanın iyi kontrol edildiği görülmektedir. Ancak alfa seviyesi azaldıkça I. Tip hata oranları nominal düzeyi aşmaktadır. Sinharay ve diğerleri (2017) $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ seviyelerinde, Wollack ve diğerleri (2015) ise tüm alfa seviyelerinde süreklilik düzeltmesi ile hesaplanan I. Tip hata oranlarının kontrol edildiğini belirtmiştir. Araştırma bulguları $\alpha = .001$ seviyesi dışında Sinharay ve diğerleriyle benzerlik göstermekte; Wollack ve diğerlerinin ise I. Tip hatayı çok daha iyi kontrol ettiği görülmektedir. Qin (2016) süreklilik düzeltmesi ile EDI_WTR'nin hiçbir koşulda ve hiçbir alfa seviyesinde I. Tip hatayı kontrol edemediğini ifade etmiştir ancak EDI_WTR'nin SBİ'nin özel bir kullanımı olduğu göz ardı edilmemelidir.

Yetenek düzeylerine göre I. Tip hata oranlarının değişimi incelendiğinde $\alpha = .05$ için tüm yetenek düzeylerinde I. Tip hata oranlarının nominal düzeyin altında kaldığı görülmektedir. Orta yetenek düzeyi ve altında, 1, 2 ve 3. yetenek düzeylerinde, en büyük alfa seviyelerinde I. Tip hata oranları iyi kontrol edilirken alfa seviyesi azaldıkça I. Tip hata oranlarının nominal düzeyi aştığı görülmektedir. 4. yetenek düzeyinde en düşük alfa seviyesi dışında; 5. yetenek düzeyinde ise tüm alfa seviyelerinde I. Tip hata oranlarının iyi kontrol edilmektedir. Süreklilik düzeltmesi uygulandığı halde I. Tip hatanın düşük yetenek düzeylerinde daha düşük alfa seviyelerinde iyi kontrol edilememesi problemleri bir durumdur ancak süreklilik düzeltmesi ile daha fazla hata oranının nominal düzeyin altında kalması da önemlidir. Wollack ve diğerlerinin (2015) en düşük yetenek düzeyinin en büyük alfa seviyesi dışında tüm yetenek düzeyleri ve alfa seviyelerinde I. Tip hata oranlarını iyi kontrol ettiği görülmektedir. Sinharay ve Johnson'ın (2016) dizi sonu silmeleri için neredeyse tüm yetenek düzeylerinde ve alfa seviyelerinde I. Tip hata oranlarını iyi kontrol ettiği görülmektedir. Tesadüfi silmeler için ise tüm yetenek düzeylerinde ve tüm alfa seviyelerinde I. Tip hata oranlarının nominal düzeyin altında kaldığı görülmektedir. Sinharay ve Johnson'ın dizi sonu tesadüfi silmeleri için hesapladığı I. Tip hata oranlarının birbirine oldukça benzer olduğu göz önünde bulundurulduğunda, araştırma

bulgularının Sinharay ve Johnson'ın bulgularıyla karşılaştırılmasının uygun olduğu düşünülmektedir. Buna göre Tablo 19, Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) bulgularıyla karşılaştırıldığında I. Tip hata oranlarının ilgili araştırmalarda daha iyi kontrol edildiği görülmektedir.

160 Maddelik Veri Seti

160 maddelik veri seti için süreklilik düzeltmesi ile hesaplanan I. Tip hata oranları süreklilik düzeltmesi yapılmadan hesaplanan hata oranlarının yer aldığı Tablo 18 ile karşılaştırıldığında bazı koşullar için hata oranlarının kontrol edilebildiği görülmektedir. I. Tip hata oranları yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden incelendiğinde, yalnızca en büyük alfa seviyesinde I. Tip hata oranlarının kontrol edildiği görülmektedir.

I. Tip hata oranlarının yetenek düzeylerine göre değişimi incelendiğinde, 1. yetenek düzeyinin en büyük alfa seviyesinde hatanın iyi kontrol edildiği görülmektedir. 4. yetenek düzeyinde yalnızca $\alpha = .01$ seviyesinde; 5. yetenek düzeyinde ise en büyük üç alfa seviyesinde I. Tip hata oranları nominal düzeyi aşmamaktadır. 2 ve 3. yetenek düzeyinde ise tüm alfa seviyelerinde I. Tip hata oranlarının kontrol edilemediği görülmektedir. Sinharay ve diğerlerinin (2017) 100 madde üzerinden yürüttükleri çalışmada ise I. Tip hata oranlarının $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ düzeylerinde nominal düzeyi aşmadığı görülmektedir. 500 örneklem büyüklüğünden 5000 örneklem büyüklüğüne geçildiğinde araştırma bulgularının Sinharay ve diğerlerine kısmen de olsa benzerlik göstermeye başladığı ifade edilmiştir ancak 5000 örneklem büyüklüğünden 50000 örneklem büyüklüğüne geçildiğinde araştırma bulgularının benzerliğinde bir artış görülmemiştir. Süreklilik düzeltmesi uygulandığı halde genel olarak 160 maddelik veri seti için I. Tip hata oranlarının nominal düzeyin çoğunu aştığı görülmektedir. Tablo 19 bulguları, 5000 test alan için bulguların yer aldığı Tablo 15 ile karşılaştırıldığında genel olarak hata oranlarında az miktarda düşüş olduğu ve yalnızca bir koşul için daha I. Tip hatanın kontrol edildiği görülmektedir. 160 maddelik veri seti için de I. Tip hatanın düşük yetenek düzeylerinde iyi kontrol edilememesi Wollack ve diğerlerinin (2015) de belirttiği gibi istenmeyen bir durumdur. Ayrıca süreklilik düzeltmesinin madde sayısının az olduğu durumlarda I. Tip hatanın şişme eğilimde olmasından dolayı kullanımının önerildiği Wollack ve diğerlerinin (2015) göz önünde bulundurulduğunda, 160 madde sayısının oldukça yüksek olmasına rağmen I. Tip hatanın 5000 test alanında olduğu gibi hala yeterince kontrol edilememesi dikkat çekicidir.

Süreklilik düzeltmesi yapılarak hesaplanan I. Tip hata oranlarında diğer örneklemelerde olduğu gibi 50 maddelik veri setinde, 160 maddelik veri setine göre hata oranları daha iyi kontrol edilmektedir. Süreklilik düzeltmesi, beklendiği gibi, tüm koşullarda I. Tip hata oranının düşmesine neden olmaktadır. Bu bakımdan bulgular 500 ve 5000 test alan bulguları ile benzerlik göstermektedir.

I. Tip hatanın değerlendirilmesi için Bradley'in (1978) esnek referansı kullanıldığında 50 maddelik veri seti için $\alpha = .05$ seviyesindeki tüm yetenek düzeylerinde, alfa seviyesi kriterine göre kabul edilebilir olan I. Tip hata oranlarının referans aralığının altında kalarak kontrol edilemediği görülmektedir. Qin (2016) tarafından Bradley'in (1978) esnek referansı kullanıldığında süreklilik düzeltmesi uygulanan EDI_WTR'nin yalnızca $\alpha = .05$ düzeyinde I. Tip hatayı kontrol ettiği belirtilmiştir. Araştırma bulguları Qin (2016) ile benzerlik göstermemekle birlikte Qin'in SBI'nin özel hali olan EDI_WTR'yi kullandığının da göz önünde bulundurulması gerekmektedir. 5. yetenek düzeyinde nominal düzeye göre kabul edilebilir olan tüm alfa seviyelerinin referans aralığının dışında kaldığı görülmektedir. 4. yetenek düzeyinde $\alpha = .0001$ seviyesinden düşük alfa seviyelerindeki hata oranlarının referans aralığının altında kaldığı ancak $\alpha = .0001$ düzeyinde I. Tip hata oranlarının kontrol edildiği görülmektedir. 50 maddelik veri seti için I. Tip hata oranlarının değerlendirilmesinde esnek referans aralığının kullanılması 500 test alana ilişkin Tablo 11 bulgulardan farklı ancak 5000 test alana ilişkin Tablo 15 bulgularına benzer olarak, kontrol altına alınan hata oranı sayısını azaltacak yönde etki göstermektedir. 160 maddelik veri seti için ise esnek referans aralığı kullanıldığında genel olarak kontrol edilen hata oranlarının sayısı artmaktadır. Yetenek düzeyi ayrımı yapılmaksızın I. Tip hata oranları incelendiğinde $\alpha = .05$ ve $\alpha = .01$ düzeylerinde hata oranlarının iyi kontrol edildiği görülmektedir. Tüm yetenek düzeylerinde $\alpha = .05$ seviyesindeki I. Tip hata oranlarının kontrol edildiği; en düşük yetenek düzeyi için $\alpha = .01$ seviyesinde de hata oranlarının referans aralığı aşmadığı görülmektedir. 4. yetenek düzeyinde ise en büyük üç alfa seviyesinde de hata oranları kontrol edilmektedir.

50000 test alan için SBI'nin süreklilik düzeltmesi yapılmaksızın güç oranları Tablo 20'de yer almaktadır.

Tablo 20

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 50000)

50 Madde									160 Madde									
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	
%10 (5 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%10 (16 Silme)	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2	1.000	1.000	1.000	1.000	1.000	0.999	0.994		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	0.998	0.983	0.967	0.895	0.724		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	0.997	0.911	0.817	0.537	0.428	0.235	0.085		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995
	5	<u>0.844</u>	<u>0.294</u>	<u>0.156</u>	<u>0.033</u>	<u>0.016</u>	0.003	0.000		5	1.000	1.000	1.000	1.000	0.998	0.949	0.660	0.660
	Toplam	0.968	0.841	0.794	0.711	0.682	0.626	0.561		Toplam	1.000	1.000	1.000	1.000	1.000	0.990	0.931	
%20 (10 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%20 (32 Silme)	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	1.000	1.000	1.000	0.999	0.991		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	1.000	0.998	0.990	0.906	0.839	0.632	0.368		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.987</u>	<u>0.676</u>	<u>0.474</u>	<u>0.150</u>	<u>0.081</u>	0.014	0.000		5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	Toplam	0.997	0.935	0.893	0.811	0.784	0.729	0.672		Toplam	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
%30 (15 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%30 (48 Silme)	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	1.000	0.999	0.996	0.964	0.924	0.765	0.511		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.984</u>	<u>0.699</u>	<u>0.521</u>	<u>0.201</u>	<u>0.110</u>	0.017	0.000		5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Toplam	0.997	0.940	0.903	0.833	0.807	0.756	0.702		Toplam	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
Değişken Silme	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	Değişken Silme	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2	1.000	1.000	1.000	1.000	1.000	0.999	0.992		2	1.000	1.000	1.000	1.000	1.000	1.000	0.999	
	3	0.936	0.846	0.806	0.697	0.658	0.574	0.451		3	0.971	0.903	0.868	0.780	0.744	0.657	0.523	
	4	—	—	—	—	—	—	—		4	—	—	—	—	—	—	—	—
	5	—	—	—	—	—	—	—		5	—	—	—	—	—	—	—	—
	Toplam	0.979	0.949	0.935	0.899	0.886	0.857	0.815		Toplam	0.990	0.968	0.956	0.927	0.915	0.886	0.841	

(Devam ediyor)

Tablo 20 (Devam)

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 50000)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
Toplam	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	Toplam	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	1.000	1.000	1.000	1.000	0.999	0.997		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	0.984	0.961	0.951	0.920	0.906	0.867	0.791		3	0.993	0.976	0.967	0.945	0.936	0.914	0.881
	4	<i>0.999</i>	<i>0.970</i>	<i>0.935</i>	<i>0.802</i>	<i>0.730</i>	<i>0.544</i>	<i>0.321</i>		4	1.000	1.000	1.000	1.000	1.000	1.000	0.998
	5	<u>0.938</u>	<u>0.556</u>	<u>0.384</u>	<u>0.128</u>	<u>0.069</u>	0.012	0.000		5	1.000	1.000	1.000	1.000	0.999	0.983	0.886
	Toplam	<i>0.986</i>	<i>0.912</i>	<i>0.876</i>	<i>0.804</i>	<i>0.779</i>	<i>0.730</i>	<i>0.673</i>		Toplam	0.998	0.995	0.993	0.988	0.986	0.978	0.954

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Yetenek düzeylerine göre ayrılmadan toplam test alanlar üzerinden I. Tip hata kontrol edilemediği için istatistiksel güç değerlendirilememektedir. İstatistiksel güç tahrifat miktarına göre ayrılmaksızın toplam test alanlar üzerinden yetenek düzeylerine göre incelendiğinde, 1. yetenek düzeyinde tahrifatlı bireylerin $\alpha = .05$ seviyesinde mükemmel düzeyde, 5. yetenek düzeyinde ise yüksek düzeyde doğru belirlendiği görülmektedir. 5. yetenek düzeyi için diğer alfa seviyelerinde istatistiksel güç kabul edilebilir düzeyde değildir.

İstatistiksel güç tahrifat miktarına göre incelendiğinde 5. yetenek düzeyi için tahrifat oranı arttıkça yetenek düzeylerinde istatistiksel gücün arttığı görülmektedir. 1. yetenek düzeyinde en düşük tahrifat oranında dahi istatistiksel güç mükemmel olduğu için tahrifat oranı arttıkça gücün değişimi gözlenememektedir. %10 oranında tahrifat (5 silme) yapıldığı durumda $\alpha = .05$ düzeyinde 1. yetenek düzeyinde tahrifatlı bireyler mükemmel düzeyde, 5. yetenek düzeyinde ise tatmin edici düzeyde belirlenmektedir. 5. yetenek düzeyi için $\alpha = .01$ seviyesinde gücün keskin bir şekilde düştüğü ve diğer alfa seviyelerinde istatistiksel gücün kabul edilebilir düzeyde olmadığı görülmektedir. %20 oranında tahrifat (10 silme) yapıldığı durumda 1. yetenek düzeyinde $\alpha = .05$ düzeyinde istatistiksel güç mükemmel düzeyde; 5. yetenek düzeyinde ise $\alpha = .05$ seviyesinde güç çok yüksek iken diğer alfa seviyelerinde istatistiksel güç kabul edilebilir düzeyde değildir. %30 oranında tahrifat (15 silme) yapıldığı durumda %20 oranında tahrifat (10 silme) yapıldığı duruma benzer olarak, 1. yetenek düzeyinde $\alpha = .05$ düzeyinde istatistiksel gücün mükemmel olduğu; 5. yetenek düzeyinde ise $\alpha = .05$ seviyesinde çok yüksek iken diğer alfa seviyelerinde istatistiksel gücün kabul edilebilir düzeyde olmadığı görülmektedir. Değişken sayıda silme yapıldığı durumda da 1. yetenek düzeyinde $\alpha = .05$ seviyesinde istatistiksel güç mükemmeldir. Yüksek yetenek düzeylerinde ise tahrifata uğrayan test alan olmadığı için istatistiksel güç hesaplanmamıştır.

160 Maddelik Veri Seti

160 maddelik veri seti için hiçbir yetenek düzeyinde hiçbir alfa seviyesinde I. Tip hata kontrol edilemediği için istatistiksel güç dikkate alınmamaktadır.

Tablo 20 bulguları, 5000 test alan için bulguların yer aldığı Tablo 16 ile karşılaştırıldığında kabul edilebilir istatistiksel güç değerlerinin sayısının daha az olduğu görülmektedir. Tablo19 genel olarak incelendiğinde benzer araştırmalarda da olduğu gibi, tahrifat oranı arttıkça istatistiksel gücün arttığı ancak alfa seviyesi azaldıkça istatistiksel gücün

düştüğü görülmektedir (Qin, 2016; Sinharay vd., 2017; Wollack vd., 2015). Wollack ve diğerlerinin (2016) de belirttiği üzere yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir.

Tablo 20’de yer alan 50 maddelik veri seti için istatistiksel güç değerleri Bradley’in esnek referans aralığına göre incelendiğinde ise yetenek düzeylerine göre ayırım yapılmadığı durumda en büyük alfa seviyesinde istatistiksel gücün tüm tahrifat oranı koşullarında ve toplam test alanlarda çok yüksek olduğu görülmektedir. 5. yetenek düzeyinde en büyük alfa seviyelerindeki istatistiksel güç değerlerinin I. Tip hata oranlarının kontrol edilememesi nedeniyle değerlendirme dışında kaldığı ancak $\alpha = .0005$ seviyesinde istatistiksel gücün değerlendirilebildiği ancak onun da tüm silme koşulları için kabul edilebilir düzeyin oldukça altında olduğu görülmektedir. 4. yetenek düzeyi için $\alpha = .05$ seviyesinde hata oranlarının referans aralığında kalması nedeniyle değerlendirilebilir olduğu ve istatistiksel gücün tüm koşullarda neredeyse mükemmel olduğu görülmektedir. 160 maddelik veri seti için I. Tip hata oranları Bradley’in esnek referans aralığına göre değerlendirildiğinde de kontrol edilemediğinden istatistiksel güç hiçbir koşulda değerlendirilememektedir.

50000 test alan için süreklilik düzeltmesi yapılarak uygulanan SBİ’nin istatistiksel gücü Tablo 21’de yer almaktadır.

Tablo 21

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 50000)

50 Madde									160 Madde									
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	
%10 (5 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%10 (16 Silme)	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	<u>1.000</u>	1.000	1.000	0.998	0.994	0.963		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>1.000</u>	<u>0.995</u>	<u>0.985</u>	0.918	0.871	0.722	0.515		3	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>0.959</u>	<u>0.701</u>	<u>0.553</u>	<u>0.284</u>	<u>0.207</u>	<u>0.089</u>	0.028		4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.977
	5	<u>0.455</u>	<u>0.083</u>	<u>0.036</u>	<u>0.006</u>	<u>0.002</u>	<u>0.000</u>	<u>0.000</u>		5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	0.996	0.981	0.844	0.459	
	Toplam	<u>0.883</u>	<u>0.756</u>	<u>0.715</u>	0.642	0.616	0.561	0.501		Toplam	<u>1.000</u>	<u>1.000</u>	1.000	0.999	0.996	0.969	0.887	
%20 (10 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%20 (32 Silme)	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	0.996	0.962		3	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>1.000</u>	<u>0.983</u>	<u>0.947</u>	<u>0.776</u>	<u>0.676</u>	<u>0.462</u>	0.229		4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	5	<u>0.876</u>	<u>0.389</u>	<u>0.233</u>	<u>0.048</u>	<u>0.021</u>	<u>0.001</u>	<u>0.000</u>		5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.995
	Toplam	<u>0.975</u>	<u>0.874</u>	<u>0.836</u>	0.765	0.739	0.692	0.638		Toplam	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.999	
%30 (15 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%30 (48 Silme)	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	0.999	0.995		3	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>1.000</u>	<u>0.996</u>	<u>0.985</u>	<u>0.895</u>	<u>0.822</u>	<u>0.628</u>	0.389		4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	5	<u>0.888</u>	<u>0.470</u>	<u>0.308</u>	<u>0.082</u>	<u>0.034</u>	<u>0.003</u>	<u>0.000</u>		5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.998
	Toplam	<u>0.978</u>	<u>0.893</u>	<u>0.859</u>	0.795	0.771	0.726	0.677		Toplam	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	
Değişken Silme	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	Değişken Silme	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	<u>1.000</u>	1.000	0.999	0.998	0.992	0.978		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	0.998
	3	<u>0.785</u>	<u>0.653</u>	<u>0.610</u>	0.503	0.463	0.385	0.284		3	<u>0.877</u>	<u>0.774</u>	<u>0.731</u>	<u>0.633</u>	<u>0.584</u>	<u>0.488</u>	<u>0.370</u>	
	4	—	—	—	—	—	—	—		4	—	—	—	—	—	—	—	—
	5	—	—	—	—	—	—	—		5	—	—	—	—	—	—	—	—
	Toplam	<u>0.928</u>	<u>0.884</u>	<u>0.870</u>	0.834	0.820	0.792	0.754		Toplam	<u>0.959</u>	<u>0.925</u>	0.910	0.878	0.861	0.829	0.789	

(Devam ediyor)

Tablo 21 (Devam)

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 50000)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
Toplam	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	Toplam	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	2	<u>1.000</u>	<u>1.000</u>	1.000	1.000	0.999	0.996	0.985		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>0.946</u>	<u>0.912</u>	<u>0.899</u>	0.855	0.834	0.776	0.689		3	<u>0.969</u>	0.944	0.933	0.908	0.896	0.872	0.843
	4	<u>0.986</u>	<u>0.893</u>	<u>0.828</u>	<u>0.652</u>	<u>0.568</u>	<u>0.393</u>	0.215		4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	0.992
	5	<u>0.739</u>	<u>0.314</u>	<u>0.192</u>	<u>0.045</u>	<u>0.019</u>	<u>0.001</u>	<u>0.000</u>		5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	0.999	0.994	0.948	0.817
	Toplam	<u>0.942</u>	<u>0.848</u>	<u>0.814</u>	0.751	0.727	0.682	0.630		Toplam	<u>0.993</u>	<u>0.987</u>	<u>0.985</u>	<u>0.979</u>	<u>0.976</u>	<u>0.963</u>	<u>0.933</u>

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

İstatistiksel güç tahrifat oranlarına ve yetenek düzeylerine göre ayrılmaksızın toplam test alanlar üzerinden incelendiğinde, en büyük üç alfa seviyesinde istatistiksel gücün oldukça yüksek olduğu ve alfa seviyesi düştükçe istatistiksel güç azaldığı görülmektedir. İstatistiksel güç tahrifat miktarına göre ayrılmaksızın toplam test alanlar üzerinden yetenek düzeylerine göre incelendiğinde, genel olarak yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir. 1. yetenek düzeyinde en büyük üç alfa seviyesinde tahrifatlı bireyler mükemmel düzeyde belirlenmektedir ancak daha düşük alfa seviyelerinde hatalar iyi kontrol edilemediği için istatistiksel güç değerlendirilememiştir. 2. yetenek düzeyinde $\alpha = .05$ ve $\alpha = .01$ seviyelerinde istatistiksel gücün değerlendirilebildiği ve bu seviyelerde gücün mükemmel olduğu görülmektedir. 3. yetenek düzeyinde en büyük üç alfa seviyesinde istatistiksel gücün değerlendirilebildiği ve bu seviyelerde istatistiksel gücün oldukça yüksek olduğu görülmektedir. 4. yetenek düzeyinde $\alpha = .05$ seviyesinde güç çok yüksek olmakla birlikte, alfa seviyesi küçüldükçe istatistiksel güç düşmektedir ancak $\alpha = .005$ seviyesinde güç hala tatmin edici düzeydedir. 5. yetenek düzeyinde $\alpha = .05$ seviyesinde gücün kabul edilebilir düzeyde olduğu ancak diğer alfa seviyelerinde I. Tip hata kontrol edilmesine rağmen kabul edilebilir istatistiksel gücün elde edilemediği görülmektedir.

İstatistiksel güç tahrifat miktarına göre incelendiğinde ise 500 ve 5000 test alan için de geçerli olan, Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) da belirttiği üzere tahrifat oranı arttıkça yetenek düzeylerinde istatistiksel gücün arttığı görülmektedir. %10 oranında tahrifat (5 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde Sinharay ve diğerlerinin (2017) aksine istatistiksel gücün $\alpha = .05$ seviyesinde tatmin edici olduğu, alfa seviyesi azaldıkça gücün düştüğü ancak $\alpha = .005$ seviyesinde ise hala kabul edilebilir düzeyde olduğu belirlenmiştir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde orta yetenek düzeyi ve altında en büyük alfa seviyelerinde tahrifatlı bireylerin neredeyse mükemmel düzeyde belirlendiği görülmektedir. 4. yetenek düzeyinde $\alpha = .05$ seviyesinde istatistiksel gücün yüksek, $\alpha = .01$ seviyesinde ise kabul edilebilir düzeyde olduğu görülürken alfa seviyesi küçüldükçe istatistiksel gücün kabul edilemez düzeyde düştüğü görülmektedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edilmesine rağmen hiçbir alfa seviyesinde kabul edilebilir istatistiksel güç elde edilemediği görülmektedir. Wollack ve diğerleri (2015) tarafından 5 silme yapıldığı durumda yalnızca 1. yetenek düzeyinde ve $\alpha = .05$ düzeyinde çok yüksek istatistiksel güç; 1. yetenek düzeyi için $\alpha = .01$ ve 2. yetenek düzeyi için $\alpha = .05$ düzeylerinde kabul edilebilir

istatistiksel güç rapor edilmiştir. Araştırma bulguları, $\alpha = .05$ ve $\alpha = .01$ seviyelerinde en üst yetenek düzeyine kadar yüksek istatistiksel güç elde edilmesi ile Wollack ve diğerlerinden (2015) farklılaşmaktadır. Ancak en üst yetenek düzeyinde Wollack ve diğerlerinden (2015) daha yüksek istatistiksel güç elde edilmesine rağmen güç oranları kabul edilebilir değildir. 5. yetenek düzeyinde kabul edilebilir istatistiksel güç elde edilememesi açısından bulguların benzerlik gösterdiği görülmektedir. Sinharay ve Johnson'ın (2016) bulgularında ise 5 silme yapıldığı durumda istatistiksel güç değerlerinin hiçbir yetenek düzeyinde ve alfa seviyesinde kabul edilebilir olmadığı görülmektedir.

%20 oranında tahrifat (10 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün en büyük üç alfa seviyesinde de tatmin edici olduğu görülmektedir. Sinharay ve diğerlerinde (2017) ise 8 silme yaparak hesaplanan istatistiksel güç yalnızca $\alpha = .05$ seviyesinde kabul edilebilir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde en yüksek yetenek düzeyine kadar büyük alfa seviyelerinde tahrifatlı bireylerin mükemmel düzeyde belirlendiği görülmektedir. 4. yetenek düzeyinde büyük alfa seviyelerinde tahrifatlı bireylerin büyük oranda doğru belirlendiği, $\alpha = .001$ seviyesinde dahi istatistiksel gücün kabul edilebilir düzeyde olduğu görülmektedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edilmesine rağmen yalnızca $\alpha = .05$ seviyesinde istatistiksel gücün tatmin edici olduğu görülmektedir. Sinharay ve Johnson'ın (2016) 10 silme yapıldığı durumda kabul edilebilir istatistiksel güç değerlerini 1. yetenek düzeyinde $\alpha = .05$ ve $\alpha = .01$ düzeylerinde; 2. yetenek düzeyinde ise yalnızca $\alpha = .05$ düzeyinde elde ettiği görülmektedir. Araştırma bulguları, Sinharay ve Johnson'ın düşük yetenek düzeylerinin daha düşük alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilememesi bulgularıyla benzerlik göstermektedir. Wollack ve diğerlerinin (2015) çalışmasında ise 10 silme yapıldığı durumda $\alpha = .05$ seviyesinde 3. yetenek düzeyine kadar istatistiksel güç mükemmel yakınken, 3. yetenek düzeyinde güç çok yüksek ve 4. yetenek düzeyinde ise kabul edilebilir düzeydedir. En düşük yetenek düzeyinde neredeyse en küçük alfa seviyelerine kadar istatistiksel gücün oldukça yüksek olduğu; 2. yetenek düzeyinde istatistiksel gücün büyük alfa seviyelerinde kabul edilebilir olduğu görülmektedir. Araştırma bulguları, en üst yetenek düzeyinin en büyük alfa seviyesinde yüksek istatistiksel güç elde edilmesi ve düşük yetenek düzeylerinin küçük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden farklılaşmaktadır.

%30 oranında tahrifat (15 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün $\alpha = .05$, $\alpha = .01$ ve $\alpha = .005$ seviyelerinde tatmin edici olduğu görülmektedir. İstatistiksel güç yetenek düzeylerine göre

incelendiğinde orta yetenek düzeyi ve altında I. Tip hatanın iyi kontrol edildiği en büyük alfa seviyelerinde istatistiksel gücün mükemmel olduğu görülmektedir. 4. yetenek düzeyinde en büyük alfa seviyesinde istatistiksel güç mükemmel iken alfa seviyesi azaldıkça güç düşmekte ancak $\alpha = .0005$ seviyesinde bile istatistiksel güç tatmin edici görünmektedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edilmesine rağmen, 10 silme yapıldığı durumdakine benzer olarak, yalnızca $\alpha = .05$ seviyesinde istatistiksel gücün çok yüksek olduğu ve diğer alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilemediği görülmektedir. Wollack ve diğerlerinin (2015) çalışmasında 15 silme yapıldığı durumda $\alpha = .05$ seviyesinde en düşük yetenek düzeylerinde istatistiksel gücün neredeyse mükemmel olduğu, 3. yetenek düzeyinde ise çok yüksek olduğu görülmektedir. 1. yetenek düzeyinde en düşük alfa seviyesinde bile istatistiksel güç tatmin edicidir. 2. yetenek düzeyinde istatistiksel gücün $\alpha = .0005$ seviyesine kadar kabul edilebilir olduğu görülmektedir. Araştırma bulguları 10 silme yapıldığı durumdakine benzer şekilde, en üst yetenek düzeyinin en büyük alfa seviyesinde yüksek istatistiksel güç elde edilmesi ve düşük yetenek düzeylerinin küçük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden (2015) farklılaşmaktadır.

Değişken sayıda silme yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün en büyük üç alfa seviyesinde de yüksek olduğu görülmektedir. İstatistiksel güç düşük yetenek düzeylerinin büyük alfa seviyelerinde mükemmeldir. Orta yetenek düzeyinde ise yalnızca $\alpha = .05$ seviyesinde kabul edilebilir istatistiksel güç elde edilmiştir. Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) çalışmalarında da görüldüğü gibi en yüksek yetenek düzeylerinde tahrifata uğrayan test alan olmadığı için istatistiksel güç hesaplanmamıştır. Wollack ve diğerlerinin (2015) çalışmasında değişken sayıda silme yapıldığı durumda, en düşük alfa seviyesi dışında 1. yetenek düzeyinde tüm alfa seviyelerinde istatistiksel gücün kabul edilebilir olduğu ancak 2. ve 3. yetenek düzeylerinde kabul edilebilir istatistiksel gücün olmadığı görülmektedir. Araştırma bulguları, diğer yetenek düzeylerinde de kabul edilebilir düzeyde istatistiksel güç elde edilmesi ve 1. yetenek düzeyinde düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden farklılaşmaktadır. Sinharay ve Johnson'ın (2016) bulgularında ise değişken silme yapıldığı durumda kabul edilebilir istatistiksel güç yalnızca $\alpha = .05$ düzeyinde 1. yetenek düzeyinde elde edilmiştir. Genel olarak Sinharay ve Johnson'ın (2016) tüm koşullar için istatistiksel güç değerleri oldukça düşüktür. Sinharay ve Johnson'ın (2016) tahrifatlı birey yüzdesinin (%50), araştırmadaki tahrifatlı birey yüzdesinden (yaklaşık %8) oldukça yüksek olmasının istatistiksel gücü etkileyebileceği düşünülmektedir. Qin'in (2016)

araştırma bulguları EDI_WTR'nin bazı koşullarında tahrifat yüzdesinin artmasının istatistiksel gücü bir miktar düşürdüğünü göstermektedir.

Tüm tahrifat miktarı koşulları için yetenek düzeyi arttıkça kontrol edilen I. Tip hata oranı sayısının arttığı ancak istatistiksel gücün düştüğü görülmektedir. Yüksek yetenek düzeylerinde tahrifat oranı arttıkça istatistiksel gücün arttığı görülmektedir. Düşük yetenek düzeylerinde en düşük tahrifat oranı koşulunda bile tahrifatlı bireyler mükemmel şekilde belirlendiği için gücün değişimi gözlenmemiştir. Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) da belirttiği üzere tahrifat oranı arttıkça istatistiksel gücün arttığı ancak alfa seviyesi azaldıkça ve aynı tahrifat oranı için yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir. Wollack ve diğerlerine (2015) göre tahrifatlı bir madde için, düşük yetenek düzeyindeki test alanın doğru yanıt olasılık ortalamasına sahip olarak maddeyi doğru yanıtlama olasılığı düşük olduğundan kolaylıkla SBİ ile tahrifatlı olarak belirlenmektedir. Yüksek yetenek düzeyindeki test alanların ise doğru yanıt olasılık ortalaması yüksek olduğundan maddeyi doğru yanıtlanması beklenmekte ve SBİ tarafından tahrifatlı olarak işaretlenmemektedir. Sonuç olarak tahrifatlı test alanların yüksek yetenek düzeylerinde SBİ ile belirlenmesi zorlaşmaktadır.

Süreklilik düzeltmesi yapılması, benzer araştırmalarda olduğu gibi, kontrol edilen I. Tip hata oranı sayısını artırmakta ancak istatistiksel gücün düşmesine neden olmaktadır (Qin, 2016; Sinharay vd., 2017; Wollack vd., 2015). Süreklilik düzeltmesi yapılmadığı durumda I. Tip hatanın nominal düzeyi aşmadığı koşullar için süreklilik düzeltmesi yapılmasının istatistiksel gücü ne kadar düşürdüğü incelenmiştir. Buna göre tahrifat oranı dikkate alınmadığında istatistiksel gücün, 1. yetenek düzeyinde değişmediği, 5. yetenek düzeyinde ise 0.05 ile 0.24 arasında azaldığı görülmektedir. 500 ve 5000 test alandakine benzer şekilde silme koşullarının tümünde 1. yetenek düzeyi için süreklilik düzeltmesi yapılmasının istatistiksel gücü değiştirmediği görülmektedir. 5. yetenek düzeyi için ise süreklilik düzeltmesinin yapılması istatistiksel gücü %10 oranında tahrifat (5 silme) koşulunda 0.01 ile 0.39 arasında; %20 oranında tahrifat (10 silme) koşulunda 0.06 ile 0.29 arasında; %30 oranında tahrifat (15 silme) koşulunda 0.08 ile 0.23 arasında azaltmıştır. 5000 test alanın bulgularıyla karşılaştırıldığında, 50000 test alan için istatistiksel gücün değişim ranjının az miktarda daha büyük olduğu belirlenmiştir. İstatistiksel gücün süreklilik düzeltmesine bağlı değişiminin tahrifat oranı arttıkça azaldığı görülmektedir. SBİ ile tahrifatlı bireylerin belirlenmesinde süreklilik düzeltmesi uygulanmasının I. Tip hatanın daha iyi kontrol edilmesini sağlayarak kabul edilebilir istatistiksel güç koşulunu artırdığı ve tahrifata uğraması muhtemel olan düşük yetenek düzeylerinde istatistiksel gücü neredeyse hiç etkilemediği belirlenmiştir.

Tablo 21 bulguları 5000 test alanına ilişkin bulguların yer aldığı Tablo 17 ile karşılaştırıldığında, istatistiksel gücün tüm silme koşulları için, düşük yetenek düzeylerinde önemli bir değişiklik göstermediği, üst yetenek düzeylerinde ise bir miktar düştüğü belirlenmiştir.

160 Maddelik Veri Seti

160 maddelik veri seti için istatistiksel güç tahrifat miktarına ve yetenek düzeylerine göre ayrılmaksızın toplam test alanları üzerinden incelendiğinde, I. Tip hatanın yalnızca $\alpha = .05$ düzeyinde kontrol edildiği ve bu düzeyde istatistiksel gücün mükemmel yakın olduğu görülmektedir. İstatistiksel güç tahrifat miktarına göre ayrılmaksızın toplam test alanları üzerinden yetenek düzeylerine göre incelendiğinde, 1. yetenek düzeyinde $\alpha = .05$ düzeyinde, 4. yetenek düzeyinde $\alpha = .01$ seviyesinde, 5. yetenek düzeyinde ise en büyük üç alfa seviyesinde tahrifatlı bireylerin mükemmel şekilde belirlendiği görülmektedir.

İstatistiksel güç tahrifat miktarına göre incelendiğinde ise tüm sabit silme koşullarında I. Tip hatanın kontrol edildiği durumlarda tahrifatlı bireyler mükemmel şekilde belirlenmektedir. Değişken silme koşulunda ise yetenek düzeyi ayrımı yapılmaksızın tahrifatlı bireylerin çok yüksek oranda belirlendiği görülmektedir. En düşük yetenek düzeyinde ise I. Tip hatanın kontrol edilebildiği tek durum olan $\alpha = .05$ düzeyinde istatistiksel gücün mükemmel olduğu görülmektedir. 500 ve 5000 test alan bulgularına benzer şekilde 50 maddelik veri seti ile karşılaştırıldığında daha az sayıda istatistiksel güç yorumlanabilmektedir. Buna karşılık, tüm test alanları üzerinden ve yetenek düzeyleri için aynı tahrifat miktarı oranlarında karşılaştırma yapıldığında 160 maddelik veri seti için 4. ve 5. yetenek düzeyinde istatistiksel gücün her durumda daha yüksek olduğu görülmektedir. En düşük yetenek düzeyinde her iki madde seti için de $\alpha = .05$ seviyesinde güç mükemmeldir. 5000 test alanı üzerinden istatistiksel güce ait bulguların yer aldığı Tablo 17 ile karşılaştırıldığında ise farklı bir yetenek düzeyinde, 4. yetenek düzeyinde, bir tane kabul edilebilir istatistiksel güç daha elde edildiği görülmüştür.

Tablo 21’de yer alan 50 maddelik veri seti için Bradley’in esnek referans aralığına göre $\alpha = .05$ seviyesinde tüm yetenek düzeylerinde, 5. yetenek düzeyinde tüm alfa seviyelerinde ve 4. yetenek düzeyinin neredeyse tüm alfa seviyelerinde istatistiksel güç değerleri I. Tip hata oranlarının kontrol edilememesi nedeniyle değerlendirme dışında kalmıştır. Esnek referans aralığının kullanılması hiçbir tahrifat oranında ve hiçbir yetenek düzeyinde istatistiksel gücün değerlendirilebilir hale gelmesini sağlamamıştır. 160 maddelik veri seti için esnek referans

aralığı kullanıldığında değerlendirilebilen istatistiksel güç sayısının 5. yetenek düzeyi hariç, tüm yetenek düzeylerinde ve toplam test alanlar düzeyinde arttığı görülmektedir. Yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden $\alpha = .05$ ve $\alpha = .01$ düzeylerinde tahrifatlı bireylerin çok yüksek oranda doğru belirlendiği görülmektedir. Tüm silme koşullarında $\alpha = .05$ düzeyinde tüm yetenek düzeylerinde istatistiksel gücün değerlendirilebildiği ve değişken silme koşulu dışında tüm koşullarda gücün mükemmel olduğu görülmektedir. Ayrıca 1. yetenek düzeyinde $\alpha = .01$ yetenek düzeyinde güç değerlendirilebilir hale gelmiş ve bu düzeyde güç mükemmeldir. Üst yetenek düzeylerinde de daha düşük alfa seviyelerinde istatistiksel güç değerlendirilebilir hale gelmiş ve belirtilen düzeylerde tahrifatlı bireyler mükemmel düzeyde belirlenmiştir. Tablo 21 bulguları, esnek referans aralığına göre Tablo 17 bulgularıyla da benzerlik göstermektedir.

99990 Örneklem Büyüklüğü İçin SBI'nin İstatistiksel Güç ve I. Tip Hata Bulguları

99990 örneklem büyüklüğü için I. Tip hata oranlarına ilişkin bulgular Tablo 22 ve Tablo 23'te, istatistiksel güce ilişkin bulgular ise Tablo 24 ve Tablo 25'te yer almaktadır.

Tablo 22

Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi Yapılmaksızın SBI'nin I. Tip Hataları (n= 99990)

50 Madde								160 Madde							
Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
1	<u>0.029</u>	0.022	0.0186	0.0139	0.01256	0.01017	0.007881	1	0.092	0.054	0.0480	0.0356	0.03099	0.02259	0.015597
2	<i>0.075</i>	0.047	0.0386	0.0243	0.02041	0.01469	0.009897	2	0.196	0.119	0.0914	0.0524	0.04273	0.02742	0.016295
3	0.095	0.037	0.0265	0.0159	0.01282	0.00807	0.004590	3	0.261	0.104	0.0709	0.0318	0.02313	0.01225	0.005266
4	<i>0.052</i>	<i>0.013</i>	0.0083	0.0034	0.00254	0.00150	0.000924	4	0.232	0.060	0.0332	0.0098	0.00625	0.00291	0.001177
5	<u>0.014</u>	<u>0.002</u>	<u>0.0010</u>	<u>0.0003</u>	<u>0.00029</u>	0.00016	0.000077	5	0.160	0.020	0.0091	0.0032	0.00231	0.00105	0.000215
Toplam	<i>0.053</i>	0.024	0.0185	0.0115	0.00965	0.00686	0.004638	Toplam	0.189	0.071	0.0503	0.0264	0.02094	0.01315	0.007650

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 22'ye göre 50 maddelik veri setinde süreklilik düzeltmesi yapılmaksızın I. Tip hata oranlarına ilişkin bulgular Tablo 14'te yer alan 5000 ve Tablo 18'de yer alan 50000 test alan üzerinden elde edilen bulgulara tamamen benzerlik göstermektedir. Hata oranları yetenek düzeyi ayırımı yapılmaksızın tüm test alanlar üzerinden incelendiğinde, tüm alfa seviyelerinde nominal düzeyin aşıldığı ve I. Tip hata oranlarının iyi kontrol edilemediği görülmektedir. Bu durum da, aslında tahrifatlı olmayan test alanların alfa düzeyini aşan oranda tahrifatlı olarak belirlendiğini göstermektedir. Benzer şekilde Qin (2016), $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ manidarlık düzeylerinde yürüttüğü çalışmada, tüm araştırma koşullarında I. Tip hatanın nominal seviyeleri aştığını belirtmiştir. Wollack ve diğerleri (2015) ise yalnızca düşük alfa seviyelerinde hata oranlarının kontrol edilemediğini belirtmiştir.

I. Tip hata oranlarının yetenek düzeylerine göre değişimi incelendiğinde, 1. yetenek düzeyinin en büyük alfa seviyesinde hatanın iyi kontrol edildiği görülmektedir. 5. yetenek düzeyinde ise $\alpha = .0001$ ve $\alpha = .00001$ düzeyleri dışında diğer alfa seviyelerinde I. Tip hata oranlarının nominal düzeyi aşmadığı görülmektedir. 2, 3 ve 4. yetenek düzeyinde hiçbir alfa seviyesinde, I. Tip hata oranlarının kontrol edilemediği görülmektedir. Wollack ve diğerlerinde (2015) de en düşük yetenek düzeyinde I. Tip hatanın yalnızca $\alpha = .05$ düzeyinde kontrol edilebildiği ancak Tablo 22'nin aksine diğer tüm yetenek düzeylerinde I. Tip hatanın iyi kontrol edildiği görülmektedir. Sinharay ve Johnson (2016) araştırmasında, tahrifatlı test alan yüzdesini %50 olarak belirleyip, I. Tip hata oranlarını dizi sonu ve tesadüfi silmelere göre ayrı ayrı hesaplamıştır. Dizi sonu silmeleri için Tablo 22 bulgularına benzer şekilde orta ve düşük yetenek düzeylerinde hiçbir alfa düzeyinde hatanın kontrol edilemediğini belirtmiştir. Tesadüfi silmeler için ise Tablo 22 bulgularının aksine, 1. yetenek düzeyinin en yüksek alfa seviyesi dışındaki tüm seviyelerde I. Tip hata oranlarının nominal düzeyi aşmadığı görülmektedir. Sinharay ve Johnson (2016) tarafından iyicil silme türlerine göre ayrılarak hesaplanan I. Tip hata oranlarının birbirinden oldukça farklılaştığı göz önünde bulundurulduğunda, yürütülen araştırmada iyicil silme türlerine göre ayrılmadan hesaplanan I. Tip hata oranı bulgularıyla doğrudan karşılaştırmanın yanıltıcı olabileceği düşünülmektedir. Ayrıca araştırmada kullanılan tahrifatlı birey yüzdesinden (%8) oldukça yüksek oranda (%50) yapılan tahrifatın Sinharay ve Johnson'ın I. Tip hata oranlarının farklılaşmasına neden olabileceği düşünülmektedir. Tablo 22 genel olarak incelendiğinde I. Tip hata oranlarının alfa seviyelerinin çoğunda nominal düzeyi aştığı görülmektedir. Uygulamada test tahrifatının çoğunlukla düşük yetenekli bireylerde

gerçekleşmesi beklendiğinden I. Tip hatanın düşük yetenek düzeylerinde iyi kontrol edilememesi Wollack ve diğerlerinin (2015) de belirttiği gibi istenmeyen bir durumdur.

160 Maddelik Veri Seti

160 maddelik veri seti üzerinden hesaplanan I. Tip hata oranlarının hiçbir yetenek düzeyinde ve hiçbir alfa seviyesinde kontrol edilemediği görülmektedir. Test alanlar aslında tahrifatlı olmadığı halde, alfa düzeyini aşan oranda tahrifatlı olarak belirlenmektedir. Daha küçük örneklemlerde de görüldüğü gibi 50 maddelik veri setinde, 160 maddelik veri setine göre I. Tip hata oranlarının daha iyi kontrol edildiği görülmektedir.

50 maddelik veri setinde en yüksek yetenek düzeyi için, nominal düzeye göre kabul edilebilir olan büyük alfa seviyelerindeki I. Tip hata oranlarının Bradley'in (1978) esnek referansı kullanıldığında referans aralığın dışında kalarak kontrol edilemediği görülmektedir. Bununla birlikte toplam test alanlar üzerinden ve 2. yetenek düzeyinde $\alpha = .05$ seviyesinde; 4. yetenek düzeyinde ise $\alpha = .05$ ve $\alpha = .01$ seviyelerinde I. Tip hata oranlarının referans aralığı içerisinde kaldığı görülmektedir. Tablo 22'deki araştırma bulgularından farklı olarak Qin (2016), esnek referans aralığı kullanıldığında süreklilik düzeltmesi uygulanmayan SBI'nin hiçbir koşulda I. Tip hatayı kontrol etmediği; SBI'nin özel bir kullanımı olan EDI_WTR'nin ise $\alpha = .05$ düzeyinde hiçbir koşulda I. Tip hatayı kontrol edemezken $\alpha = .01$ düzeyinde tüm koşullarda I. Tip hatayı kontrol ettiğini belirtmiştir. 160 maddelik veri setinde diğer örneklemlere benzer şekilde, I. Tip hata oranlarının tüm koşullarda referans aralığın dışında olduğu görülmektedir.

Aynı veri setlerine süreklilik düzeltmesi yapılarak hesaplanan I. Tip hata oranları Tablo 23'te yer almaktadır.

Tablo 23

Madde Sayıları, Manidarlık Düzeyleri ve Yetenek Düzeylerine Göre Süreklilik Düzeltmesi İle SBI'nin I. Tip Hataları (n= 99990)

50 Madde								160 Madde							
Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
1	<u>0.011</u>	<u>0.006</u>	<i>0.0051</i>	0.0033	0.00290	0.00212	0.001481	1	<u>0.032</u>	<i>0.014</i>	0.0105	0.0059	0.00497	0.00310	0.001807
2	<u>0.020</u>	<u>0.009</u>	<i>0.0073</i>	0.0047	0.00411	0.00297	0.001831	2	<i>0.059</i>	0.023	0.0169	0.0084	0.00621	0.00338	0.001644
3	<u>0.017</u>	<u>0.006</u>	<u>0.0037</u>	0.0017	0.00131	0.00072	0.000459	3	<i>0.065</i>	0.019	0.0120	0.0051	0.00365	0.00191	0.000783
4	<u>0.007</u>	<u>0.002</u>	<u>0.0011</u>	<i>0.0005</i>	<i>0.00030</i>	<i>0.00011</i>	<i>0.000009</i>	4	<i>0.052</i>	<u>0.010</u>	<i>0.0063</i>	0.0031	0.00242	0.00134	0.000470
5	<u>0.001</u>	<u>0.000</u>	<u>0.0000</u>	<u>0.0000</u>	<u>0.00000</u>	<u>0.00000</u>	<u>0.000000</u>	5	<u>0.028</u>	<u>0.005</u>	<u>0.0036</u>	0.0020	0.00147	0.00058	0.000126
Toplam	<u>0.011</u>	<u>0.005</u>	<u>0.0034</u>	0.0020	0.00171	0.00117	0.000749	Toplam	<u>0.047</u>	<i>0.014</i>	0.0098	0.0049	0.00373	0.00205	0.000960

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 23'e göre 50 maddelik veri setinde süreklilik düzeltilmesi yapılarak hesaplanan I. Tip hata oranlarına ilişkin bulgular genel olarak Tablo 19'de yer alan 50000 test alan üzerinden elde edilen bulgulara büyük ölçüde benzerlik göstermektedir.

Tablo 23'e göre süreklilik düzeltilmesinin, birçok koşulda I. Tip hata oranının iyi kontrol edilmesini sağladığı görülmektedir. 50 maddelik veri setinde yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden süreklilik düzeltilmesi yapılarak hesaplanan I. Tip hata oranları, en büyük üç alfa seviyesinde nominal düzeyi aşmayarak iyi kontrol edilmektedir. Alfa seviyesi azaldıkça nominal düzey aşılmakta ve I. Tip hata kontrol edilememektedir. Wollack ve diğerleri (2015) tüm test alanlar üzerinden süreklilik düzeltilmesi ile hesaplanan I. Tip hata oranlarının tüm alfa seviyelerinde kontrol edildiğini; Sinharay ve diğerleri (2017) de Wollack ve diğerlerine (2015) benzer olarak $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ seviyelerinde hatanın iyi kontrol edildiğini belirtmiştir. Qin (2016) SBİ'nin süreklilik düzeltilmesi ile I. Tip hata oranlarını raporlaştırmamış olup SBİ'nin birden fazla yanlıştan doğruya silmesi olanlar için kullanılan versiyonu olan EDI_WTR'nin süreklilik düzeltilmesi ile I. Tip hatasını raporlaştırmıştır. Qin, EDI_WTR'nin hiçbir koşulda ve hiçbir alfa seviyesinde I. Tip hatayı kontrol edemediğini belirtmiştir ancak EDI_WTR, SBİ'nin özel bir kullanımı olduğunun unutulmaması gerekmektedir.

I. Tip hata oranlarının yetenek düzeylerine göre değişimi incelendiğinde $\alpha = .05$ için tüm yetenek düzeylerinde I. Tip hatanın iyi kontrol edildiği görülmektedir. Orta yetenek düzeyi ve altında en büyük alfa seviyelerinde I. Tip hata oranları nominal düzeyi aşmazken alfa seviyesi azaldıkça I. Tip hatanın iyi kontrol edilemediği görülmektedir. 4. yetenek düzeyinde $\alpha = .0001$ seviyesi dışında; 5. yetenek düzeyinde ise tüm alfa seviyelerinde I. Tip hata oranlarının iyi kontrol edildiği görülmektedir. I. Tip hatanın düşük yetenek düzeylerinde düşük alfa seviyelerinde iyi kontrol edilememesi problemleri bir durum olmakla birlikte süreklilik düzeltilmesi kontrol edilen I. Tip hata oranının artması önemlidir. Wollack ve diğerlerinin (2015) yetenek düzeylerine göre I. Tip hata kontrolü incelendiğinde, 1. yetenek düzeyinde $\alpha = .05$ seviyesi dışında tüm yetenek düzeyleri ve alfa seviyelerinde hata oranlarının iyi kontrol edildiği görülmektedir. Sinharay ve Johnson'ın (2016) I. Tip hata oranlarını dizi sonu ve tesadüfi silmelere göre ayrı ayrı raporlaştırdığı daha önce ifade edilmiştir. Dizi sonu silmeleri için, $\alpha = .05$, $\alpha = .01$, $\alpha = .001$ düzeylerinde tüm yetenek düzeylerinde, $\alpha = .0001$ düzeyinde ise 3. yetenek düzeyi dışında tüm yetenek düzeylerinde I. Tip hatanın kontrol edildiği görülmektedir. Tesadüfi silmeler için ise tüm yetenek düzeylerinde ve tüm alfa seviyelerinde I. Tip hatanın kontrol

edildiği görülmektedir. Sinharay ve Johnson (2016) tarafından süreklilik düzeltilmesi ile iyicil silme türlerine göre ayrılarak hesaplanan I. Tip hata oranlarının birbirine oldukça benzer olduğu, yalnızca $\alpha = .0001$ seviyesinde farklılaştığı göz önünde bulundurulduğunda, araştırma bulgularının Sinharay ve Johnson'ın bulgularıyla karşılaştırılmasının uygun olduğu düşünülmektedir. Buna göre Tablo 23, Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) bulgularıyla karşılaştırıldığında $\alpha = .05$ ve $\alpha = .01$ seviyelerinde bulgularının uyumlu olduğu ancak I. Tip hata oranlarının ilgili araştırmalarda daha iyi kontrol edildiği görülmektedir.

160 Maddelik Veri Seti

160 maddelik veri seti için süreklilik düzeltilmesi ile hesaplanan I. Tip hata oranları süreklilik düzeltilmesi yapılmadan hesaplanan hata oranlarının yer aldığı ve hiçbir koşulda hata oranlarının kontrol edilemediğinin görüldüğü Tablo 22 ile karşılaştırıldığında bazı koşullar için hata oranlarının kontrol edilebildiği görülmektedir. Ayrıca Tablo 23 bulguları, 50000 test alanına ilişkin bulguların yer aldığı Tablo 19 bulgularına tamamen benzerlik göstermektedir. Yetenek düzeyi ayrımı yapılmaksızın tüm test alanları üzerinden yalnızca $\alpha = .05$ seviyesinde I. Tip hata oranlarının kontrol edildiği görülmektedir.

I. Tip hata oranlarının yetenek düzeylerine göre değişimi incelendiğinde, 1. yetenek düzeyinin en büyük alfa seviyesinde hatanın iyi kontrol edildiği görülmektedir. 4. yetenek düzeyinde yalnızca $\alpha = .01$ seviyesinde; 5. yetenek düzeyinde ise en büyük üç alfa seviyesinde I. Tip hata oranlarının nominal düzeyi aşmadığı görülmektedir. 2 ve 3. yetenek düzeyinde ise hiçbir koşulda I. Tip hata oranları kontrol edilememiştir. Sinharay ve diğerlerinin (2017) 100 madde üzerinden yürüttükleri çalışma da I. Tip hata oranlarının $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ düzeylerinde kontrol edildiği görülmektedir. 500 örneklem büyüklüğünden 5000 örneklem büyüklüğüne geçildiğinde araştırma bulgularının Sinharay ve diğerlerine kısmen de olsa benzerlik göstermeye başladığı ifade edilmiştir ancak Sinharay ve diğerleri ile neredeyse aynı olan 99990 örneklem büyüklüğünden araştırma bulgularının benzerliğinde bir artış görülmemiştir. Süreklilik düzeltilmesi uygulandığı halde genel olarak 160 maddelik veri seti için I. Tip hata oranlarının çoğunlukla iyi kontrol edilmediği görülmektedir. Tablo 23 bulguları, 50000 test alanı için bulguların yer aldığı Tablo 19 ile karşılaştırıldığında genel olarak hata oranlarında önemli bir değişiklik olmadığı görülmektedir. 160 maddelik veri seti için de I. Tip hatanın düşük yetenek düzeylerinde iyi kontrol edilememesi Wollack ve diğerlerinin (2015) de belirttiği gibi istenmeyen bir durumdur. Ayrıca süreklilik düzeltilmesinin madde sayısının az olduğu durumlarda I. Tip hatanın şişme eğiliminde olmasından dolayı kullanımının önerildiği

Wollack ve diğerlerinin (2015) göz önünde bulundurulduğunda, 160 madde sayısının oldukça yüksek olmasına rağmen I. Tip hatanın 50000 test alanda olduğu gibi hala yeterince kontrol edilememesi dikkat çekicidir.

Süreklilik düzeltilmesi yapılarak hesaplanan I. Tip hata oranlarında daha küçük örneklerde görülene benzer şekilde, 50 maddelik veri setinde 160 maddelik veri setine göre hata oranları daha iyi kontrol edilmektedir. Süreklilik düzeltilmesi, beklendiği gibi, tüm koşullarda I. Tip hata oranının düşmesine neden olmaktadır. Bu bakımdan bulgular 500, 5000 ve 50000 test alan bulguları ile benzerlik göstermektedir.

I. Tip hatanın değerlendirilmesi için Bradley'in (1978) esnek referansı kullanıldığında 50 maddelik veri seti için $\alpha = .05$ seviyesindeki tüm yetenek düzeylerinde, alfa seviyesi kriterine göre kabul edilebilir olan I. Tip hata oranlarının referans aralığının altında kalarak kontrol edilemediği görülmektedir. Qin (2016) tarafından Bradley'in (1978) esnek referansı kullanıldığında süreklilik düzeltilmesi uygulanan EDI_WTR'nin yalnızca $\alpha = .05$ düzeyinde I. Tip hatayı kontrol ettiği belirtilmiştir. Araştırma bulguları Qin (2016) ile benzerlik göstermemekle birlikte Qin'in SBI'nin özel hali olan EDI_WTR'yi kullandığının göz önünde bulundurulması gerekmektedir. 5. yetenek düzeyinde nominal düzeye göre kabul edilebilir olan hiçbir alfa seviyesinde I. Tip hata oranları kontrol edilememektedir. 4. yetenek düzeyinde $\alpha = .001$ ve daha düşük alfa seviyelerindeki hata oranlarının referans aralığında kaldığı ve I. Tip hata oranlarının kontrol edildiği görülmektedir. 50 maddelik veri seti için I. Tip hata oranlarının değerlendirilmesinde esnek referans aralığının kullanılması 50000 test alana ilişkin Tablo 19 bulgularına benzer olarak, kontrol altına alınan hata oranı sayısını azaltacak yönde etki göstermektedir. Ancak düşük yetenek düzeylerinde $\alpha = .005$ seviyelerinde I. Tip hatanın kontrol edilmesini sağlaması önemlidir. 160 maddelik veri seti için esnek referans aralığı kullanıldığında ise genel olarak referans aralık içerisinde kalan hata oranlarının sayısı artmaktadır. Yetenek düzeyi ayrımı yapılmaksızın I. Tip hata oranları incelendiğinde $\alpha = .05$ seviyesine ek olarak $\alpha = .01$ seviyesinde de hata oranlarının iyi kontrol edildiği görülmektedir. Esnek referans aralığı dikkate alındığında $\alpha = .05$ seviyesinde 2, 3 ve 4. yetenek düzeylerinde I. Tip hata oranlarının manidarlık düzeyini aşmadığı ve böylece tüm yetenek düzeylerinde $\alpha = .05$ seviyesindeki I. Tip hata oranlarının kontrol edildiği görülmektedir. Ayrıca en düşük yetenek düzeyi için $\alpha = .01$ seviyesinde ve 4. yetenek düzeyi için $\alpha = .005$ seviyesinde de hata oranları kontrol edilmektedir.

99990 test alan için SBI'nin süreklilik düzeltilmesi yapılmaksızın güç oranları Tablo 24'te yer almaktadır.

Tablo 24

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 99990)

50 Madde									160 Madde									
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	
%10 (5 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%10 (16 Silme)	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	1.000	1.000	1.000	1.000	0.999	0.990		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	0.998	0.977	0.956	0.864	0.689		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>0.998</u>	<u>0.899</u>	0.800	0.509	0.401	0.208	0.073		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995
	5	<u>0.850</u>	<u>0.312</u>	<u>0.174</u>	<u>0.034</u>	<u>0.016</u>	0.002	0.000		5	1.000	1.000	1.000	0.999	0.996	0.934	0.605	
	Toplam	<u>0.970</u>	<u>0.842</u>	<u>0.794</u>	<u>0.704</u>	<u>0.675</u>	<u>0.615</u>	<u>0.551</u>		Toplam	1.000	1.000	1.000	1.000	0.999	0.987	0.920	
%20 (10 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%20 (32 Silme)	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	3	1.000	1.000	1.000	1.000	1.000	0.999	0.984		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	4	<u>1.000</u>	<u>0.998</u>	0.988	0.894	0.820	0.610	0.330		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.990</u>	<u>0.706</u>	<u>0.502</u>	<u>0.168</u>	<u>0.090</u>	0.014	0.001		5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997
	Toplam	<u>0.998</u>	<u>0.941</u>	<u>0.898</u>	<u>0.812</u>	<u>0.782</u>	<u>0.725</u>	<u>0.663</u>		Toplam	1.000	1.000	1.000	1.000	1.000	1.000	0.999	
%30 (15 Silme)	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	%30 (48 Silme)	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	4	<u>1.000</u>	<u>1.000</u>	0.999	0.961	0.917	0.753	0.479		4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	<u>0.991</u>	<u>0.752</u>	<u>0.560</u>	<u>0.225</u>	<u>0.127</u>	0.021	0.000		5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	Toplam	<u>0.998</u>	<u>0.950</u>	<u>0.912</u>	<u>0.837</u>	<u>0.809</u>	<u>0.755</u>	<u>0.695</u>		Toplam	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
Değişken Silme	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	Değişken Silme	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	1.000	1.000	1.000	1.000	0.997	0.990		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	3	0.964	0.895	0.860	0.750	0.708	0.595	0.456		3	0.976	0.914	0.881	0.798	0.758	0.675	0.548	
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000		4	—	—	—	—	—	—	—	
	5	—	—	—	—	—	—	—		5	—	—	—	—	—	—	—	
	Toplam	<u>0.988</u>	<u>0.965</u>	<u>0.953</u>	<u>0.917</u>	<u>0.903</u>	<u>0.864</u>	<u>0.815</u>		Toplam	0.992	0.971	0.960	0.933	0.919	0.892	0.849	

(Devam ediyor)

Tablo 24 (Devam)

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi Yapılmaksızın SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 99990)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
Toplam	1	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	Toplam	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	<u>1.000</u>	1.000	1.000	1.000	1.000	0.999	0.995		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	0.991	0.974	0.964	0.932	0.916	0.865	0.782		3	0.994	0.979	0.970	0.949	0.939	0.919	0.887
	4	<u>0.999</u>	<u>0.966</u>	0.929	0.788	0.713	0.524	0.294		4	1.000	1.000	1.000	1.000	1.000	1.000	0.998
	5	<u>0.944</u>	<u>0.590</u>	<u>0.412</u>	<u>0.142</u>	<u>0.078</u>	0.012	0.000		5	1.000	1.000	1.000	1.000	0.999	0.978	0.867
	Toplam	<u>0.989</u>	0.920	0.882	0.806	0.780	0.726	0.666		Toplam	0.999	0.995	0.993	0.989	0.986	0.978	0.952

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 24 bulguları Tablo 20’de yer alan 50000 test alan üzerinden elde edilen bulgulara tamamen benzerlik göstermekle birlikte istatistiksel gücün az miktarda daha yüksek olduğu görülmektedir. Yetenek düzeylerine göre ayrılmadan toplam test alanlar üzerinden I. Tip hata kontrol edilemediği için istatistiksel güç değerlendirilememektedir. İstatistiksel güç tahrifat miktarına göre ayrılmaksızın toplam test alanlar üzerinden yetenek düzeylerine göre incelendiğinde, $\alpha = .05$ seviyesinde tahrifatlı bireylerin 1. yetenek düzeyinde mükemmel düzeyde, 5. yetenek düzeyinde ise yüksek düzeyde doğru belirlendiği görülmektedir. 5. yetenek düzeyi için diğer alfa seviyelerinde istatistiksel güç kabul edilebilir düzeyde değildir.

İstatistiksel güç tahrifat miktarına göre incelendiğinde, 5. yetenek düzeyi için tahrifat oranı arttıkça aynı yetenek düzeyinde istatistiksel gücün arttığı görülmektedir. 1. yetenek düzeyinde en düşük tahrifat oranında dahi istatistiksel güç mükemmel olduğu için tahrifat oranı arttıkça gücün değişimi gözlenememektedir. %10 oranında tahrifat (5 silme) yapıldığı durumda $\alpha = .05$ düzeyinde 1. yetenek düzeyinde tahrifatlı bireyler mükemmel düzeyde, 5. yetenek düzeyinde ise tatmin edici düzeyde belirlenmektedir. 5. yetenek düzeyi için $\alpha = .01$ seviyesinde gücün keskin bir şekilde düşmesiyle birlikte diğer alfa seviyelerinde de güç azalmaya devam etmekte ve en büyük alfa seviyesi dışında kabul edilebilir istatistiksel güç elde edilememektedir. %20 oranında tahrifat (10 silme) yapıldığı durumda 1. yetenek düzeyinde $\alpha = .05$ düzeyinde istatistiksel gücün mükemmel olduğu; 5. yetenek düzeyinde ise güç $\alpha = .05$ seviyesinde mükemmel yakın ve $\alpha = .01$ seviyesinde kabul edilebilir düzeydeyken diğer alfa seviyelerinde istatistiksel gücün kabul edilebilir düzeyde olmadığı görülmektedir. %30 oranında tahrifat (15 silme) yapıldığı durumda %00 oranında tahrifat (10 silme) yapıldığı duruma benzer olarak, 1. yetenek düzeyinde $\alpha = .05$ seviyesinde tahrifatlı test alanların mükemmel şekilde belirlendiği; 5. yetenek düzeyinde ise güç $\alpha = .05$ seviyesinde mükemmel yakın ve $\alpha = .01$ seviyesinde kabul edilebilir düzeydeyken diğer alfa seviyelerinde istatistiksel gücün kabul edilebilir düzeyde olmadığı görülmektedir. Değişken sayıda silme yapıldığı durumda da 1. yetenek düzeyinde $\alpha = .05$ seviyesinde tahrifatlı bireyler mükemmel şekilde belirlenmektedir. En yüksek yetenek düzeyinde ise tahrifata uğrayan test alan olmadığı için istatistiksel güç hesaplanmamıştır.

Tablo 24 bulguları, 5000 test alan için bulguların yer aldığı Tablo 16 ile karşılaştırıldığında kabul edilebilir istatistiksel güç değerlerinin sayısının daha az olduğu; 50000 test alan için bulguların yer aldığı Tablo 20 ile karşılaştırıldığında ise kabul edilebilir istatistiksel güç değerlerinin sayısının daha fazla olduğu görülmektedir. Tablo 24 genel olarak

incelendiğinde benzer arařtırmalarda da olduđu gibi, tahrifat oranı arttıkça istatistiksel g¼c¼n arttıđı ancak alfa seviyesi azaldıkça istatistiksel g¼c¼n d¼řt¼đ¼ g¼r¼lmektedir (Qin, 2016; Sinharay vd., 2017; Wollack vd., 2015). Wollack ve diđerlerinin (2016) de belirttiđi üzere yetenek d¼zeyi arttıkça istatistiksel g¼c¼n bir miktar d¼řt¼đ¼ g¼r¼lmektedir.

160 Maddelik Veri Seti

160 maddelik veri seti iin t¼m yetenek d¼zeylerinde ve alfa seviyelerinde I. Tip hata kontrol edilemediđi iin istatistiksel g¼c¼ dikkate alınmamaktadır.

Tablo 24'te yer alan 50 maddelik veri seti iin istatistiksel g¼c¼ deđerleri Bradley'in esnek referans aralıđına g¼re incelendiđinde ise yetenek d¼zeylerine g¼re ayırım yapılmadıđı durumda en b¼y¼k alfa seviyesinde istatistiksel g¼c¼n t¼m silme kořullarında ve toplam test alanlarda ok y¼ksek olduđu g¼r¼lmektedir. 5. yetenek d¼zeyinde en b¼y¼k alfa seviyelerindeki istatistiksel g¼c¼ deđerlerinin I. Tip hata oranlarının referans aralıđın altında kalması nedeniyle deđerlendirme dıřında kaldıđı ancak $\alpha = .0005$ seviyesinde istatistiksel g¼c¼n deđerlendirilebildiđi ancak onun da kabul edilebilir d¼zeyin olduka altında olduđu g¼r¼lmektedir. 4. yetenek d¼zeyi iin en b¼y¼k iki alfa seviyesinde hata oranlarının referans aralıđında kalması nedeniyle deđerlendirilebilir olduđu ve bu d¼zeyde istatistiksel g¼c¼n neredeyse t¼m silme kořullarında m¼kemmele yakın olduđu g¼r¼lmektedir. Ayrıca 2. yetenek d¼zeyinde esnek referans aralıđı ile $\alpha = .05$ seviyesinde istatistiksel g¼c¼ deđerlendirilebilir haldedir ve burada istatistiksel g¼c¼ t¼m kořullarda m¼kemmeldir. Esnek referans aralıđının kullanılması ile en ¼st yetenek d¼zeyinde birok alfa seviyesinde istatistiksel g¼c¼n deđerlendirilememesi problemleri bir durum gibi g¼z¼kmekle birlikte tahrifatın daha ok d¼ř¼k yetenek d¼zeylerinde yapılması beklendiđinden, 2. yetenek d¼zeyinde istatistiksel g¼c¼n deđerlendirilebilir hale gelmesi ¼nemli g¼r¼lmektedir. 160 maddelik veri seti iin I. Tip hata oranları Bradley'in esnek referans aralıđına g¼re deđerlendirildiđinde de kontrol edilemediđinden istatistiksel g¼c¼ hibir kořulda deđerlendirilememektedir.

99990 test alan iin s¼reklilik d¼zeltmesi yapılarak uygulanan SBİ'nin istatistiksel g¼c¼ Tablo 25'te yer almaktadır.

Tablo 25

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 99990)

50 Madde									160 Madde									
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	
%10 (5 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%10 (16 Silme)	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	0.999	0.998	0.990	0.957		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>1.000</u>	<u>0.993</u>	<u>0.979</u>	0.895	0.841	0.690	0.474		3	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>0.957</u>	<u>0.685</u>	<u>0.530</u>	<u>0.257</u>	<u>0.183</u>	<u>0.076</u>	<u>0.022</u>		4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.979
	5	<u>0.473</u>	<u>0.090</u>	<u>0.039</u>	<u>0.005</u>	<u>0.002</u>	<u>0.000</u>	<u>0.000</u>		5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	0.993	0.973	0.809	0.406	
	Toplam	<u>0.886</u>	<u>0.754</u>	<u>0.709</u>	0.631	0.605	0.551	0.491		Toplam	<u>1.000</u>	<u>1.000</u>	1.000	0.999	0.995	0.962	0.877	
%20 (10 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%20 (32 Silme)	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	0.994	0.949		3	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>1.000</u>	<u>0.980</u>	<u>0.939</u>	<u>0.757</u>	<u>0.656</u>	<u>0.429</u>	<u>0.197</u>		4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	5	<u>0.894</u>	<u>0.418</u>	<u>0.251</u>	<u>0.053</u>	<u>0.022</u>	<u>0.002</u>	<u>0.000</u>		5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.990
	Toplam	<u>0.979</u>	<u>0.880</u>	<u>0.838</u>	0.762	0.736	0.685	0.629		Toplam	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	0.998
%30 (15 Silme)	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	%30 (48 Silme)	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	0.994		3	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	<u>1.000</u>	<u>0.997</u>	<u>0.987</u>	<u>0.891</u>	<u>0.816</u>	<u>0.609</u>	<u>0.359</u>		4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	5	<u>0.922</u>	<u>0.508</u>	<u>0.342</u>	<u>0.091</u>	<u>0.041</u>	<u>0.004</u>	<u>0.000</u>		5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	0.995
	Toplam	<u>0.984</u>	<u>0.901</u>	<u>0.866</u>	0.797	0.772	0.723	0.671		Toplam	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	0.999
Değişken Silme	1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	Değişken Silme	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	
	2	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	0.998	0.996	0.989	0.969		2	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	3	<u>0.850</u>	<u>0.717</u>	<u>0.658</u>	0.528	0.480	0.378	0.262		3	<u>0.887</u>	<u>0.793</u>	<u>0.749</u>	<u>0.652</u>	<u>0.612</u>	<u>0.521</u>	<u>0.403</u>	
	4	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>		4	—	—	—	—	—	—	—	
	5	—	—	—	—	—	—	—		5	—	—	—	—	—	—	—	
	Toplam	<u>0.950</u>	<u>0.906</u>	<u>0.886</u>	0.842	0.825	0.789	0.743		Toplam	<u>0.962</u>	<u>0.931</u>	0.916	0.884	0.871	0.840	0.801	

(Devam ediyor)

Tablo 25 (Devam)

Madde Sayısı ve Tahrifat Miktarı Koşullarının Farklı Manidarlık Düzeylerinde Süreklilik Düzeltmesi İle SBI'nin Yetenek Düzeylerine Göre İstatistiksel Güç Oranları (n= 99990)

		50 Madde							160 Madde								
Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$	Tahrifat Miktarı	Yetenek Düzeyi	$\alpha=.05$	$\alpha=.01$	$\alpha=.005$	$\alpha=.001$	$\alpha=.0005$	$\alpha=.0001$	$\alpha=.00001$
Toplam	1	<u>1.000</u>	<u>1.000</u>	<i>1.000</i>	1.000	1.000	1.000	1.000	Toplam	1	<u>1.000</u>	<u>1.000</u>	1.000	1.000	1.000	1.000	1.000
	2	<u>1.000</u>	<u>1.000</u>	<i>1.000</i>	0.999	0.998	0.995	0.981		2	<i>1.000</i>	1.000	1.000	1.000	1.000	1.000	1.000
	3	<u>0.963</u>	<u>0.927</u>	<u>0.909</u>	0.856	0.830	0.765	0.670		3	<i>0.972</i>	0.948	0.937	0.913	0.903	0.880	0.851
	4	<u>0.986</u>	<u>0.887</u>	<u>0.818</u>	<u>0.635</u>	<u>0.552</u>	<u>0.371</u>	<u>0.193</u>		4	<i>1.000</i>	<u>1.000</u>	<i>1.000</i>	1.000	1.000	1.000	0.993
	5	<u>0.763</u>	<u>0.339</u>	<u>0.211</u>	<u>0.050</u>	<u>0.022</u>	<u>0.002</u>	<u>0.000</u>		5	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	0.998	0.991	0.936	0.797
	Toplam	<u>0.950</u>	<u>0.855</u>	<u>0.818</u>	0.749	0.724	0.676	0.621		Toplam	<u>0.994</u>	<u>0.988</u>	0.986	0.980	0.977	0.963	0.932

Not: I. Tip hata oranının nominal düzeyi aşmadığı koşullar altı çizili, Bradley'in (1978) esnek referans aralığına göre kabul edilebilir olan koşullar ise italik olarak belirtilmiştir.

50 Maddelik Veri Seti

Tablo 25 bulguları, Tablo 21’de yer alan 50000 test alan üzerinden elde edilen bulgularla genel olarak benzerlik göstermektedir. İstatistiksel güç tahrifat miktarına ve yetenek düzeylerine göre ayrılmaksızın toplam test alanlar üzerinden incelendiğinde, alfa seviyesi düştükçe istatistiksel güç azalmakla birlikte, en büyük üç alfa seviyesinde istatistiksel gücün yüksek olduğu görülmektedir. İstatistiksel güç tahrifat miktarına göre ayrılmaksızın toplam test alanlar üzerinden incelendiğinde, genel olarak yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir. 1. ve 2. yetenek düzeyinde $\alpha = .05$ ve $\alpha = .01$ seviyelerinde tahrifatlı bireyler mükemmel düzeyde belirlenmektedir ancak daha düşük alfa seviyelerinde hatalar iyi kontrol edilemediği için istatistiksel güç değerlendirilememiştir. 3. yetenek düzeyinde en büyük üç alfa seviyesinde istatistiksel gücün değerlendirilebildiği ve bu seviyelerde gücün oldukça yüksek olduğu görülmektedir. 4. yetenek düzeyinde $\alpha = .05$ seviyesinde güç çok yüksek olmakla birlikte, alfa seviyesi katılaştıkça istatistiksel güç düşmektedir ancak $\alpha = .005$ seviyesinde güç hala tatmin edici düzeydedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hata kontrol edilmektedir ancak yalnızca $\alpha = .05$ seviyesinde kabul edilebilir gücün elde edildiği görülmektedir.

İstatistiksel güç tahrifat miktarına göre incelendiğinde diğer örneklerde de görüldüğü ve Wollack ve diğerleri (2015) ile Sinharay ve Johnson’ın (2016) da belirttiği üzere tahrifat oranı arttıkça yetenek düzeylerinde istatistiksel gücün arttığı görülmektedir. %10 oranında tahrifat (5 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün en büyük üç alfa seviyesinde kabul edilebilir düzeyde olduğu belirlenmiştir. Sinharay ve diğerlerinin (2017) çalışmalarında ise 4 silme yaparak hesapladıkları istatistiksel gücün $\alpha = .05$, $\alpha = .01$ ve $\alpha = .001$ düzeylerinin hiçbirinde kabul edilebilir olmadığı görülmektedir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde orta yetenek düzeyi ve altında, hatanın kontrol edildiği koşullarda en büyük alfa seviyelerinin neredeyse tamamında tahrifatlı bireylerin mükemmel düzeyde belirlendiği görülmektedir. 4. yetenek düzeyinde $\alpha = .05$ seviyesinde istatistiksel gücün oldukça yüksek olduğu ancak alfa seviyesi küçüldükçe istatistiksel gücün kabul edilemez düzeyde düştüğü görülmektedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edilmesine rağmen hiçbir alfa seviyesinde kabul edilebilir istatistiksel güç elde edilemediği görülmektedir. Wollack ve diğerleri (2015) tarafından 5 silme yapıldığı durumda yalnızca 1. yetenek düzeyinde ve $\alpha = .05$ düzeyinde çok yüksek istatistiksel güç; 1. yetenek düzeyi için $\alpha = .01$ ve 2. yetenek düzeyi için $\alpha = .05$ düzeylerinde kabul edilebilir istatistiksel güç rapor edilmiştir. Araştırma

bulguları, $\alpha = .05$ ve $\alpha = .01$ seviyelerinde en üst yetenek düzeyine kadar yüksek istatistiksel güç elde edilmesi ile Wollack ve diğerlerinden farklılaşmaktadır. Ancak en üst yetenek düzeyinde Wollack ve diğerlerinden daha yüksek istatistiksel güç elde edilmesine rağmen Wollack ve diğerlerine benzer şekilde güç oranları kabul edilebilir değildir. Sinharay ve Johnson'ın (2016) ise 5 silme yapıldığı durumda hiçbir yetenek düzeyinde ve alfa seviyesinde kabul edilebilir istatistiksel güç elde edilemediği görülmektedir.

%20 oranında tahrifat (10 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün en büyük üç alfa seviyesinde de tatmin edici olduğu görülmektedir. Sinharay ve diğerlerinin (2017) ise 8 silme için hesapladıkları istatistiksel gücün yalnızca $\alpha = .05$ düzeyinde kabul edilebilir olduğu göz önünde bulundurulduğunda yürütülen araştırmada istatistiksel gücün daha iyi kontrol edildiği görülmektedir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde 4. yetenek düzeyine kadar hata oranlarının kontrol edildiği büyük alfa seviyelerinde tahrifatlı bireylerin mükemmel düzeyde belirlendiği görülmektedir. 4. yetenek düzeyinde büyük alfa seviyelerinde tahrifatlı bireylerin çok büyük oranda doğru belirlendiği, alfa seviyesi azaldıkça gücün düştüğü ancak $\alpha = .001$ seviyesinde dahi istatistiksel gücün kabul edilebilir düzeyde olduğu görülmektedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hata oranlarının nominal düzeyin altında kalmasına rağmen yalnızca $\alpha = .05$ seviyesinde istatistiksel gücün tatmin edici olduğu ve diğer alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilemediği görülmektedir. Sinharay ve Johnson'ın (2016) 10 silme yapıldığı durumda kabul edilebilir istatistiksel güç değerlerini $\alpha = .05$ düzeyinde 1. ve 2. yetenek düzeyinde; $\alpha = .01$ düzeyinde ise yalnızca 1. yetenek düzeyinde elde ettiği görülmektedir. Araştırma bulguları, Sinharay ve Johnson'ın düşük yetenek düzeylerinin daha düşük alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilmemesi bulgularıyla benzerlik göstermektedir. Wollack ve diğerlerinin (2015) çalışmasında ise 10 silme yapıldığı durumda en büyük alfa seviyesinde 3. yetenek düzeyine kadar istatistiksel güç mükemmel yakınen, 3. yetenek düzeyinde güç çok yüksek ve 4. yetenek düzeyinde ise kabul edilebilir düzeydedir. 1. yetenek düzeyinde neredeyse en düşük alfa seviyelerine kadar istatistiksel gücün oldukça yüksek olduğu; 2. yetenek düzeyinde istatistiksel gücün büyük alfa seviyelerinde kabul edilebilir olduğu görülmektedir. Araştırma bulguları, en üst yetenek düzeyinin en büyük alfa seviyesinde yüksek istatistiksel güç elde edilmesi ve düşük yetenek düzeylerinin düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden (2015) farklılaşmaktadır.

%30 oranında tahrifat (15 silme) yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün $\alpha = .05$, $\alpha = .01$ ve $\alpha =$

.005 seviyelerinde tatmin edici olduğu görülmektedir. İstatistiksel güç yetenek düzeylerine göre incelendiğinde 1, 2 ve 3. yetenek düzeylerinde I. Tip hatanın iyi kontrol edildiği en büyük alfa seviyelerinde tahrifatlı test alanların mükemmel düzeyde belirlendiği görülmektedir. 4. yetenek düzeyinde en büyük alfa seviyesinde istatistiksel güç mükemmel iken alfa seviyesi azaldıkça güç düşmekte ancak $\alpha = .0005$ seviyesinde bile gücün tatmin edici olduğu görülmektedir. 5. yetenek düzeyinde tüm alfa seviyelerinde I. Tip hatanın iyi kontrol edilmesine rağmen, 10 silme yapıldığı durumdakine benzer olarak, yalnızca en büyük alfa seviyesinde istatistiksel gücün çok yüksek olduğu ve diğer alfa seviyelerinde kabul edilebilir istatistiksel güç elde edilemediği görülmektedir. Wollack ve diğerlerinin (2015) çalışmasında 15 silme yapıldığı durumda en büyük alfa seviyesinde en düşük yetenek düzeylerinde istatistiksel gücün neredeyse mükemmel olduğu, 3. yetenek düzeyinde ise çok yüksek olduğu görülmektedir. 1. yetenek düzeyinde en düşük alfa seviyesinde bile istatistiksel güç tatmin edicidir. 2. yetenek düzeyinde istatistiksel gücün $\alpha = .001$ seviyesinde dahi kabul edilebilir olduğu görülmektedir. Araştırma bulguları, en üst yetenek düzeyinin en büyük alfa seviyesinde yüksek istatistiksel güç elde edilmesi ve düşük yetenek düzeylerinin düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden (2015) farklılaşmaktadır.

Değişken sayıda silme yapıldığı durumda, yetenek düzeylerine göre ayırım yapılmaksızın tüm test alanlar üzerinden incelendiğinde istatistiksel gücün en büyük üç alfa seviyesinde tatmin edici olduğu görülmektedir. İstatistiksel güç düşük yetenek düzeylerinde hatanın kontrol edildiği büyük alfa seviyelerinde mükemmeldir. Orta yetenek düzeyinde ise güç $\alpha = .05$ seviyesinde tatmin edici iken $\alpha = .01$ seviyesinde kabul edilebilir düzeydedir. 500 ve 50000 test alanda değişken silmeler yalnızca ilk üç yetenek düzeyinde yapılırken, 5000 ve 99990 test alanda değişken silmelerin 4. yetenek düzeyinde de yapıldığı görülmektedir. Ancak 4. yetenek düzeyinde değişken tahrifata uğrayan birey sayısının çok az olduğu belirlenmiştir. 5000 test alanın bulgularının yer aldığı Tablo 17'ye benzer şekilde, 4. yetenek düzeyi için istatistiksel güç mükemmel düzeydedir. Genellikle istatistiksel gücün yetenek düzeyi arttıkça düştüğü göz önüne alındığında bu durumun 4. yetenek düzeyinde tahrifatlı birey sayısının az olmasından kaynaklanabileceği düşünülmektedir. Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) çalışmalarında da görüldüğü gibi en yüksek yetenek düzeyinde tahrifata uğrayan test alan olmadığı için istatistiksel güç hesaplanmamıştır. Wollack ve diğerlerinin (2015) çalışmasında değişken sayıda silme yapıldığı durumda, en düşük alfa seviyesi dışında 1. yetenek düzeyinde tüm alfa seviyelerinde istatistiksel gücün kabul edilebilir olduğu ancak 2. ve 3. yetenek düzeylerinde kabul edilebilir istatistiksel gücün olmadığı görülmektedir. Araştırma bulguları, diğer yetenek düzeylerinde de yüksek istatistiksel güç elde edilmesi ve 1.

yetenek düzeyinde düşük alfa seviyelerinde istatistiksel güç elde edilememesiyle Wollack ve diğerlerinden (2015) farklılaşmaktadır. Sinharay ve Johnson (2016) tarafından ise değişken silme yapıldığı durumda kabul edilebilir istatistiksel gücün yalnızca $\alpha = .05$ düzeyinde 1. yetenek düzeyinde elde edildiği raporlaştırılmıştır. Genel olarak Sinharay ve Johnson'ın (2016) tüm koşullar için istatistiksel güç değerleri oldukça düşüktür. Sinharay ve Johnson'ın (2016) tahrifatlı birey yüzdesinin (%50), araştırmadaki tahrifatlı birey yüzdesinden (yaklaşık %8) oldukça yüksek olmasının istatistiksel gücü etkileyebileceği düşünülmektedir. Qin (2016) tarafından EDI_WTR kullanılarak yürütülen araştırmanın bazı koşullarında tahrifat yüzdesinin artmasının istatistiksel gücü bir miktar düşürdüğü görülmektedir.

Tüm tahrifat miktarı koşulları için yetenek düzeyi arttıkça kontrol edilen I. Tip hata oranı sayısının arttığı ancak istatistiksel gücün düştüğü görülmektedir. Yüksek yetenek düzeylerinde tahrifat oranı arttıkça istatistiksel gücün arttığı görülmektedir. Düşük yetenek düzeylerinde en düşük tahrifat oranı koşulunda bile güç mükemmel olduğu için bir değişim gözlenmemiştir. Wollack ve diğerleri (2015) ile Sinharay ve Johnson'ın (2016) da belirttiği üzere tahrifat oranı arttıkça istatistiksel gücün arttığı ancak alfa seviyesi azaldıkça ve aynı tahrifat oranı için yetenek düzeyi arttıkça istatistiksel gücün düştüğü görülmektedir. Wollack ve diğerleri (2015) yüksek yetenek düzeylerinde istatistiksel gücün düşük çıkmasını SBİ'nin hesaplanmasındaki doğru yanıt olasılık ortalamasına bağlı olarak açıklamaktadır. Buna göre, düşük yetenek düzeyindeki test alanlar için, silinmiş maddeler kolay olmadıkça, test alanın beklenen doğru yanıt olasılık ortalamasına sahip olarak maddeyi doğru yanıtlaması beklenmemektedir. Bu durum ilgili maddeyi doğru yanıtlamış test alanlar için SBİ'nin manidar çıkararak test alanı tahrifatlı olarak işaretlemesine neden olmaktadır. Yüksek yetenek düzeyindeki test alanlarda ise, silinmiş maddelerin doğru yanıtlanma olasılığının, maddeler çok zor olmadıkça, doğru yanıt olasılık ortalamasından yüksek olması ve dolayısıyla maddenin doğru yanıtlanması beklenmektedir. Bu durum, madde tahrifat sonucunda doğru yanıtlanmış olsa bile SBİ'nin manidar çıkmamasına ve test alanın tahrifatlı olarak işaretlenmemesine neden olmaktadır. Sonuç olarak tahrifatlı test alanların yüksek yetenek düzeylerinde SBİ ile belirlenmesi zorlaşmaktadır.

Süreklilik düzeltmesi yapılması, benzer araştırmalarda olduğu gibi, I. Tip hata oranının nominal düzeyi aşmadığı koşul sayısını artırmakta ancak istatistiksel gücün düşmesine neden olmaktadır (Sinharay ve Johnson, 2016; Wollack vd, 2015). Süreklilik düzeltmesi yapılmadığı durumda I. Tip hatanın kontrol edildiği koşullar için süreklilik düzeltmesi yapılmasının istatistiksel gücü ne kadar düşürdüğü incelenmiştir. Buna göre tahrifat oranı dikkate alınmadığında istatistiksel gücün, 1. yetenek düzeyinde değişmediği, 5. yetenek düzeyinde ise

0.06 ile 0.25 arasında azaldığı görülmektedir. 500, 5000 ve 50000 test alandakine benzer şekilde silme koşullarının tümünde 1. yetenek düzeyi için süreklilik düzeltmesi yapılmasının istatistiksel gücü deęiřtirmedięi görülmektedir. 5. yetenek düzeyi için ise süreklilik düzeltmesinin yapılması istatistiksel gücü %10 oranında tahrifat (5 silme) koşulunda 0.01 ile 0.38 arasında; %20 oranında tahrifat (10 silme) koşulunda 0.07 ile 0.29 arasında; %30 oranında tahrifat (15 silme) koşulunda 0.07 ile 0.24 arasında azaltmıřtır. 5000 test alanın bulgularıyla karřılařtırıldığında, 99990 test alan için istatistiksel gücün deęiřim ranjının bir miktar daha fazla olduđu, 50000 test alanın bulguları ile karřılařtırıldığında ise deęiřim ranjının neredeyse aynı kaldıđı belirlenmiřtir. İstatistiksel gücün süreklilik düzeltmesine bađlı deęiřiminin tahrifat oranı arttıkça azaldıđı görülmektedir. SBİ ile tahrifatlı bireylerin belirlenmesinde süreklilik düzeltmesi uygulanmasının I. Tip hata kontrolünü artırarak kabul edilebilir istatistiksel güç koşulu da arttırdıđı ve tahrifata uğraması muhtemel olan düşük yetenek düzeylerinde istatistiksel gücü neredeyse hiç etkilemediđi belirlenmiřtir.

Tablo 25 bulguları 50000 test alana iliřkin bulguların yer aldıđı Tablo 21 ile karřılařtırıldığında, istatistiksel gücün tüm silme koşulları için önemli bir deęiřiklik göstermediđi belirlenmiřtir.

160 Maddelik Veri Seti

160 maddelik veri seti için istatistiksel güç tahrifat miktarına ve yetenek düzeylerine göre ayrılmaksızın toplam test alanlar üzerinden incelendiđinde, I. Tip hatanın kontrol edildiđi tek durum olan $\alpha = .05$ seviyesinde istatistiksel gücün mükemmel yakın olduđu görülmektedir. İstatistiksel güç tahrifat miktarına göre ayrılmaksızın toplam test alanlar üzerinden yetenek düzeylerine göre incelendiđinde, 1. yetenek düzeyinde $\alpha = .05$ seviyesinde, 4. yetenek düzeyinde $\alpha = .01$ seviyesinde, 5. yetenek düzeyinde ise en büyük üç alfa seviyesinde tahrifatlı bireylerin mükemmel şekilde belirlendiđi görülmektedir.

İstatistiksel güç tahrifat miktarına göre incelendiđinde ise tüm sabit silme koşullarında I. Tip hatanın kontrol edildiđi durumlarda güç mükemmeldir. Deęiřken silme koşulunda ise yetenek düzeyi ayrımı yapılmaksızın tahrifatlı bireylerin çok yüksek oranda dođru belirlendiđi görülmektedir. 1. yetenek düzeyinde ise I. Tip hatanın kontrol edilebildiđi tek durum olan $\alpha = .05$ düzeyinde istatistiksel gücün, diđer sabit silme koşullarına benzer şekilde mükemmel olduđu görülmektedir. 50 maddelik veri setinden farklı olarak 4. yetenek düzeyinde deęiřken sayıda silme yapılamamıřtır. Diđer örneklem büyüklüklerinde de görüldüđu gibi, 50 maddelik veri seti ile karřılařtırıldığında daha az sayıda istatistiksel güç yorumlanabilmektedir. 50000

test alan üzerinden istatistiksel güce ait bulguların yer aldığı Tablo 21 ile karşılaştırıldığında bulguların tamamen benzerlik gösterdiği görülmektedir.

Tablo 25'te yer alan bulgular Bradley'in (1978) esnek referans aralığına göre değerlendirildiğinde bulguların, Tablo 21'de yer alan 50000 test alan için bulgulara 50 maddelik veri setinde büyük ölçüde, 160 maddelik veri setinde ise tamamen benzerlik gösterdiği görülmektedir.

Tablo 25'te yer alan 50 maddelik veri seti için Bradley'in esnek referans aralığına göre tüm koşullar için $\alpha = .05$ seviyesinde tüm yetenek düzeylerinde, 5. yetenek düzeyinde tüm alfa seviyelerinde ve 4. yetenek düzeyinin büyük alfa seviyelerinde istatistiksel güç değerleri I. Tip hata oranlarının referans aralığın dışında kalması nedeniyle değerlendirilememiştir. Esnek referans aralığının kullanılması en düşük yetenek düzeylerinde $\alpha = .005$ seviyesinde istatistiksel gücün değerlendirilebilir hale gelmesini sağlamıştır ve bu düzeylerde güç tüm koşullar için mükemmeldir. 4. yetenek düzeyinde ise en düşük alfa seviyelerinde I. Tip hata oranları referans aralığın içerisinde kaldığı için değerlendirilebilir hale gelmiştir ancak bu düzeyde değişken silme koşulu dışında hiçbir koşulda güç kabul edilebilir düzeyde değildir. 4. yetenek düzeyi için değişken silme koşulunda ise $\alpha = .0001$ seviyesinde güç mükemmel iken $\alpha = .00001$ seviyesinde tahrifatlı hiçbir test alan doğru belirlenmemektedir.

160 maddelik veri seti için esnek referans aralığı kullanıldığında değerlendirilebilen istatistiksel güç sayısının 5. yetenek düzeyi hariç, tüm yetenek düzeylerinde ve toplam test alanlar düzeyinde arttığı görülmektedir. Yetenek düzeyi ayrımı yapılmaksızın tüm test alanlar üzerinden $\alpha = .05$ ve $\alpha = .01$ düzeylerinde tahrifatlı bireylerin mükemmel yakın düzeyde doğru belirlendiği görülmektedir. Tüm silme koşulları için $\alpha = .05$ seviyesinde tüm yetenek düzeylerinde istatistiksel gücün değerlendirilebildiği ve değişken silme koşulu dışında tüm koşullarda gücün mükemmel olduğu görülmektedir. Ayrıca 1. yetenek düzeyinde $\alpha = .01$ yetenek düzeyinde gücün değerlendirilebilir hale gelmesi ve bu düzeyde gücün mükemmel olması önemlidir.

BÖLÜM 4

SONUÇLAR VE ÖNERİLER

Bu bölümde araştırma sorularına yanıt aramak için yapılan analizlerle elde edilen bulgulardan yola çıkılarak ulaşılan sonuçlara ve elde edilen sonuçlar kapsamında yapılan önerilere yer verilmiştir.

Sonuçlar

Bu çalışma kapsamında test tahrifatının belirlenmesinde kullanılan MTK'ya dayalı SBİ'nin, simülatif veriler üzerinden farklılaştırılan koşullar (örneklem büyüklüğü, yetenek düzeyi, madde sayısı ve tahrifat oranları) altında I. Tip hata oranları ve istatistiksel gücü araştırılmıştır. Bu doğrultuda süreklilik düzeltmesi uygulanan ve uygulanmayan SBİ'nin ürettiği I. Tip hata oranları ve istatistiksel güç değerleri; 500, 5000, 50000 ve 99990 örneklem büyüklüklerinde beş farklı yetenek düzeyine göre, 50 ve 160 maddelik veri setlerinde, %10, %20 ve %30 oranlarında sabit tahrifat ve değişken tahrifat için .00001, .0001, .0005, .001, .005, .01 ve .05 α seviyelerinde incelenmiş ve aşağıdaki sonuçlara ulaşılmıştır:

1. SBİ'de süreklilik düzeltmesi uygulanmadığı durumlarda 50 maddelik veri setlerinde genel olarak büyük alfa seviyelerinde I. Tip hata oranı kontrolü sağlanmıştır. Tüm örneklem büyüklükleri için 160 maddelik veri setlerinde tahrifatlı olmadığı halde tahrifatlı olarak belirlenen test alanların oranının belirlenen alfa düzeylerini aşması, I. Tip hata oranlarının iyi kontrol edilemediğini göstermiştir.
2. SBİ'de süreklilik düzeltmesi uygulanması tüm alfa seviyelerinde, tahrifatlı olmadığı halde SBİ tarafından tahrifatlı olarak belirlenen test alan oranlarını düşürerek I. Tip hata oranlarının düşmesine neden olurken, tahrifatlı test alanların da süreklilik düzeltmesi uygulanmayan SBİ'ye göre daha düşük oranlarda doğru tespit edilerek istatistiksel gücün düşmesine neden olmuştur. SBİ'de süreklilik düzeltmesi uygulanması daha fazla alfa seviyesinde I. Tip hata oranı kontrolünü sağladığı için daha fazla koşulda kabul edilebilir istatistiksel güç elde edilmiştir.

3. 50 maddelik veri setinde I. Tip hata oranlarının her örneklem büyüklüğü için 160 maddelik veri setine göre daha iyi kontrol edildiği, yalancı pozitif oranların daha düşük olduğu görülmüştür.
4. Örneklem büyüklüğü arttıkça I. Tip hata oranlarının azalma eğiliminde olduğu belirlenmiştir ancak örneklem büyüklüğü 50000 ve 99990 için bulgular oldukça benzerdir.
5. I. Tip hata oranlarının düşük yetenek düzeylerinde çoğunlukla büyük alfa seviyelerinde kontrol edilebildiği; yüksek yetenek düzeylerinde ise daha düşük alfa seviyelerinde de I. Tip hata oranlarının kontrol edilebildiği görülmüştür.
6. I. Tip hata oranının kontrol edilebildiği koşullarda düşük yetenek düzeylerinde istatistiksel güç oldukça yüksek iken, yetenek düzeyi arttıkça istatistiksel gücün düştüğü ve tahrifata uğrayan test alanların daha düşük oranlarda doğru tespit edilebildiği belirlenmiştir. Ayrıca yüksek yetenek düzeylerinde I. Tip hata oranları daha düşük alfa seviyelerinde kontrol edilmesine karşın bu seviyelerde istatistiksel güç kabul edilebilir düzeyde değildir.
7. Süreklilik düzeltmesinin uygulanması daha çok üst yetenek düzeylerinde istatistiksel gücün, tahrifatlı test alanların doğru tespit edilme oranının, düşmesine neden olmuştur. Düşük yetenek düzeylerinde tahrifatlı test alanların doğru tespit edilme oranı, yüksek yetenek düzeylerine göre süreklilik düzeltmesinden daha az etkilenmiştir.
8. Alfa seviyesi düştükçe istatistiksel güç tüm koşullarda düşmüş ve en düşük alfa seviyelerinde nominal düzeye göre hiçbir koşulda kabul edilebilir istatistiksel güç elde edilememiştir.
9. Tüm örneklem büyüklükleri için 50 maddelik veri setlerinde I. Tip hata oranları 160 maddelik veri setine göre daha iyi kontrol edilebildiğinden daha fazla koşulda istatistiksel güç değerlendirilmiştir. Ancak benzer koşullarda her iki madde sayısı için de hatanın kontrol edildiği durumlarda 160 maddelik veri seti için istatistiksel gücün 50 maddelik veri setine eşit ya da daha büyük olduğu belirlenmiştir.
10. İstatistiksel gücün tüm örneklem üzerinden değil yetenek düzeylerine göre incelendiğinde daha ayrıntılı bilgi verdiği belirlenmiştir. Ayrıca tüm örneklemelerde yetenek düzeylerine göre istatistiksel güçteki değişimin benzer eğilimler gösterdiği görülmüştür.
11. Örneklem büyüklüğü arttıkça istatistiksel güçte artış olduğu ancak 50000 ve 99990 örneklem büyüklükleri için bulguların oldukça benzer olduğu görülmüştür.

12. 50 maddelik veri seti içinsüreklilik düzeltmesi uygulandığında %10 oranında sabit tahrifat durumunda bile en düşük yetenek düzeylerinin büyük alfa seviyelerinde istatistiksel güç oldukça yüksek çıkmıştır.
13. Sabit tahrifat oranı arttıkça tahrifata uğrayan bireyler daha doğru tespit edilmiş ve istatistiksel güç artmıştır.
14. Değişken tahrifat durumunda düşük yetenek düzeyleri için büyük alfa seviyelerinde istatistiksel güç kabul edilebilir düzeydedir.
15. 50 maddelik veri setlerinde I. Tip hata oranlarının “Nominal düzeye” ve “Bradley’in esnek referans aralığına” göre değerlendirilmesi karşılaştırıldığında, genellikle üst yetenek düzeylerinde nominal düzeye göre kontrol edilebilen düşük alfa seviyelerindeki hata oranlarının Bradley’in esnek referans aralığı göz önünde bulundurulduğunda kontrol edilemediği görülmüştür. Birbirine zıt bir durum gibi görünmesine karşın, pratikte istatistiksel güç olarak karşılaştırıldığında, nominal düzey referans alındığında da düşük alfa seviyelerinde istatistiksel güç kabul edilebilir düzeyde değildir. Bu durum da uygulamada Bradley’in esnek referans aralığı ile benzer yorumlamalara neden olmuştur. Düşük yetenek düzeyleri için ise Bradley’in esnek referans aralığı göz önünde bulundurulduğunda, nominal düzeye göre kabul edilen en büyük alfa seviyelerindeki istatistiksel gücün kabul edilebilir değerinin altında kaldığı ancak daha düşük alfa seviyelerinde istatistiksel gücün kabul edilebilir olduğu görülmüştür.
16. 160 maddelik veri seti için Bradley’in esnek referans aralığının kullanılması, nominal düzeye göre I. Tip hata oranlarının daha iyi kontrol edilmesine, dolayısıyla daha fazla koşulda kabul edilebilir istatistiksel güç elde edilmesine neden olmuştur.

Öneriler

Bu çalışma sürecinden ve elde edilen bulgulara dayalı olarak geliştirilen öneriler uygulayıcılara ve araştırmacılara dönük olmak üzere iki alt başlık altında sunulmaktadır.

Uygulayıcılara Öneriler

1. Türkiye’de uygulanan yüksek riskli sınavlarda, sınav yönetimi sürecinin rutin bir parçası olarak yetkililer tarafından test tahrifatının belirlenmesine yönelik çalışmaların yürütülmesi önerilmektedir. Ancak test tahrifatına yönelik ithamların ciddi

yükümlülükleri olup adli yargı süreçlerini de gerektirebileceğinden tahrifata yönelik tespitlerin, sadece yalancı pozitif oranları da içerebilen SBİ gibi istatistiklerin sonuçlarına bağlı olarak değil, çok yönlü araştırmalar sonucunda yapılması gerektiği unutulmamalıdır.

2. Daha fazla koşul için I. Tip hata oranının kontrol edilmesini sağladığından SBİ'nin süreklilik düzeltilmesi uygulanarak kullanılması önerilmektedir.
3. Düşük madde sayısında I. Tip hata oranı kontrolü daha fazla olduğu için madde sayısı fazla olan uygulamalarda SBİ kullanımını önerilmemektedir.
4. SBİ'nin 500 örneklem büyüklüğünde kullanımında I. Tip hata oranı kontrolünün birçok koşulda sağlanamadığı göz önünde bulundurularak, 500'den küçük örneklem büyüklükleri için kullanımı önerilmemektedir.
5. SBİ düşük yetenek düzeyleri için küçük alfa seviyelerinde genellikle I. Tip hata oranı kontrolünü sağlayamadığından, tahrifatın belirlenmesinde büyük alfa düzeylerinin dikkate alınmasının daha uygun olduğu düşünülmektedir. Ancak büyük alfa düzeylerinin, düşük alfa düzeylerine göre daha fazla hata içerdiği göz önünde bulundurulmalıdır.
6. İstatistiksel gücün yetenek düzeylerinde farklılık gösterdiği belirlendiği için tahrifata uğrayan bireylerin belirlenmesinde yetenek düzeylerine göre ayırım yapılması önerilmektedir. Buna bağlı olarak da yüksek yetenek düzeylerindeki bireylere göre tahrifata uğraması daha muhtemel olan düşük yetenek düzeylerinde, bireylerin daha yüksek oranlarda doğru belirlendiğinin göz önünde bulundurulması önerilmektedir.

Araştırmacılara Öneriler

1. Farklı tahrifat oranlarında SBİ'nin I. Tip hata oranları ve istatistiksel gücü belirlenebilir.
2. SBİ'nin I. Tip hata oranı kontrolünün madde sayısına göre farklılık gösterdiği belirlendiğinden, madde sayısının 50'nin altında ve 50 ile 160 arasında olduğu koşullar için SBİ'nin I. Tip hata oranı ve istatistiksel gücünü belirleme çalışmaları yürütülebilir.
3. Araştırma kapsamında yetenek kestirimi STM'ye göre yapılmıştır. SBİ'nin I. Tip hata ve istatistiksel gücü farklı yetenek kestirim modelleri (2 PLM, 3 PLM, nonparametrik modeller vb.) için belirlenebilir.

4. Arařtırma kapsamında madde ve yetenek parametreleri de simülatif olarak üretilmiř ve simülatif veriler üzerinde çalıřılmıřtır. Gerçek veriler üzerinde ve gerçek parametrelerden üretilen simülatif veriler üzerinde çalıřılabilir.
5. Çalıřmada yetenek düzeyleri, her bir yetenek düzeyinde eřit sayıda test alanın yer alacađı řekilde belirlenmiřtir. Bu durum varyansın uç yetenek düzeylerinde daha büyük olmasına neden olabileceđi için yetenek düzeyleri yetenek aralıklarına göre belirlenerek I. Tip hata ve istatistiksel güç çalıřmaları yapılabilir.
6. Arařtırma kapsamında iyicil silmeler simüle edilmiřtir. İyicil silmelerin gerekenden fazla oranda simüle edilmiř olması SBİ'nin gücünü etkileyebilir. Bu nedenle gerçek sınav uygulamalarından elde edilen optik formlar üzerinden yalnızca hileli silmeler simüle edilerek tahrifat yapılıp SBİ'nin I. Tip hata oranları ve istatistiksel gücü belirlenebilir.
7. Gerçek uygulamalarda madde parametreleri tahrifatlı veriler üzerinden kestirileceđi için madde parametrelerinin gerçek deđerleri yansıtmaması beklenmektedir. Bu durumun SBİ'nin istatistiksel gücü ve I. Tip hata oranı üzerindeki etkisi arařtırılabilir.
8. I. Tip hata oranlarının tahrifat sayısına göre deđiřimini incelemek üzere, I. Tip hata oranları tahrifat sayılarına göre ayrılarak belirlenebilir.
9. Arařtırma kapsamında iyicil silme türlerine göre ayırım yapılmaksızın I. Tip hata oranları incelenmiřtir. İyicil silme türlerine göre ayrılarak I. Tip hata oranlarının incelenmesi bir bařka arařtırmanın konusu olabilir. Ayrıca Türkiye'de uygulanan geniř ölçekli testlerde iyicil silmelerin olasılık dađılımına iliřkin bilgilere ulařılamadıđı için arařtırma kapsamında iyicil silmeler simüle edilirken yurtdıřındaki uygulamaların olasılık dađılımlarından yararlanılmıřtır. Türkiye'deki geniř ölçekli testlerde iyicil silmelerin olasılık dađılımına iliřkin arařtırmalar yürütülebilir.
10. I. Tip hata oranlarının deđerlendirilmesinde nominal düzey ve Bradley'in esnek referans aralıđı için, aynı alfa seviyelerinde kontrol sađlandıđı görüldüđü gibi bazı alfa seviyeleri için farklı kararlar verildiđi de görüldüđünden iki kriter birlikte deđerlendirilebilir.



KAYNAKLAR

- Baker, F. B. (2016). Testlerin Karakteristik Özelliklerinin Belirlenmesi (M. İlhan, Çev.). N. Güler (Ed.), *Madde Tepki Kuramının Temelleri* (ss. 149-162). Ankara: Pegem Akademi.
- Baker, F. & Kim, S. (2004). *Item response theory parameter estimation techniques*, (2nd ed.). New York: Taylor & Francis Group, LLC.
- Başaran, S. (2005). *Diğer Ülkelerde Lise Bitirme Sınavları ve Türk Eğitim Sistemi İçin Lise Bitirme Sınavı Önerisi*. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.
- Belov, D. I. (2014). *Detection of aberrant answer changes via kullback–leibler divergence* (Law School Admission Council Research Report). Newtown, PA: LSAC, Inc.
- Belov, D. I. (2015). Robust detection of examinees with aberrant answer changes. *Journal of Educational Measurement*, 52, 437-456.
- Belov, D. I. (2016). Comparing the performance of eight item preknowledge detection statistics. *Applied Psychological Measurement*, 40, 83–97.
- Bishop, S. & Egan, K. (2017). Detecting erasures and unusual gain scores: Understanding the status quo. In G.J. Cizek and J.A. Wollack (Eds.), *Handbook of quantitative methods for detecting cheating on tests*. Taylor & Francis.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 46, 443-459.
- Bradley, J. (1978). Robustness?. *British Journal Of Mathematical And Statistical Psychology*, 31(2), 144-152. doi: 10.1111/j.2044-8317.1978.tb00581.x
- Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., Kim, K., Falk, C. F., ..., Ogreden, O. (2018). Package ‘mirt’. <https://cran.r-project.org/web/packages/mirt/index.html> adresinden 28.12.2019 tarihinde alınmıştır.
- Chen, W. & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational And Behavioral Statistics*, 22(3), 265-289. doi: 10.3102/10769986022003265
- Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum.
- Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, 10(4), 417. doi:10.2307/3001616
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New Jersey: Lawrence Erlbaum.

- Çokluk, O., Şekercioğlu, G. ve Büyüköztürk, S. (2010). *Sosyal Bilimler İçin Çok Değişkenli İstatistik*. Ankara: Pegem Akademi.
- Davey, T., Nering, M. & Thompson, T. (1997). *Realistic simulation of item response data*. Iowa City, IA.: American Coll. Testing Program
- de Ayala, R. J. & Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model. *Applied Psychological Measurement*, 23(1), 3-19. doi:10.1177/01466219922031130
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford Publication, Inc.
- DeMars, C. (2016). Varsayımlar (M. G. Şahin, Çev.). H. Kelecioğlu (Ed.), *Madde Tepki Kuramı* (ss. 38-60). Ankara: Nobel Yayıncılık.
- Dinç, E. ve Akşit, İ. (2015). Ortaöğretime Geçiş Sınavları ile İlgili Köşe Yazılarının İncelenmesi. *Türkiye Sosyal Araştırmalar Dergisi*, 2, 59-82.
- Embretson, S. E. ve Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Flens, G., Smits, N., Terwee, C. B., Dekker, J., Huijbrechts, I. & de Beurs, E. (2017). Development of a computer adaptive test for depression based on the Dutch-Flemish version of the PROMIS item bank. *Evaluation & the Health Professions*, 40(1), 79-105. doi: 10.1177/0163278716684168
- Fremer, J. J. ve Ferrera, S. (2013). Security in large-scale paper and pencil testing. In A. Wollack & J. J. Fremer (Eds.). *Handbook of test security* (pp. 17-38). New York: Routledge.
- GOSA. (2016). 2011 *Erasure Analysis Process Overview*. The Governor's Office of Student Achievement. Retrieved from: http://gosa.georgia.gov/sites/gosa.georgia.gov/files/related_files/document/Erasure%20Analysis%20Process%20Overview%20Final_021016.pdf in 28.02.2016.
- Gündeğer, C. ve Doğan, N. (2016). Bireyselleştirilmiş Bilgisayarlı Sınıflama Testlerinde Madde Havuzu Özelliklerinin Test Uzunluğu ve Sınıflama Doğruluğu Üzerindeki Etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, doi: 10.16986/HUJE.2016024284
- Haick, A. (2003). *Testing irregularities: Are we getting accurate scores?*. (Unpublished doctoral dissertation). University of La Verne, La Verne, California.
- Hambleton, R., K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Holland, P. W. (1996). *Assessing unusual agreement between the incorrect answers of two examinees using the K-Index: Statistical theory and empirical support*. Princeton, New Jersey: Educational Testing Service.

- Jacob, B. A. & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118 (3), 843-877.
- Jang, E. & Roussos, L. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal Of Educational Measurement*, 44(1), 1-21. doi: 10.1111/j.1745-3984.2007.00024.x
- Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, 21(1),31-41.
- Karasar, N. (2007). *Bilimsel Araştırma Yöntemi*. Ankara: Nobel Yayın Dağıtım.
- Kim, S. & Lee, W. (2004). *IRT scale linking methods for mixed-format tests* (ACT Research Report 2004-5). Iowa City, IA: Act, Inc.
- Koğar, H. (2018). *An examination of parametric and nonparametric dimensionality assessment methods with exploratory and confirmatory mode*. *Journal of Education and Learning*, 7(3), 148. doi:10.5539/jel.v7n3p148
- Kumandaş, H. (2013). *Yükseköğretime Öğrenci Seçmede ve Yerleştirmede Kullanılan Sınavların Oluşturduğu Risk Faktörlerinin Okul Başarısı Üzerindeki Etkileri*. (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara, Türkiye.
- Lee, Y. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q3 statistics in an EFL reading comprehension test. *Language Testing*, 21(1), 74-100. doi: 10.1191/0265532204lt260oa
- Milli Eğitim Bakanlığı. (2011a). Milli Eğitim Bakanlığı merkezi sistem sınav yönergesi. Milli Eğitim Bakanlığı. Erişim adresi: http://mevzuat.meb.gov.tr/html/25822_0.html (Erişim tarihi 03. 03. 2016)
- Milli Eğitim Bakanlığı. (2011b). 7. sınıflar seviye belirleme sınavı 2011 yılı sayısal bilgiler. Milli Eğitim Bakanlığı. Erişim adresi: <http://www.meb.gov.tr/duyurular/duyurular2011/EGITEK/SBS7/SBS7SayısalBilgiler.pdf> (Erişim tarihi 03. 03. 2016)
- MI. (2012). 2011 *Erasure analysis report new jersey assessment of skills and knowledge (NJ ASK 3-8)*. Measurement Incorporated. Retrieved from: <https://education.state.nj.us/broadcasts/2012/AUG/16/7683/2011%20NJ%20ASK%20Erasure%20Analysis%20Report%20FINAL.PDF> in 28.02.2016.
- Monseur, C., Baye, A., Lafontaine, D. & Quittre, V. (2011). PISA test format assessment and the local independence assumption. In M. Von Davier ve D. Hasted (Ed.), *IERI Monograph Serises: Issues and Methodologies in Large-Sclae Assessments* (Vol. 4, pp. 131-155). Princeton, NJ: IEA-ETS Research Institute.
- Morris TP. White IR & Crowther MJ. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074–2102. doi: 10.1002/sim.8086.

- Mroch, A. A., Lu, Y., Huang, C. & Harris, D. J. (2012). *Patterns of erasure behavior for a large-scale assessment*. Paper presented at the 2012 Conference on Statistical Detection of Potential Test Fraud, Lawrence, KS. Retrieved from: https://img1.wsimg.com/blobby/go/71ff6051-7987-4285-89a3-874c4bfcc71c/downloads/1d2lkha9p_311182.pdf?ver=1595864143672 in 5.17.2017.
- Ogasawara, H. (2001). Standart errors of item response theory equating/linking by response function methods. *Applied Psychological Measurement*, 25 (1), 53- 67.
- Olejnik, S. (1984). Planning educational research. *The Journal of Experimental Education*, 53(1), 40-48. doi: 10.1080/00220973.1984.10806360
- Ostini, R., & Nering, M. (2006). *Polytomous item response theory models*. Thousand Oaks: Sage.
- Ölçme, Seçme ve Yerleştirme Merkezi. (2007a). Sınavda Uyulması Gereken Kurallar (2007 KPDS Mayıs Dönemi). <http://www.osym.gov.tr/belge/1-7871/sinavda-uyulmasi-gereken-kurallar.html>. (Erişim tarihi 28. 02. 2016)
- Ölçme, Seçme ve Yerleştirme Merkezi. (2007b). Temel İlke ve Koşullar (2007 KPDS Mayıs Dönemi). <http://www.osym.gov.tr/belge/1-7861/temel-ilkeler-ve-kosullar.html>. (Erişim tarihi 28. 02. 2016)
- Ölçme, Seçme ve Yerleştirme Merkezi. (2007c). 2007- ÖSS'ye İlişkin Sayısal Bilgiler. <http://osym.gov.tr/belge/1-8315/2007-ossye-iliskin-sayisal-bilgiler.html>. (Erişim tarihi 28. 02. 2016)
- Ölçme, Seçme ve Yerleştirme Merkezi. (2010a). 2010-KPSS Lisans: Eğitim Bilimleri Testinin İptali basın açıklaması. <http://www.osym.gov.tr/belge/1-12135/2010-kpss-lisans-egitim-bilimleri-testinin-iptali-17092-.html>. (Erişim tarihi 22. 02. 2016)
- Ölçme, Seçme ve Yerleştirme Merkezi. (2010b). 2010 YGS'ye İlişkin Sayısal Bilgiler. <http://www.osym.gov.tr/belge/1-7871/sinavda-uyulmasi-gereken-kurallar.html>. (Erişim tarihi 28. 02. 2016)
- Ölçme, Seçme ve Yerleştirme Merkezi. (2014). ÖSYM Mevzuatı-Kanun-Yönetmelikler. <http://dokuman.osym.gov.tr/pdfdokuman/2014/Genel/OSYM-Mevzuati-KanunveYonetmelikler06112014.pdf>. (Erişim tarihi 01. 03. 2016)
- Ölçme, Seçme ve Yerleştirme Merkezi. (2015). Yaklaşan 2015-ÖSYS ve Bazı Konularda Bilgilendirme basın açıklaması. <http://www.osym.gov.tr/belge/1-23225/basin-aciklamasi-yaklasan-2015-osys-ve-bazi-konularda-b-.html>. (Erişim tarihi 22. 02. 2016)
- Ölçme, Seçme ve Yerleştirme Merkezi. (2016). Sınav Uygulama Yönergesi. ÖSYM Sınav Görevlileri Yönetimi Daire Başkanlığı. https://gis.osym.gov.tr/osym_genel_sinav_uygulama.pdf (Erişim tarihi 03. 03. 2016)
- Paek, I. & Young, M. J. (2005). Investigation of student growth recovery in a fixed- item linking procedure with a fixed-person prior distribution for mixed-format test data. *Applied Measurement in Education*, 18 (2), 199-215.

- Pagano, R. R. (2013). *Understanding statistics in the behavioral sciences*. Belmont: Wadsworth, Cengage Learning.
- Qin, S. (2016). *Multilevel detection of possible test tampering through erasure analysis* (Doctoral dissertation, University of Georgia). Retrieved from https://getd.libs.uga.edu/pdfs/qin_shanshan_201605_phd.pdf
- Qualls, A. L. (2001). Can knowledge of erasure behavior be used as an indicator of possible cheating? *Educational Measurement: Issues and Practice*, 20(1), 9–16.
- Reeve, B., Hays, R., Bjorner, J., Cook, K., Crane, P., Teresi, J., ... & Cella, D. (2007). Psychometric evaluation and calibration of health-related quality of life item banks. *Medical Care*, 45(Suppl 1), S22-S31. doi:10.1097/01.mlr.0000250483.85507.04
- Reif, M. (2015). Package ‘mcIRT’. <https://mran.microsoft.com/snapshot/2017-12-11/web/packages/mcIRT/index.html> adresinden 17.03.2016 tarihinde alınmıştır.
- Robitzsch, A. (2019). Package ‘sirt’. <https://cran.r-project.org/web/packages/sirt/index.html> adresinden 10.01.2020 tarihinde alınmıştır.
- Roussos, L. & Ozbek, O. Y. (2006). Formulation of the DETECT population parameter and evaluation of detect estimator bias. *Journal of Educational Measurement*, 43(3), 215-243. doi: 10.1111/j.1745-3984.2006.00014.x
- Serlin, R. (2000). Testing for robustness in Monte Carlo studies. *Psychological Methods*, 5(2), 230-240. doi: 10.1037/1082-989x.5.2.230
- Sinharay, S. & Johnson, M. S. (2016). Three new methods for analysis of answer changes. *Educational and Psychological Measurement*, 77(1), 54-81.
- Sinharay, S., Duong, M. Q. & Wood, S. W. (2017). A new statistic for detection of aberrant answer changes. *Journal of Educational Measurement*, 54(2), 200–217. doi:10.1111/jedm.12141
- Smits, N., Cuijpers, P. & van Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry Research*, 188(1), 147-155. doi: 10.1016/j.psychres.2010.12.001
- Soltana, G., Sabetzadeh, M., & Briand, L. C. (2017). *Synthetic data generation for statistical testing*. Paper presented at 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE). Retrieved from: <https://dl.acm.org/doi/pdf/10.5555/3155562.3155670> in 10.10.2020.
- Sotaridona, L. S. & Meijer, R. R. (2002). Statistical properties of the k-index for detecting answer copying. *Journal of Educational Measurement*, 39, 115-132.
- Sotaridona, L. S. ve Meijer, R. R. (2003). Two new statistics to detect answer copying. *Journal of Educational Measurement*, 40, 53-69.

- Sünbül, O. & Yormaz, S. (2018). Investigating the performance of omega index according to item parameters and ability levels. *Eurasian Journal of Educational Research*, 74, 207-226, doi: 10.14689/ejer.2018.74.11
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Pearson/Allyn & Bacon.
- Tan, Q., Cai, Y., Li, Q., Zhang, Y. & Tu, D. (2018). Development and validation of an item bank for depression screening in the Chinese population using computer adaptive testing: A simulation study. *Frontiers In Psychology*, 9, doi: 10.3389/fpsyg.2018.01225
- Thiessen, B. (2006). *Thesis equivalency project-educator cheating: Classification, explanation, and detection*. 03.06.2017 tarihinde <http://bradthiessen.com/html5/docs/cheating.pdf> adresinden alınmıştır.
- Thissen, D. & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 73-140). Mahwah, NJ: Erlbaum.
- Turgut, F. (1983). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Saydam Matbaacılık.
- van der Linden, W. J. ve Jeon, M. (2012). Modeling answer changes on test items. *Journal of Educational and Behavioral Statistics*, 37, 180-199.
- van der Linden, W. J. & Sotaridona, L. (2006). Detecting answer copying when the regular response process follows a known response model. *Journal of Educational and Behavioral Statistics*, 31, 283-304.
- Wollack, J. A. (1996). *Detection of answer copying using item response theory*. (Doctor of Philosophy). Available from ProQuest Dissertations and Theses database. (UMI No. 9622477).
- Wollack, J. A. & Cohen, A. S. (1998). Detection of answer copying with unknown item and trait parameters. *Applied Psychological Measurement*, 22, 144-152.
- Wollack, J. A. & Fremer, J. J. (2013). Introduction: The test security threat. In J. A. Wollack & J. J. Fremer (Ed.). *Handbook of test security* (pp. 1-13). New York: Routledge.
- Wollack, J. A., Cohen, A. S. & Eckerly, C. A. (2015). Detecting test tampering using item response theory. *Educational and Psychological Measurement*, 75 (6), 931-953.
- Wollack, J. A. & Eckerly, C. A. (2017). Detecting test tampering at the group level. G. J. Cizek ve J. A. Wollack (Ed.). *Handbook of quantitative methods for detecting cheating on tests* (s. 214-231). New York: Routledge.
- Yen, W. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125-145. <https://doi.org/10.1177/014662168400800201>

- Yen, W. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal Of Educational Measurement*, 30(3), 187-213. doi: 10.1111/j.1745-3984.1993.tb00423.x
- Yormaz, S. (2014). *Kopya Belirleme İndekslerinin Çeşitli Koşullar Altında I. Tip Hata Oranlarının ve Güçlerinin Belirlenmesi*. (Yayımlanmamış Yüksek Lisans Tezi). Mersin Üniversitesi, Mersin, Türkiye.
- Zhang, J. (2007). Conditional covariance theory and detect for polytomous items. *Psychometrika*, 72(1), 69-91. doi: 10.1007/s11336-004-1257-7
- Zhang, J. & Stout, W. (1999). The theoretical detect index of dimensionality and its application to approximate simple structure. *Psychometrika*, 64(2), 213-249. doi: 10.1007/bf02294536
- Zopluoglu, C. & Davenport, E.C, Jr. (2012). The empirical power and type I error rates of the GBT and ω indices in detecting answer copying on multiple-choice tests. *Educational and Psychological Measurement*, 72 (6), 975–1000.
- Zopluoglu, C. (2013). Copydetect: An R package for computing statistical indices to detect answer copying on multiple-choice examinations. *Applied Psychological Measurement*, 37(1), 93–95.



EK 1. Etik Kurul Onayı

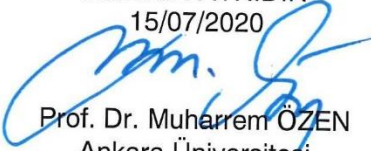
ANKARA ÜNİVERSİTESİ SOSYAL BİLİMLER ALT ETİK KURULU KARAR ÖRNEĞİ

Karar Tarihi : 15/07/2020
Toplantı Sayısı : 6
Karar Sayısı : 124

124- Üniversitemiz Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı doktora öğrencisi **Esra Kınay Çiçek**'in "Test Tahrifatının Olduğu Durumlarda Kullanılan Silme Belirleme İndeksinin Ürettiği I.Tip Hata ve İstatistiksel Gücün Farklı Koşullar Altında İncelenmesi" başlıklı tezi ile ilgili 07/07/2020 tarihli "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

Araştırmanın aydınlatılmış onam formunun uygun biçimde hazırlandığı, görüşme sorularının yansız olduğu ve ayrımcı ifadeler içermediği, yöntemsel tutarlığın bulunduğu, dolayısıyla araştırma tasarımının etik açıdan uygun olduğu görülmekle birlikte; Üniversitemiz Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı doktora öğrencisi **Esra Kınay Çiçek**'in "Test Tahrifatının Olduğu Durumlarda Kullanılan Silme Belirleme İndeksinin Ürettiği I.Tip Hata ve İstatistiksel Gücün Farklı Koşullar Altında İncelenmesi" başlıklı tezi tamamlanmış bir tez çalışması olduğundan Etik Kurul onay belgesi verilememektedir.

ASLININ AYNIDIR
15/07/2020


Prof. Dr. Muharrem ÖZEN
Ankara Üniversitesi
Etik Kurulu Başkanı

EK 2. 50 ve 160 Maddelik Veri Setlerine Ait R'da Üretilen Seçenek Eğim ve Kesişim Parametreleri

Tablo 26

50 Maddelik Veri Setinde 49995 Örneklem Büyüklüğüne Ait Eğim (λ) ve Kesişim (ζ) Parametreleri

Madde	λ_1	λ_2	λ_3	λ_4	λ_5	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5
1	-1.18	1.02	-1.04	0.12	1.08	-1.29	-0.06	0.60	0.55	0.19
2	-1.08	1.37	0.63	-0.92	-0.01	-1.27	-1.19	0.03	0.44	1.99
3	-1.38	0.71	-1.57	1.95	0.29	-0.43	-0.60	1.07	-0.97	0.93
4	-1.68	0.77	0.27	-1.21	1.86	1.35	-0.96	0.85	-1.41	0.18
5	-1.65	-0.31	1.39	-0.54	1.12	1.37	-1.33	-0.37	1.04	-0.70
6	-1.73	0.31	2.05	0.57	-1.20	1.07	1.18	-0.66	-0.13	-1.45
7	-1.00	-0.16	-0.62	2.28	-0.51	2.90	0.47	-0.72	-1.77	-0.88
8	-1.10	-1.13	0.10	1.97	0.16	1.95	-0.45	-1.47	0.05	-0.09
9	-1.60	0.68	1.13	-0.03	-0.17	-1.51	0.81	-1.10	0.52	1.27
10	-1.32	-0.85	2.17	0.38	-0.38	0.25	-0.47	0.31	-0.33	0.24
11	-0.59	1.09	-0.35	0.67	-0.81	-1.42	1.30	0.05	0.84	-0.77
12	-1.78	0.87	-0.44	0.31	1.03	-0.57	0.54	1.71	0.21	-1.90
13	-0.58	-0.94	0.61	-0.60	1.51	-1.05	0.20	-0.40	-0.47	1.71
14	-0.56	0.95	0.30	0.37	-1.07	0.69	1.24	-0.55	-0.91	-0.48
15	-1.51	-0.35	0.03	-0.24	2.07	-0.28	-1.98	-0.44	1.61	1.09
16	-1.07	0.28	0.90	0.33	-0.44	0.27	0.92	-0.46	-1.01	0.28
17	-1.31	0.41	-1.10	-0.62	2.62	1.45	1.26	-2.20	-0.51	0.00
18	-0.85	1.13	-0.95	0.17	0.51	1.54	-0.26	0.26	-0.55	-0.98
19	-1.11	0.55	-0.11	0.57	0.10	0.22	-1.68	0.68	1.17	-0.40
20	-0.95	0.64	-0.05	1.27	-0.91	0.51	-0.87	0.67	0.39	-0.70
21	-1.75	0.00	1.26	0.74	-0.25	0.18	1.07	-1.22	-0.61	0.58
22	-1.46	0.73	-0.20	2.38	-1.46	0.16	0.50	0.38	-0.34	-0.70
23	-0.88	-0.50	1.76	0.61	-0.99	2.33	-0.91	-0.34	-1.09	0.01
24	-1.44	1.42	-0.68	1.82	-1.12	-0.68	0.24	0.02	0.22	0.19
25	-1.38	1.79	-0.60	-0.96	1.14	-0.68	-0.73	1.71	1.39	-1.69
26	-1.07	1.51	-0.37	-0.99	0.92	-0.23	-0.31	-0.05	1.11	-0.53
27	-0.93	0.46	-0.89	1.74	-0.37	-0.61	0.16	-1.02	0.42	1.05
28	-1.93	0.34	-0.87	0.04	2.41	0.75	-0.61	1.49	-1.32	-0.31
29	-1.84	0.63	-0.08	0.62	0.68	0.23	-0.12	-1.58	0.82	0.65
30	-0.78	0.03	0.74	-0.67	0.68	0.86	0.05	-1.56	-0.64	1.29
31	-0.89	-0.39	-0.34	1.45	0.17	0.82	0.17	-0.28	0.10	-0.81
32	-1.52	1.01	1.00	0.62	-1.12	0.69	-0.49	0.68	-0.68	-0.20
33	-1.88	0.42	0.10	1.45	-0.10	0.34	-1.41	-0.72	-1.47	3.27
34	-1.58	0.43	-0.70	1.13	0.71	-1.01	-0.33	0.09	0.18	1.07
35	-1.36	1.05	-0.71	0.43	0.59	1.02	0.67	-2.87	0.77	0.41
36	-1.86	0.10	1.25	-0.27	0.78	0.75	-0.90	1.54	0.18	-1.57
37	-0.63	0.67	-0.64	-0.07	0.68	0.28	0.17	-0.35	-0.21	0.11
38	-2.12	-0.50	1.48	0.32	0.81	-1.40	-0.63	1.10	0.56	0.37

39	-0.89	0.13	-1.46	2.24	-0.03	-0.51	0.66	-1.36	-0.24	1.45
40	-1.15	0.10	2.59	-1.10	-0.43	0.04	-0.18	-1.31	1.16	0.28
41	-1.75	0.11	-0.71	1.22	1.13	1.40	-2.56	0.20	-0.14	1.10
42	-1.92	0.35	1.15	-0.15	0.56	-0.42	-0.58	0.65	-0.64	0.99
43	-1.67	1.08	-1.24	1.77	0.06	-1.62	0.98	0.15	-0.13	0.62
44	-1.63	0.68	0.45	-1.30	1.81	0.32	-1.12	0.07	0.00	0.73
45	-0.92	1.55	0.63	-1.22	-0.05	0.32	-2.01	1.44	0.78	-0.53
46	-1.85	0.75	-0.17	0.16	1.11	0.33	-1.05	1.00	-0.73	0.45
47	-0.69	0.36	0.94	-0.21	-0.39	-0.09	0.15	0.33	0.68	-1.07
48	-1.58	-0.55	1.60	0.61	-0.08	2.79	0.19	-1.37	-1.11	-0.50
49	-1.17	1.20	0.49	-0.13	-0.39	-0.29	0.62	-0.86	1.25	-0.72
50	-0.85	-0.18	-0.42	1.28	0.17	-1.59	-0.28	-2.26	2.93	1.20



Tablo 27

160 Maddelik Veri Setinde 49995 Örneklem Büyüklüğüne Ait Eğim (λ) ve Kesişim (ζ) Parametreleri

Madde	λ_1	λ_2	λ_3	λ_4	λ_5	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5
1	-0.53	0.72	-0.78	0.31	0.28	-0.09	-0.82	1.59	-0.88	0.20
2	-0.86	0.39	1.06	0.61	-1.21	1.00	-1.36	-0.99	0.84	0.51
3	-1.58	0.27	0.71	0.11	0.49	-0.48	-0.97	-0.06	0.08	1.43
4	-0.87	0.75	0.40	0.37	-0.65	1.22	0.36	-1.59	-0.54	0.54
5	-1.27	-0.39	1.05	0.55	0.04	-0.34	0.66	1.01	0.09	-1.42
6	-1.69	0.16	0.88	-0.83	1.48	0.91	-0.58	-0.22	0.20	-0.31
7	-1.13	0.11	0.16	1.10	-0.24	0.65	-0.12	-0.45	-0.41	0.34
8	-1.66	0.38	0.40	0.47	0.40	-1.39	0.40	1.20	0.03	-0.24
9	-0.60	-0.50	0.29	1.36	-0.55	-0.87	-0.53	0.78	-0.79	1.40
10	-1.10	0.66	-0.91	1.04	0.32	-0.22	-0.54	0.69	-1.06	1.13
11	-1.60	0.33	1.41	-0.23	0.09	0.40	-1.12	0.85	0.19	-0.32
12	-0.82	-0.18	-0.31	0.53	0.78	1.14	-1.17	-0.25	0.96	-0.69
13	-1.21	-1.13	0.46	0.29	1.59	0.66	-1.77	0.35	0.53	0.24
14	-1.59	0.87	0.93	-0.57	0.35	-2.48	0.08	0.37	1.34	0.69
15	-1.37	0.65	0.44	-0.15	0.42	0.50	0.11	-0.42	0.02	-0.22
16	-1.25	-0.18	0.31	0.57	0.56	0.31	-0.21	0.02	0.11	-0.24
17	-0.75	-0.10	0.87	-1.01	0.99	-1.55	0.42	-0.21	1.30	0.05
18	-0.93	0.79	0.39	-0.07	-0.18	-0.27	0.41	0.68	-1.21	0.39
19	-1.68	0.30	1.08	0.30	0.01	0.06	-0.27	-0.50	-0.37	1.07
20	-0.58	-0.61	1.21	0.09	-0.11	-0.68	0.01	1.02	-0.66	0.31
21	-0.54	-0.09	0.78	-0.08	-0.07	-3.05	0.06	-0.15	2.15	0.99
22	-1.55	0.33	0.50	0.73	0.00	-1.03	1.01	1.17	-0.84	-0.31
23	-1.43	0.27	0.78	-1.07	1.44	1.58	-0.49	-0.15	-1.66	0.73
24	-0.96	0.30	-0.35	-0.91	1.92	-0.36	-1.46	1.10	0.75	-0.03
25	-0.62	0.96	0.77	-0.53	-0.58	1.14	0.27	-0.48	-1.09	0.15
26	-1.23	0.85	2.15	-1.05	-0.71	0.54	-0.34	0.13	-1.08	0.76
27	-0.70	0.07	0.20	0.36	0.06	1.38	-0.47	-0.87	0.51	-0.55
28	-1.40	1.10	0.46	-0.12	-0.05	0.22	0.99	-0.63	-0.12	-0.46
29	-1.51	-0.31	0.61	1.16	0.05	0.37	-0.18	-0.08	-0.99	0.88
30	-1.65	1.49	0.88	-0.01	-0.71	-1.22	1.03	2.25	-0.85	-1.21
31	-1.31	0.53	0.81	0.46	-0.50	-1.02	0.39	1.50	-0.43	-0.44
32	-1.42	0.66	0.07	0.34	0.35	-0.21	1.44	1.31	-1.18	-1.36
33	-0.75	-0.55	0.33	0.43	0.54	-1.30	0.87	-0.22	-0.20	0.86
34	-0.76	0.33	0.47	-0.12	0.08	1.65	-1.02	-1.83	0.74	0.47
35	-1.09	0.07	0.77	0.63	-0.38	-1.10	1.08	-0.46	0.07	0.41
36	-1.50	1.34	-0.69	-0.80	1.65	0.29	-0.75	-1.14	1.11	0.48
37	-1.44	-0.27	-0.29	0.46	1.53	-0.11	-0.54	0.57	-0.22	0.30
38	-1.46	-0.04	2.02	1.20	-1.72	0.54	-1.76	1.51	0.45	-0.74
39	-1.37	0.74	-0.16	0.99	-0.19	-1.78	1.69	-1.68	1.71	0.07
40	-0.81	-0.35	0.92	0.57	-0.33	0.72	-2.68	1.42	0.23	0.31

41	-1.37	-1.44	0.31	1.54	0.96	-0.13	0.31	0.92	-0.34	-0.75
42	-0.57	-0.01	0.94	-0.31	-0.05	1.17	0.54	0.12	-0.59	-1.23
43	-0.55	0.35	0.08	0.27	-0.15	-0.29	-0.35	0.81	0.59	-0.75
44	-0.60	0.27	-0.34	0.16	0.51	-1.02	2.83	-0.25	-1.21	-0.34
45	-0.67	0.87	0.48	-1.77	1.10	0.35	-0.44	-0.01	-0.73	0.82
46	-0.79	0.17	1.68	1.08	-2.14	-0.18	-0.15	-0.38	0.16	0.54
47	-1.33	1.44	-1.13	0.18	0.84	-0.35	-0.45	0.60	0.79	-0.58
48	-0.97	0.06	0.64	-0.04	0.32	1.41	1.62	-0.91	-0.84	-1.28
49	-1.04	0.45	0.80	-0.09	-0.12	1.47	-0.18	-1.61	1.21	-0.89
50	-1.44	0.73	0.24	-0.15	0.61	0.56	-1.21	1.63	0.03	-1.01
51	-0.50	0.07	0.71	-0.10	-0.18	0.31	-1.33	0.32	0.07	0.62
52	-1.61	1.07	-0.21	1.09	-0.35	-0.77	-0.95	1.63	0.55	-0.46
53	-0.89	-0.12	0.03	-0.82	1.81	0.12	-1.75	0.31	0.14	1.18
54	-0.94	0.69	-0.50	-0.07	0.82	-1.13	0.82	1.84	-0.04	-1.48
55	-1.55	-0.22	0.02	1.27	0.48	0.04	0.03	-0.29	-0.83	1.05
56	-0.71	0.41	0.58	0.10	-0.38	0.38	1.06	-1.22	1.22	-1.43
57	-0.51	-0.21	1.80	-0.61	-0.47	-0.17	0.50	-0.87	0.92	-0.38
58	-0.74	-0.57	1.28	-0.10	0.14	1.81	0.91	-1.07	-1.20	-0.46
59	-1.70	0.09	1.21	0.13	0.28	-0.98	-0.86	0.65	-0.51	1.71
60	-0.76	0.35	0.49	0.11	-0.19	-0.09	0.36	-0.18	-0.20	0.10
61	-1.51	0.74	0.69	0.90	-0.82	2.32	0.58	0.16	-1.75	-1.32
62	-0.84	1.04	-0.05	-0.21	0.05	0.13	0.42	0.16	-0.13	-0.57
63	-1.62	1.43	-0.20	-0.03	0.42	0.59	1.44	-0.45	-1.04	-0.54
64	-0.81	0.98	-0.54	0.28	0.08	-0.25	-0.01	0.16	-0.32	0.41
65	-1.36	-0.17	-0.09	1.11	0.51	0.21	0.27	-0.05	1.47	-1.90
66	-0.54	-0.22	0.53	0.15	0.08	1.02	-0.02	-0.11	0.22	-1.11
67	-1.23	0.76	-0.01	0.98	-0.50	0.67	-0.73	1.09	-0.34	-0.69
68	-0.67	-0.63	0.44	0.82	0.03	-0.44	-0.57	1.54	0.21	-0.74
69	-1.07	-1.30	1.09	1.08	0.20	0.44	-0.23	-0.03	-0.17	0.00
70	-1.41	-1.30	-0.45	2.30	0.85	-0.65	0.18	0.20	0.35	-0.09
71	-1.28	0.17	0.81	-0.11	0.41	1.80	0.44	-1.06	-0.43	-0.75
72	-0.54	0.64	0.06	0.49	-0.65	-0.12	-0.36	0.13	-0.24	0.58
73	-0.85	2.10	-0.11	-0.13	-1.00	1.75	-0.87	-1.55	0.28	0.38
74	-0.68	0.96	-0.61	-0.23	0.57	-0.97	0.58	0.77	0.15	-0.53
75	-0.63	-0.13	0.61	0.15	0.00	0.37	-0.52	-0.53	0.80	-0.13
76	-1.27	-0.31	0.94	0.24	0.40	0.50	1.00	-0.28	-1.02	-0.20
77	-0.88	0.90	0.51	-1.31	0.79	1.57	0.48	0.67	-0.84	-1.89
78	-1.14	-0.32	0.86	-0.28	0.88	1.59	0.29	-1.58	-0.51	0.20
79	-1.42	0.13	0.90	0.78	-0.38	-0.80	0.97	-1.06	-0.21	1.10
80	-1.23	-0.03	0.47	-0.34	1.13	-0.32	0.31	1.95	-0.09	-1.84
81	-0.46	-0.53	1.33	-1.50	1.16	-0.81	-0.86	1.08	-0.09	0.68
82	-0.77	1.70	-0.72	-0.62	0.41	-0.46	-1.04	-0.79	-0.53	2.81
83	-1.61	-0.75	1.55	0.93	-0.12	0.44	0.42	-0.11	0.03	-0.79
84	-1.68	1.20	-0.23	-0.14	0.85	-0.37	0.53	1.67	-0.42	-1.40
85	-0.80	1.29	0.53	-1.13	0.11	-1.04	-0.06	-0.23	0.21	1.11

86	-0.60	0.47	0.05	-0.18	0.26	1.13	0.35	-0.91	-1.05	0.47
87	-1.28	1.20	-0.31	0.37	0.02	-0.15	-0.30	0.53	0.52	-0.60
88	-0.65	0.69	0.01	-0.26	0.21	0.92	-1.06	1.04	-0.69	-0.21
89	-0.83	0.24	0.41	-0.02	0.19	-1.00	0.72	-0.63	0.90	0.00
90	-1.25	0.37	0.23	0.76	-0.12	0.11	0.21	-0.58	-0.93	1.20
91	-0.97	0.89	-0.30	-0.01	0.40	-0.49	-0.73	1.15	-1.31	1.37
92	-1.27	1.04	0.69	-0.61	0.15	-0.18	-0.45	0.28	-0.82	1.16
93	-1.12	1.38	-0.08	-0.63	0.45	1.26	-0.16	0.56	0.14	-1.79
94	-0.67	-0.64	-0.12	1.51	-0.08	0.28	1.74	-0.67	-0.96	-0.40
95	-1.41	1.25	-0.94	0.56	0.54	0.03	-1.27	0.44	0.44	0.36
96	-1.32	0.28	-1.42	1.60	0.87	0.33	-0.40	-0.80	0.06	0.82
97	-0.89	0.16	0.51	0.37	-0.15	0.54	0.01	-1.06	-0.70	1.20
98	-0.54	1.38	-1.25	0.85	-0.44	-0.45	0.42	-1.49	0.68	0.85
99	-0.68	1.01	-1.09	0.94	-0.19	0.83	0.38	-0.71	1.52	-2.03
100	-1.31	0.46	-0.28	0.89	0.24	0.61	0.38	0.12	0.54	-1.66
101	-0.58	-0.23	1.29	-0.61	0.14	0.94	0.06	0.03	-0.98	-0.05
102	-0.91	-0.51	0.68	0.13	0.60	-0.21	0.72	-0.53	-0.18	0.19
103	-0.80	0.21	0.73	-0.54	0.41	0.40	0.62	1.41	-0.93	-1.50
104	-0.56	-0.05	0.37	0.67	-0.44	0.88	-1.03	0.09	0.02	0.03
105	-1.34	-0.15	-0.45	0.37	1.57	2.14	-0.84	-0.79	-0.44	-0.07
106	-0.83	0.88	0.02	0.36	-0.43	0.82	-0.45	0.89	-0.21	-1.05
107	-1.21	0.07	0.03	1.07	0.04	-0.06	0.18	-1.00	0.91	-0.04
108	-1.19	-0.05	0.76	1.09	-0.61	0.36	-1.77	0.57	0.29	0.55
109	-1.07	0.41	0.10	-0.51	1.06	0.89	0.61	0.08	-0.68	-0.90
110	-1.37	1.15	0.23	-0.40	0.40	3.24	-0.71	-0.05	-1.82	-0.65
111	-0.74	0.09	-1.08	0.45	1.29	0.11	1.46	-1.69	0.05	0.06
112	-0.62	0.33	0.20	-0.18	0.27	-1.62	1.70	-0.72	0.10	0.53
113	-0.43	-1.22	-0.11	0.89	0.87	1.14	-0.17	-1.48	0.00	0.50
114	-1.07	0.38	-0.68	0.39	0.98	1.42	-1.53	0.38	0.85	-1.13
115	-0.72	0.29	0.76	0.53	-0.85	-0.36	-1.11	-0.27	0.56	1.18
116	-0.60	-1.16	0.64	0.10	1.02	-0.59	0.68	-0.47	-0.57	0.94
117	-1.60	-0.07	0.29	1.31	0.07	-1.22	1.08	-0.43	0.03	0.54
118	-1.35	-0.63	0.31	1.86	-0.19	1.49	0.66	0.22	-1.78	-0.59
119	-1.53	-0.39	1.44	1.06	-0.58	0.10	-0.95	0.40	-1.33	1.78
120	-1.25	0.56	0.97	-0.01	-0.27	0.21	0.44	0.53	-1.05	-0.14
121	-1.12	0.53	-1.26	0.80	1.05	-0.29	-1.11	0.62	-0.62	1.40
122	-0.92	0.36	0.14	0.19	0.23	0.14	1.88	-0.18	-0.23	-1.61
123	-1.40	1.53	-0.40	0.87	-0.61	0.38	-1.46	0.56	0.17	0.36
124	-0.81	0.06	0.05	-0.07	0.77	1.60	-1.75	0.63	-0.11	-0.37
125	-0.43	-1.87	1.10	-0.32	1.51	0.03	0.16	-1.34	0.02	1.13
126	-0.61	1.50	-0.81	0.17	-0.24	-1.11	-0.08	2.67	-0.56	-0.92
127	-0.72	-0.24	0.21	0.97	-0.21	-1.67	0.73	1.75	0.06	-0.86
128	-1.27	-0.64	-0.29	0.99	1.21	0.46	-0.44	-0.70	-0.22	0.90
129	-0.98	0.52	-0.12	0.66	-0.08	1.08	-0.62	0.17	-2.39	1.75
130	-0.79	-0.63	0.06	1.10	0.27	0.16	-1.82	0.47	-0.07	1.26

131	-1.34	1.29	-0.09	-0.80	0.95	1.64	1.23	-1.97	-1.16	0.26
132	-0.67	0.59	1.18	0.15	-1.25	-0.93	1.28	0.00	-2.21	1.86
133	-0.98	1.15	-0.42	0.09	0.16	-2.03	0.65	0.90	0.71	-0.23
134	-0.88	0.88	0.45	0.14	-0.60	0.13	1.23	-0.77	-0.57	-0.02
135	-1.00	1.12	-0.32	-0.63	0.82	1.89	-0.60	-0.46	0.26	-1.08
136	-0.67	0.67	0.90	-0.54	-0.35	0.18	0.37	-1.46	1.04	-0.14
137	-1.02	-0.99	1.47	0.18	0.36	-1.22	-1.41	-0.41	1.43	1.61
138	-0.90	0.09	0.62	-0.18	0.36	-0.16	0.31	-0.09	0.86	-0.92
139	-1.51	1.09	0.03	-1.02	1.40	0.83	0.48	0.58	-0.35	-1.54
140	-1.51	-0.65	2.17	0.54	-0.55	2.03	0.62	-1.09	-0.93	-0.63
141	-0.97	0.21	0.12	0.50	0.15	-1.44	0.36	-0.51	0.85	0.76
142	-1.40	0.49	0.04	0.89	-0.01	-1.23	-0.33	-0.31	1.59	0.28
143	-0.93	-0.44	-1.32	0.76	1.94	0.84	-0.53	1.88	-0.80	-1.38
144	-0.91	0.41	0.94	0.58	-1.01	-0.51	0.86	-0.73	0.88	-0.50
145	-1.19	1.31	-0.49	-0.30	0.66	-2.15	0.51	1.81	1.19	-1.36
146	-1.17	0.42	0.35	-0.30	0.70	-0.21	-1.13	0.88	0.54	-0.08
147	-1.55	-0.66	0.36	0.31	1.53	0.03	-0.15	-1.00	0.12	1.00
148	-1.49	0.21	-0.23	1.84	-0.33	-1.22	-0.71	-0.54	0.90	1.58
149	-0.59	0.44	0.09	0.07	0.00	-0.73	-0.69	-0.25	1.44	0.23
150	-1.36	1.22	0.87	-0.02	-0.70	1.90	-0.04	-1.30	-0.26	-0.30
151	-1.57	-0.49	1.08	0.61	0.37	0.05	-0.92	0.13	-1.20	1.95
152	-0.88	-0.78	-1.50	1.43	1.73	-1.15	0.52	0.50	-1.35	1.48
153	-0.49	-1.38	-0.13	1.73	0.28	1.04	0.05	0.62	-0.55	-1.15
154	-1.57	0.56	0.55	0.92	-0.47	-0.28	1.66	-0.76	-0.82	0.20
155	-1.23	-0.88	0.67	0.71	0.73	0.83	-1.12	-0.13	0.22	0.20
156	-0.87	-0.02	0.72	0.77	-0.60	1.30	1.39	-0.84	0.68	-2.53
157	-1.72	0.19	-0.29	1.16	0.65	-0.67	-1.41	1.46	-1.32	1.94
158	-1.46	-0.62	1.46	1.00	-0.38	1.87	0.53	-0.15	-1.24	-1.00
159	-1.54	0.76	-0.35	0.73	0.40	-0.80	0.09	2.25	-0.63	-0.91
160	-1.20	0.34	0.33	0.82	-0.29	0.29	0.50	-0.66	-0.24	0.11

EK 3. 50 Madde 50000 Örneklem Büyüklüğü İçin Silme Belirleme İndeksi Algoritması

Silme (Yanıt Değişirme)

1. Verilerin Üretilmesi

- 5 seçenekli çoktan seçmeli 50 maddenin 25000 öğrenci için yanıt örüntülerini üret (R'da)
- Maddelerin lambda ve zeta değerlerini veri setine aktar
- Yetenek düzeylerine göre öğrencileri eşit sayıda olacak şekilde 5 yetenek düzeyine ayır

2. İyicil Silmelerin Algoritması

a. Kaydırma Silmeleri

- 25000 öğrencinin 100 tanesini her yetenek düzeyinden 20 öğrenci olacak şekilde tesadüfi seç
- Kaydırma silmelerinin sayısını binom dağılımından hesapla ($p=.25$ ve $N=50$ madde sayısı)
- İlk kaydırma maddesinin yerini belirlemek için tesadüfi olarak $U[1, 50 - M_j + 1]$ 'den bir sayı üret. ($M_j =$ kaydırma silmelerinin sayısı)
- Her kaydırma maddesi için, i ve $i-1$ konumundaki yanıtlar karşılaştır
- Eğer farklılarsa, $i-1$ 'deki yanıt i maddesi için silinmiş yanıt (ilk yanıt) olarak kabul et
- Farklı değilse silme olarak kabul etme

b. Dizi sonu silmeleri

- 25000 öğrencinin daha önce seçilmeyen 100 tanesini her yetenek düzeyinden 20 öğrenci olacak şekilde tesadüfi seç
- Testteki son 12 madde (39-50) için tesadüfi başlangıç yanıtını tesadüfi olarak seç
- Silinmiş yanıtı, her yanıtı seçme olasılığı 0.2'ye eşit olacak şekilde tesadüfi olarak tahmin et
- Eğer mevcut son yanıt farklıysa, tahmin ettiğin yanıtı silinmiş yanıt (ilk yanıt) olarak kabul et
- Farklı değilse silme olarak kabul etme

c. Tesadüfi Silmeler

- Kaydırma ve dizi sonu için seçmediğin öğrenciler içinden (24800) tesadüfi silme sayısını binom dağılımından hesapla hesapla ($p=.02$ ve $N=50$ madde sayısı)
- Öğrencilerin hangi maddeler tesadüfi silme yapacağını $U[1,50]$ den tesadüfi sayılar seçerek belirle
- Öğrencinin son yanıtı dışındaki herhangi bir yanıtı tesadüfi olarak seç ve silinmiş yanıt (ilk yanıt) olarak kabul et

3. Hileli Silmelerin Algoritması

- A. 2 ayrı hileli silme türü (sabit silme ve değişken silme) için iyicil silmeleri içeren 25000'lik veri setini 2 defa ayrı ayrı kullan

a. Sabit Silmeler

İyicil silmeleri içeren 25000 öğrencinin veri setini al

1. 5 Silme

- 25000 öğrencinin 500 tanesini her yetenek düzeyinden 100 öğrenci olacak şekilde tesadüfi seç
- Her bir öğrencinin yanlış yanıtladığı 5 maddeyi tesadüfi olarak seç
- Her bir öğrencinin yanlış yanıtladığı 5 maddeyi doğrusu ile değiştir

- d. Seçilen öğrencilerden 5'ten daha az yanlış yanıtı olan öğrenciler varsa bunların tüm yanlış yanıtlarını doğrusu ile değiştir
- e. Sildiğin yanıtı ilk yanıtı olarak kabul et

2. 10 Silme

- a. 25000 öğrencinin daha önce seçilmeyen 500 tanesini her yetenek düzeyinden 100 öğrenci olacak şekilde tesadüfi seç
- b. Her bir öğrencinin yanlış yanıtladığı 10 maddeyi tesadüfi olarak seç
- c. Her bir öğrencinin yanlış yanıtladığı 10 maddeyi doğrusu ile değiştir
- d. Seçilen öğrencilerden 10'dan daha az yanlış yanıtı olan öğrenciler varsa bunların tüm yanlış yanıtlarını doğrusu ile değiştir
- e. Sildiğin yanıtı ilk yanıtı olarak kabul et

3. 15 Silme

- a. 25000 öğrencinin daha önce seçilmeyen 500 tanesini her yetenek düzeyinden 100 öğrenci olacak şekilde tesadüfi seç
- b. Her bir öğrencinin yanlış yanıtladığı 15 maddeyi tesadüfi olarak seç
- c. Her bir öğrencinin yanlış yanıtladığı 15 maddeyi doğrusu ile değiştir
- d. Seçilen öğrencilerden 15'ten daha az yanlış yanıtı olan öğrenciler varsa bunların tüm yanlış yanıtlarını doğrusu ile değiştir
- e. Sildiğin yanıtı ilk yanıtı olarak kabul et

b. Değişken Silme

1. Sadece iyicil silmeleri içeren 25000 öğrencinin veri setini al
2. Yetenek düzeyine göre kesme puanını belirle
3. Kesme puanının altında kalan öğrencilerden 500 tanesini her yetenek düzeyinden 100 öğrenci olacak şekilde tesadüfi seç
Üst yetenek düzeylerindeki öğrenciler seçilemeyecektir. Bu nedenle her yetenek düzeyinden öğrenci seçilemeyebilir
4. Kesme puanının 1 puan üstünü hedef puan olarak belirle
5. Seçilen öğrencilerin puanlarının hedef puana ulaşmasını sağlayacak silme sayısını belirle
6. Her öğrenci için belirlenen sayıda yanlış yanıtı doğru ile değiştir
7. Sildiğin yanıtı ilk yanıtı olarak kabul et

B. Sabit silmeleri içeren 25000'lik veri seti ile değişken silmeleri içeren içeren 25000'lik veri setini birleştirerek 50000'lik yeni veri setini oluştur

C. Silme Belirleme İndeksi analizleri için 50000'lik yeni veri setini kullan.

Analiz

4. SBİ Hesaplanması

- a. Bireylerin silme yapılmış (iyicil ve hileli) maddelerini belirle
- b. Silme yapılmış bireylerin bu maddelere verilen yanıtlarının kodunu 9 olarak değiştir ve ayrı bir dosya olarak kaydet.
- c. Bu datayı MULTILOG'a aktararak yetenek kestirimi yap ($\hat{\theta}_{j[i \in I_{E,j}]}$).
- d. Bu kestirimleri ($\hat{\theta}_{j[i \in I_{E,j}]}$) R'a aktar ve dosya adını $\theta_{\text{nonerased}}$ yap.
- e. Silinmiş maddelerdeki gözlenen değeri (bu maddelerdeki toplam doğru sayısını, yanıt anahtarı ile aynı olan yanıt sayısı) belirle,

$$X_{j,I_{E,j}} = \sum_{i \in I_{E,j}} x_{ij}$$

f. Silinmiş maddelerdeki beklenen değeri hesapla,

$$E(X_{j,I_{E,j}}) = \sum_{i \in I_{E,j}} P(x_{ij} = 1)$$

yani bireyin bu maddelerdeki doğru yanıtı işaretleme olasılığını aşağıdaki formülden hesapla:

$$p_j(x = k | \theta, \alpha, \gamma) = \frac{e^{\gamma_{jk} + \alpha_{jk}\theta}}{\sum_{h=1}^{m_j} e^{\gamma_{jh} + \alpha_{jh}\theta}}$$

- Silinmiş tüm maddeler için bireyin beklenen değerini hesapla.
- Silinmiş tüm maddeler için hesaplanan beklenen değerleri topla.

g. Silinmiş maddeler için standart hatayı hesapla

$$SE(X_{j,I_{E,j}}) = \sqrt{\sum_{i \in I_{E,j}} P(x_{ij} = 1)[1 - P(x_{ij} = 1)]}$$

h. Silme yapılmış tüm bireyler için süreklilik düzeltmesi uygulanmayan Silme Belirleme İndeksini (SBİ) Hesapla

$$SBİ = \frac{X_{j,I_{E,j}} - E(X_{j,I_{E,j}}) + C}{SE(X_{j,I_{E,j}})}$$

i. .00001, .0001, .0005, .001, .005, .01 ve .05 manidarlık düzeyleri için süreklilik düzeltmesi uygulanmayan SBİ değerini tablo z değeri ile karşılaştır.

j. Süreklilik düzeltmesi uygulanmayan SBİ değeri, tablo z değerini aşan bireyleri tahrifatlı olarak işaretle.

k. Silme yapılmış tüm bireyler için Süreklilik düzeltmesi uygulanan (c= -0.5) SBİ Hesapla

$$SBİ = \frac{X_{j,I_{E,j}} - E(X_{j,I_{E,j}}) + C}{SE(X_{j,I_{E,j}})}$$

l. .00001, .0001, .0005, .001, .005, .01 ve .05 manidarlık düzeyleri için süreklilik düzeltmesi uygulanan SBİ değerini tablo z değeri ile karşılaştır.

m. Süreklilik düzeltmesi uygulanan SBİ değeri, tablo z değerini aşan bireyleri tahrifatlı olarak işaretle.

5. Değerlendirici Ölçmeler

- a. Her bir yetenek düzeyi grubu ve toplam için aşağıdaki sayıları belirle;
- Gerçekte tahrifatlı olan ve SBİ ile tahrifatlı olarak belirlenen birey sayısı (X1)
 - Gerçekte Tahrifatlı olmayan ve SBİ ile tahrifatlı olarak belirlenmeyen birey sayısı (Y1)
 - Gerçekte Tahrifatlı olan ve SBİ tarafından tahrifatlı olarak belirlenmeyen birey sayısı (X0)
 - Gerçekte tahrifatlı olmayan ve SBİ tarafından tahrifatlı olarak belirlenen birey sayısı (Y0)
 - Gerçek tahrifatlı birey sayısı (X1 + X0)

- Gerçek tahrifatlı olmayan birey sayısı (Y1 + Y0)

		SBİ İle Belirlenen Durum		
		Tahrifatlı	Tahrifatsız	Toplam
Gerçek Durum	Tahrifatlı	X1	X0	X1+X0
	Tahrifatsız	Y0	Y1	Y1+Y0
	Toplam	X1+Y0	X0+Y1	X1+X0+Y1+Y0

b. I. Tip Hata Oranı

- I. Tip Hata Oranı = Gerçekte tahrifatlı olmayan ve SBİ tarafından tahrifatlı olarak belirlenen birey sayısı (Y0) / Tahrifatlı olmayan birey sayısı (Y1 + Y0)
- SBİ'nin I. Tip Hata oranını 7 alfa seviyesinde (.00001'den .05'e kadar) değerlendirir
- SBİ'nin I. Tip Hata oranını her yetenek düzeyi için 7 alfa seviyesinde değerlendirir

c. Testin Gücü

- Testin Gücü = Gerçekte Tahrifatlı olan ve SBİ ile tahrifatlı olarak belirlenen birey sayısı (X1) / Tahrifatlı birey sayısı (X1 + X0)
- SBİ'nin Gücünü 7 alfa seviyesinde (.00001'den .05'e kadar) değerlendirir
- SBİ'nin Gücünü her yetenek düzeyi için 7 alfa seviyesinde değerlendirir
- SBİ'nin Gücünü her yetenek düzeyi için hileli silme oranlarına göre 7 alfa seviyesinde değerlendirir

d. Bias

- Test alanları tahrifatsız ve tahrifatlı olarak belirle
- Tahrifatsız test alanlar için Bias'ı aşağıdaki formüle göre hesapla

$$Bias_{Tahrifatsız} = \frac{\sum[\theta_{\text{silinmemiş}} - \theta_{\text{tüm}}]}{\text{Tahrifatsız Test Alan Sayısı}}$$

Test alanların silme yapılmamış maddeleri üzerinden kestirilen yetenek düzeyi ile tüm maddeler üzerinden kestirilen yetenek düzeylerinin farkının toplamını al ve tahrifatsız test alan sayısına böl

- Tahrifatlı test alanlar için Bias'ı aşağıdaki formüllere göre hesapla
- Yalnızca silinmemiş maddeler üzerinden Bias hesaplamak için:

$$Bias_{Tahrifatlı_silinmemiş} = \frac{\sum[\theta_{\text{silinmemiş}} - \theta_{\text{tüm}}]}{\text{Tahrifatlı Test Alan Sayısı}}$$

Test alanların silme yapılmamış maddeleri üzerinden kestirilen yetenek düzeyi ile tüm maddeler üzerinden kestirilen yetenek düzeylerinin farkının toplamını al ve tahrifatlı test alan sayısına böl

- Tüm maddeler üzerinden Bias hesaplamak için:

$$Bias_{Tahrifatlı_tüm\ maddeler} = \frac{\sum[\theta_{\text{iyicilli}} - \theta_{\text{tüm}}]}{\text{Tahrifatlı Test Alan Sayısı}}$$

Test alanların iyicil silmeleri de içeren maddeleri üzerinden kestirilen yetenek düzeyi ile tüm maddeler üzerinden kestirilen yetenek düzeylerinin farkının toplamını al ve tahrifatlı test alan sayısına böl

e. RMSE

- Test alanları tahrifatsız ve tahrifatlı olarak belirle
- Tahrifatsız test alanlar için RMSE'yi aşağıdaki formüle göre hesapla

$$- RMSE_{Tahrifatsız} = \sqrt{\frac{\sum[(\hat{e}_{j[i \notin I_{E,j}]silinmemiş} - \theta_{tüm})^2]}{\text{Tahrifatsız Test Alan Sayısı}}}$$

Test alanların silme yapılmamış maddeleri üzerinden kestirilen hatalar ile tüm maddeler üzerinden kestirilen yetenek düzeylerinin farkının karesinin toplamını al, tahrifatsız test alan sayısına böl, sonucun karekökünü al

- Tahrifatlı test alanlar için RMSE'yi aşağıdaki formüllere göre hesapla:
- Yalnızca silinmemiş maddeler üzerinden RMSE hesaplamak için:

$$- RMSE_{Tahrifatlı_silinmemiş} = \sqrt{\frac{\sum[(\hat{e}_{j[i \notin I_{E,j}]silinmemiş} - \theta_{tüm})^2]}{\text{Tahrifatlı Test Alan Sayısı}}}$$

Test alanların silme yapılmamış maddeleri üzerinden kestirilen hatalar ile tüm maddeler üzerinden kestirilen yetenek düzeylerinin farkının karesinin toplamını al, tahrifatlı test alan sayısına böl, sonucun karekökünü al

- Tüm maddeler üzerinden RMSE hesaplamak için:

$$- RMSE_{Tahrifatlı_tüm\ maddeler} = \sqrt{\frac{\sum[(\hat{e}_{j[iyicilli]} - \theta_{tüm})^2]}{\text{Tahrifatlı Test Alan Sayısı}}}$$

Test alanların iyicil silmeleri de içeren maddeleri üzerinden kestirilen hataları ile tüm maddeler üzerinden kestirilen yetenek düzeylerinin farkının karesinin toplamını al, tahrifatlı test alan sayısına böl, sonucun karekökünü al

EK 4. Silinmemiş Maddeler Üzerinden Yetenek Kestirimine Ait Yanlılık ve RMSE Bulguları

Tablo 28

50 ve 160 Madde Üzerinden 500 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde									160 Madde								
Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar				
Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
					Yanlılık	RMSE	Yanlılık	RMSE						Yanlılık	RMSE	Yanlılık	RMSE
0	140	0.000	0.935	0	—	—	—	—	0	16	0.000	0.908	0	—	—	—	—
1	150	0.000	0.898	1	0.017	0.237	0.065	0.235	1	54	0.000	0.905	0	—	—	—	—
2	74	0.000	0.893	1	0.027	0.312	0.105	0.307	2	86	0.001	0.886	0	—	—	—	—
3	29	-0.002	0.878	1	0.043	0.364	0.134	0.357	3	90	0.000	0.890	0	—	—	—	—
4	8	-0.010	0.895	1	0.044	0.444	0.186	0.435	4	71	0.000	0.892	0	—	—	—	—
5	3	-0.021	0.974	8	0.104	0.875	0.343	0.867	5	45	0.001	0.902	1	0.012	0.177	0.051	0.175
6	3	-0.006	0.723	9	0.091	0.873	0.343	0.864	6	27	0.001	0.867	1	0.019	0.207	0.068	0.205
7	3	-0.032	0.833	4	0.103	0.842	0.365	0.832	7	12	-0.002	0.922	1	0.024	0.208	0.075	0.205
8	3	-0.022	0.967	2	0.081	0.966	0.435	0.948	8	6	0.000	0.799	1	0.033	0.248	0.100	0.245
9	4	-0.008	0.735	2	0.092	1.169	0.573	1.142	9	2	-0.003	0.776	0	—	—	—	—
10	4	-0.023	0.716	9	0.221	0.974	0.707	0.956	10	2	0.018	0.778	1	0.025	0.271	0.102	0.267
11	2	-0.063	0.635	8	0.199	0.976	0.708	0.954	11	1	-0.002	1.113	0	—	—	—	—
12	1	-0.080	0.852	4	0.247	0.940	0.726	0.922	12	0	—	—	0	—	—	—	—
13	1	-0.058	0.681	1	0.167	0.782	0.661	0.765	13	0	—	—	0	—	—	—	—
14	1	-0.018	0.662	1	0.508	0.978	0.770	0.974	14	0	—	—	0	—	—	—	—
≥15	1	-0.027	0.943	22	0.359	0.878	1.038	0.854	15	0	—	—	0	—	—	—	—

(Devam Ediyor)

Tablo 28 (Devam)

50 ve 160 Madde Üzerinden 500 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

160 Madde									
Silme Sayısı	Tahrifatsız Test Alanlar			Tahrifatlı Test Alanlar					
	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler	RMSE	Tüm Maddeler	Yanlılık	RMSE
16	0	—	—	1	0.112	0.848	0.368	0.844	
17	0	—	—	3	0.101	0.753	0.347	0.749	
18	0	—	—	5	0.109	0.893	0.369	0.888	
19	0	—	—	5	0.101	0.766	0.352	0.762	
20	0	—	—	3	0.114	0.792	0.357	0.789	
21	0	—	—	3	0.073	0.802	0.371	0.796	
22	1	-0.054	0.427	1	0.112	0.828	0.384	0.824	
23	0	—	—	0	—	—	—	—	
24	1	-0.059	1.119	0	—	—	—	—	
25	1	0.001	0.814	0	—	—	—	—	
26	1	-0.052	0.894	0	—	—	—	—	
27	1	-0.025	0.909	0	—	—	—	—	
28	1	-0.047	0.932	0	—	—	—	—	
29	1	-0.020	0.606	0	—	—	—	—	
30	1	-0.029	0.639	0	—	—	—	—	
31	1	-0.003	0.694	0	—	—	—	—	
32	1	0.002	0.532	1	0.217	0.839	0.732	0.830	
33	1	-0.025	0.680	4	0.253	0.881	0.743	0.872	

(Devam Ediyor)

Tablo 28 (Devam)

50 ve 160 Madde Üzerinden 500 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

Silme Sayısı	160 Madde							
	Tahrifatsız Test Alanlar			Tahrifatlı Test Alanlar				
	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler	RMSE	Tüm Maddeler	Yanlılık
34	1	-0.004	0.664	5	0.244	0.880	0.730	0.872
35	1	-0.021	0.829	4	0.257	0.877	0.742	0.870
36	1	0.005	0.455	3	0.277	0.783	0.751	0.777
37	0	—	—	1	0.386	0.926	0.846	0.920
38	1	-0.025	0.843	1	0.447	0.839	0.867	0.835
39	1	-0.033	0.865	0	—	—	—	—
40	0	—	—	0	—	—	—	—
41	0	—	—	0	—	—	—	—
42	0	—	—	0	—	—	—	—
43	0	—	—	0	—	—	—	—
44	0	—	—	0	—	—	—	—
45	0	—	—	0	—	—	—	—
46	0	—	—	0	—	—	—	—
47	0	—	—	0	—	—	—	—
>= 48	0	—	—	21	0.438	0.861	1.124	0.850

Tablo 29

50 ve 160 Madde Üzerinden 500 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde										160 Madde									
Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar				
Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
						Yanlılık	RMSE	Yanlılık	RMSE							Yanlılık	RMSE	Yanlılık	RMSE
0-1	1	58	0.001	1.425	0	—	—	—	—	0-8	1	80	0.001	1.305	0	—	—	—	—
	2	58	0.000	0.726	0	—	—	—	—		2	79	0.001	0.579	0	—	—	—	—
	3	55	0.000	0.241	1	0.017	0.237	0.065	0.235		3	80	0.000	0.123	2	0.017	0.186	0.055	0.184
	4	59	0.000	0.389	0	—	—	—	—		4	84	0.000	0.487	0	—	—	—	—
	5	60	-0.001	1.174	0	—	—	—	—		5	83	-0.001	1.306	0	—	—	—	—
2-4	1	21	-0.002	1.431	0	—	—	—	—	9-17	1	1	-0.014	1.441	1	0.029	1.257	0.000	1.244
	2	21	0.004	0.727	0	—	—	—	—		2	1	0.007	0.616	3	0.046	0.501	0.250	0.494
	3	23	-0.007	0.250	3	0.031	0.338	0.114	0.332		3	1	0.007	0.093	2	0.046	0.171	0.214	0.168
	4	23	-0.002	0.369	0	—	—	—	—		4	1	0.015	0.368	1	0.145	0.511	0.327	0.515
	5	22	0.001	1.144	0	—	—	—	—		5	1	-0.009	1.121	1	0.261	1.380	0.401	1.383
5-7	1	2	-0.031	1.452	4	0.029	1.465	0.416	1.440	18-26	1	1	-0.069	1.278	5	0.049	1.186	0.456	1.172
	2	2	-0.025	0.768	6	0.055	0.734	0.341	0.719		2	0	—	—	5	0.060	0.636	0.356	0.627
	3	2	0.001	0.238	4	0.079	0.266	0.308	0.255		3	1	-0.032	0.177	3	0.085	0.129	0.319	0.127
	4	2	-0.010	0.339	3	0.137	0.375	0.317	0.384		4	1	0.018	0.439	3	0.133	0.469	0.322	0.475
	5	2	-0.034	1.039	4	0.225	1.107	0.368	1.114		5	1	0.026	1.289	3	0.238	1.299	0.381	1.304

(Devam Ediyor)

Tablo 29 (Devam)

50 ve 160 Madde Üzerinden 500 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde										160 Madde									
Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar				
Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
						Yanlılık	RMSE	Yanlılık	RMSE							Yanlılık	RMSE	Yanlılık	RMSE
8-10	1	2	-0.045	1.369	5	0.065	1.368	0.693	1.331	27-35	1	1	-0.075	1.153	5	0.050	1.381	0.770	1.356
	2	2	-0.005	0.699	2	0.085	0.789	0.577	0.764		2	2	-0.048	0.618	3	0.143	0.550	0.686	0.535
	3	2	-0.004	0.245	2	0.144	0.236	0.547	0.217		3	2	-0.012	0.117	3	0.188	0.129	0.630	0.122
	4	2	-0.009	0.307	2	0.286	0.359	0.606	0.373		4	2	0.015	0.463	3	0.297	0.458	0.667	0.467
	5	2	-0.016	1.109	2	0.522	1.125	0.769	1.136		5	2	0.005	1.237	3	0.561	1.197	0.844	1.204
11-13	1	1	-0.107	1.532	3	0.046	1.535	0.803	1.486	36-44	1	0	—	—	1	0.059	1.249	0.778	1.222
	2	1	-0.069	0.632	2	0.097	0.762	0.653	0.731		2	1	-0.027	0.573	1	0.161	0.620	0.688	0.603
	3	1	0.006	0.241	3	0.180	0.235	0.614	0.213		3	1	-0.008	0.129	1	0.183	0.110	0.553	0.105
	4	1	-0.059	0.300	2	0.313	0.361	0.667	0.380		4	1	0.026	0.480	1	0.253	0.363	0.619	0.376
	5	1	-0.092	1.126	2	0.563	1.055	0.824	1.071		5	0	0.040	1.500	1	0.962	1.395	1.225	1.404
≥14	1	0	—	—	5	0.080	1.435	1.077	1.375	≥45	1	0	—	—	4	0.097	1.361	1.170	1.322
	2	0	—	—	4	0.179	0.789	0.972	0.745		2	0	—	—	4	0.241	0.577	1.031	0.554
	3	0	—	—	4	0.296	0.276	0.936	0.245		3	0	—	—	4	0.350	0.128	0.986	0.119
	4	1	0.008	0.332	4	0.430	0.295	0.948	0.317		4	0	—	—	4	0.524	0.471	1.038	0.485
	5	1	-0.018	1.532	4	0.885	0.984	1.262	1.007		5	0	—	—	4	1.069	1.149	1.459	1.160

Tablo 30

50 ve 160 Madde Üzerinden 5000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde									160 Madde								
Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar				
Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
					Yanlılık	RMSE	Yanlılık	RMSE						Yanlılık	RMSE	Yanlılık	RMSE
0	1665	0.000	1.005	0	—	—	—	—	0	182	0.000	0.966	0	—	—	—	—
1	1722	0.000	1.002	2	0.017	0.202	0.063	0.200	1	592	0.000	0.965	0	—	—	—	—
2	864	0.000	1.008	4	0.019	0.260	0.094	0.255	2	971	0.000	0.953	1	0.007	0.144	0.028	0.143
3	276	0.000	1.012	5	0.027	0.307	0.126	0.300	3	1033	0.000	0.961	1	0.008	0.148	0.030	0.147
4	70	-0.003	1.001	5	0.053	0.533	0.203	0.525	4	827	0.000	0.956	2	0.007	0.169	0.038	0.167
5	16	-0.012	0.969	43	0.101	0.962	0.348	0.953	5	536	0.000	0.961	2	0.007	0.182	0.050	0.180
6	5	-0.018	1.072	40	0.107	0.987	0.363	0.976	6	278	-0.001	0.973	2	0.017	0.215	0.071	0.213
7	4	-0.061	0.783	23	0.102	0.923	0.372	0.912	7	123	0.001	0.949	2	0.019	0.236	0.079	0.233
8	4	-0.024	0.825	11	0.112	1.002	0.469	0.985	8	52	0.000	0.966	2	0.031	0.272	0.101	0.269
9	3	-0.010	0.999	6	0.140	1.015	0.557	0.995	9	17	0.001	1.001	3	0.027	0.298	0.119	0.294
10	3	-0.047	0.811	52	0.217	1.060	0.726	1.039	10	6	0.002	0.905	2	0.023	0.337	0.137	0.333
11	3	-0.014	0.999	44	0.229	1.093	0.749	1.071	11	3	0.004	0.688	3	0.033	0.359	0.158	0.354
12	2	-0.061	0.816	20	0.216	1.246	0.801	1.214	12	0	—	—	2	0.030	0.379	0.170	0.374
13	2	-0.020	0.771	6	0.307	1.338	0.888	1.309	13	0	—	—	1	0.046	0.446	0.210	0.440
14	1	-0.016	1.005	2	0.535	1.446	1.113	1.419	14	0	—	—	2	0.038	0.459	0.226	0.453
≥15	1	-0.066	0.687	98	0.380	0.960	1.093	0.930	15	0	—	—	1	0.052	0.522	0.255	0.515

(Devam Ediyor)

Tablo 30 (Devam)

50 ve 160 Madde Üzerinden 5000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

160 Madde									
Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar					
Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		RMSE
					Yanlılık	RMSE	Yanlılık	RMSE	
16	0	—	—	6	0.103	0.877	0.350	0.873	
17	0	—	—	18	0.101	0.900	0.357	0.895	
18	0	—	—	24	0.100	0.908	0.362	0.902	
19	0	—	—	23	0.110	0.939	0.364	0.934	
20	0	—	—	18	0.108	0.996	0.373	0.991	
21	0	—	—	11	0.107	0.894	0.371	0.889	
22	1	-0.065	1.254	6	0.092	0.873	0.380	0.866	
23	1	-0.007	0.891	3	0.106	0.980	0.443	0.970	
24	0	—	—	2	0.064	0.857	0.446	0.846	
25	1	-0.041	0.990	2	0.073	0.999	0.519	0.985	
26	1	-0.044	0.943	2	0.071	1.059	0.543	1.044	
27	1	-0.042	0.954	2	0.065	1.094	0.585	1.077	
28	1	0.003	0.831	1	0.080	1.167	0.630	1.149	
29	1	-0.019	0.691	2	0.061	1.254	0.669	1.234	
30	1	0.018	0.667	1	0.066	1.356	0.724	1.334	
31	1	-0.044	0.741	1	0.063	1.379	0.728	1.355	
32	1	-0.053	0.699	9	0.231	1.066	0.756	1.054	
33	1	-0.029	0.583	21	0.233	0.980	0.748	0.969	

(Devam Ediyor)

Tablo 30 (Devam)

50 ve 160 Madde Üzerinden 5000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

160 Madde									
Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar				
Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		RMSE
					Yanlılık	RMSE	Yanlılık	RMSE	
34	1	-0.030	0.624	26	0.248	0.998	0.759	0.987	
35	2	-0.015	0.752	24	0.247	1.018	0.770	1.005	
36	1	-0.016	0.680	15	0.261	1.063	0.776	1.052	
37	1	-0.020	0.665	8	0.257	1.064	0.781	1.052	
38	1	-0.010	0.315	4	0.314	1.047	0.827	1.037	
39	0	—	—	2	0.227	0.934	0.754	0.921	
40	0	—	—	1	0.191	1.319	0.870	1.300	
41	0	—	—	0	—	—	—	—	
42	0	—	—	0	—	—	—	—	
43	0	—	—	0	—	—	—	—	
44	0	—	—	0	—	—	—	—	
45	0	—	—	0	—	—	—	—	
46	0	—	—	0	—	—	—	—	
47	0	—	—	0	—	—	—	—	
>= 48	0	—	—	100	0.441	0.924	1.158	0.911	

Tablo 31

50 ve 160 Madde Üzerinden 5000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde										160 Madde									
Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar				
Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
						Yanlılık	RMSE	Yanlılık	RMSE							Yanlılık	RMSE	Yanlılık	RMSE
0-1	1	666	0.000	1.571	0	—	—	—	—	0-8	1	911	0.001	1.467	0	—	—	—	—
	2	673	0.001	0.823	0	—	—	—	—		2	911	0.000	0.686	0	—	—	—	—
	3	668	0.000	0.291	2	0.017	0.202	0.063	0.200		3	911	0.000	0.179	12	0.016	0.215	0.064	0.212
	4	690	0.000	0.377	0	—	—	—	—		4	931	0.000	0.548	0	—	—	—	—
	5	688	-0.001	1.296	0	—	—	—	—		5	930	-0.001	1.286	0	—	—	—	—
2-4	1	245	0.003	1.576	0	—	—	—	—	9-17	1	5	-0.001	1.551	4	0.028	1.452	0.000	1.436
	2	237	0.002	0.825	0	—	—	—	—		2	5	0.001	0.689	14	0.050	0.562	0.282	0.555
	3	242	0.000	0.296	13	0.027	0.323	0.135	0.317		3	5	0.000	0.171	12	0.048	0.283	0.199	0.279
	4	241	-0.001	0.367	0	—	—	—	—		4	6	-0.003	0.519	5	0.134	0.513	0.323	0.518
	5	244	-0.004	1.293	0	—	—	—	—		5	6	-0.003	1.187	5	0.222	1.210	0.368	1.214
5-7	1	6	-0.040	1.592	19	0.028	1.593	0.434	1.565	18-26	1	2	-0.104	1.564	20	0.039	1.403	0.472	1.386
	2	5	-0.006	0.806	27	0.059	0.791	0.360	0.775		2	1	-0.039	0.785	26	0.058	0.744	0.379	0.734
	3	6	-0.031	0.290	22	0.076	0.331	0.305	0.320		3	1	-0.013	0.191	16	0.089	0.181	0.323	0.177
	4	5	-0.008	0.322	19	0.122	0.362	0.310	0.370		4	1	-0.002	0.462	15	0.137	0.539	0.323	0.544
	5	4	-0.008	1.160	19	0.255	1.245	0.392	1.252		5	1	0.010	1.225	15	0.240	1.293	0.382	1.298

(Devam Ediyor)

Tablo 31 (Devam)

50 ve 160 Madde Üzerinden 5000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde										160 Madde									
Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar				
Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
						Yanlılık	RMSE	Yanlılık	RMSE							Yanlılık	RMSE	Yanlılık	RMSE
8-10	1	2	-0.040	1.638	17	0.045	1.514	0.752	1.472	27-35	1	1	-0.131	1.327	28	0.062	1.454	0.807	1.427
	2	3	-0.022	0.837	21	0.093	0.883	0.587	0.857		2	2	-0.059	0.717	15	0.127	0.707	0.703	0.690
	3	2	0.026	0.302	10	0.159	0.322	0.586	0.303		3	2	-0.006	0.138	15	0.204	0.191	0.658	0.185
	4	2	-0.052	0.329	10	0.258	0.367	0.598	0.380		4	3	0.017	0.476	14	0.311	0.528	0.670	0.536
	5	2	-0.071	1.223	10	0.623	1.300	0.858	1.310		5	2	0.003	1.240	14	0.596	1.244	0.875	1.251
11-13	1	1	-0.045	1.599	23	0.064	1.649	0.871	1.596	36-44	1	0	—	—	8	0.054	1.679	0.926	1.643
	2	1	-0.061	0.853	11	0.101	0.862	0.673	0.830		2	1	-0.042	0.668	5	0.127	0.704	0.703	0.686
	3	2	0.014	0.308	11	0.178	0.303	0.631	0.281		3	1	-0.020	0.159	5	0.200	0.170	0.655	0.164
	4	1	0.019	0.235	11	0.274	0.332	0.638	0.348		4	1	0.001	0.438	6	0.305	0.512	0.659	0.523
	5	2	-0.053	1.249	13	0.674	1.248	0.944	1.261		5	0	—	—	6	0.727	1.313	1.004	1.322
≥14	1	0	—	—	21	0.091	1.646	1.203	1.577	≥ 45	1	0	—	—	20	0.112	1.487	1.247	1.447
	2	0	—	—	20	0.181	0.870	1.008	0.827		2	0	—	—	20	0.211	0.719	1.053	0.694
	3	0	—	—	20	0.306	0.328	0.963	0.297		3	0	—	—	20	0.347	0.195	1.011	0.184
	4	0	—	—	20	0.469	0.322	0.993	0.342		4	0	—	—	20	0.540	0.504	1.063	0.516
	5	1	-0.048	1.343	18	0.956	1.009	1.331	1.028		5	0	—	—	19	1.043	1.139	1.448	1.149

Tablo 32

50 ve 160 Madde Üzerinden 50000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

Silme Sayısı	50 Madde								160 Madde											
	Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar				Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar							
	Test Alanlar		Yanlılık		RMSE		Yanlılık		RMSE		Test Alanlar		Yanlılık		RMSE		Yanlılık		RMSE	
	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Yanlılık	RMSE	Yanlılık	RMSE	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık	RMSE		
0	16777	0.000	1.054	0	—	—	—	—	0	1829	0.000	0.990	0	—	—	—	—			
1	17102	0.000	1.053	19	0.015	0.199	0.061	0.197	1	5955	0.000	0.988	1	0.005	0.132	0.020	0.132			
2	8577	-0.001	1.049	36	0.024	0.253	0.095	0.249	2	9574	0.000	0.990	6	0.008	0.146	0.027	0.145			
3	2818	0.000	1.053	44	0.031	0.324	0.140	0.318	3	10419	0.000	0.991	12	0.007	0.157	0.032	0.156			
4	696	-0.002	1.055	50	0.043	0.443	0.198	0.434	4	8339	0.000	0.990	19	0.009	0.175	0.041	0.173			
5	156	-0.005	1.056	442	0.105	1.015	0.351	1.006	5	5260	0.000	0.991	22	0.012	0.193	0.051	0.191			
6	49	-0.022	1.044	403	0.109	1.020	0.362	1.011	6	2787	0.000	0.993	29	0.014	0.216	0.065	0.214			
7	39	-0.025	1.101	215	0.113	1.007	0.380	0.997	7	1272	0.000	0.996	27	0.018	0.244	0.081	0.241			
8	40	-0.039	1.048	101	0.137	1.054	0.462	1.042	8	511	0.000	0.995	27	0.021	0.272	0.100	0.269			
9	37	-0.040	0.987	76	0.161	1.149	0.612	1.129	9	177	0.001	0.998	25	0.026	0.303	0.120	0.299			
10	32	-0.043	1.024	520	0.237	1.068	0.735	1.049	10	66	0.001	0.997	24	0.026	0.324	0.132	0.320			
11	26	-0.042	1.049	438	0.228	1.154	0.762	1.130	11	18	0.004	1.021	17	0.029	0.357	0.153	0.352			
12	19	-0.039	0.983	200	0.239	1.267	0.810	1.237	12	5	0.000	1.069	16	0.039	0.403	0.184	0.398			
13	13	-0.060	0.892	68	0.326	1.385	0.914	1.353	13	1	0.039	0.602	15	0.038	0.442	0.206	0.437			
14	8	-0.040	0.767	24	0.677	1.476	1.203	1.453	14	0	—	—	13	0.042	0.491	0.232	0.485			
≥15	10	-0.077	0.851	963	0.392	0.988	1.108	0.959	15	0	—	—	13	0.046	0.532	0.258	0.525			

(Devam Ediyor)

Tablo 32 (Devam)

50 ve 160 Madde Üzerinden 50000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

160 Madde									
	Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar				
					Silinmemiş Maddeler		Tüm Maddeler		
Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık
16	1	-0.076	1.621	67	0.103	0.907	0.350	0.903	
17	1	-0.081	1.687	172	0.109	0.956	0.362	0.952	
18	1	-0.048	1.213	249	0.111	0.979	0.367	0.974	
19	2	-0.089	1.575	238	0.112	0.966	0.367	0.961	
20	2	-0.073	1.423	176	0.111	0.973	0.369	0.968	
21	3	-0.056	1.165	109	0.105	0.961	0.375	0.956	
22	5	-0.041	1.212	57	0.102	0.956	0.388	0.950	
23	5	-0.027	1.167	34	0.090	0.950	0.417	0.942	
24	6	-0.020	1.171	24	0.079	0.968	0.469	0.956	
25	10	-0.030	1.150	20	0.071	1.032	0.520	1.018	
26	10	-0.027	1.094	20	0.074	1.079	0.564	1.063	
27	11	-0.024	0.921	18	0.073	1.149	0.602	1.132	
28	14	-0.012	0.935	17	0.068	1.210	0.635	1.192	
29	12	-0.022	0.945	17	0.065	1.283	0.672	1.262	
30	14	-0.016	0.867	18	0.060	1.348	0.707	1.326	
31	12	-0.009	0.862	17	0.060	1.445	0.756	1.421	
32	13	-0.017	0.876	91	0.237	1.100	0.766	1.087	
33	11	-0.017	0.811	208	0.259	1.030	0.763	1.020	

(Devam Ediyor)

Tablo 32 (Devam)

50 ve 160 Madde Üzerinden 50000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

Silme Sayısı	160 Madde							
	Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar			
	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
					Yanlılık	RMSE	Yanlılık	RMSE
34	11	-0.027	0.824	270	0.263	1.029	0.766	1.019
35	9	-0.013	0.787	223	0.256	1.046	0.770	1.035
36	7	-0.006	0.729	151	0.265	1.070	0.778	1.059
37	4	0.019	0.737	77	0.273	1.090	0.788	1.078
38	5	0.004	0.724	35	0.280	1.114	0.797	1.102
39	4	-0.048	0.634	13	0.293	1.123	0.813	1.110
40	3	-0.016	0.483	5	0.561	1.299	1.050	1.290
41	3	0.006	0.785	3	1.201	1.435	1.576	1.432
42	2	-0.015	0.479	2	1.368	1.448	1.689	1.450
43	1	0.043	0.574	2	1.652	1.495	1.928	1.501
44	0	—	—	2	1.912	1.545	2.192	1.552
45	0	—	—	3	1.880	1.554	2.193	1.562
46	0	—	—	3	1.836	1.500	2.152	1.506
47	0	—	—	3	1.956	1.489	2.263	1.496
>= 48	0	—	—	991	0.4821	0.9467	1.1886	0.9345

Tablo 33

50 ve 160 Madde Üzerinden 50000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde										160 Madde									
Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar				
Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
						Yanlılık	RMSE	Yanlılık	RMSE							Yanlılık	RMSE	Yanlılık	RMSE
0-1	1	6727	0.001	1.589	0	—	—	—	—	0-8	1	9108	0.000	1.433	0	—	—	—	—
	2	6710	0.000	0.869	0	—	—	—	—		2	9113	0.000	0.679	1	0.039	0.371	0.164	0.368
	3	6709	0.000	0.298	19	0.015	0.199	0.061	0.197		3	9111	0.000	0.180	142	0.014	0.216	0.064	0.214
	4	6848	0.000	0.461	0	—	—	—	—		4	9309	0.000	0.619	0	—	—	—	—
	5	6886	-0.001	1.399	0	—	—	—	—		5	9307	0.000	1.403	0	—	—	—	—
2-4	1	2391	0.003	1.592	0	—	—	—	—	9-17	1	58	0.001	1.467	42	0.031	1.424	0.000	1.409
	2	2405	0.001	0.873	0	—	—	—	—		2	51	0.003	0.666	137	0.050	0.571	0.277	0.564
	3	2404	0.000	0.302	130	0.033	0.345	0.149	0.338		3	52	-0.001	0.177	99	0.051	0.256	0.206	0.253
	4	2465	-0.002	0.454	0	—	—	—	—		4	53	0.001	0.611	43	0.141	0.611	0.323	0.616
	5	2425	-0.005	1.395	0	—	—	—	—		5	56	-0.002	1.399	42	0.251	1.376	0.392	1.379
5-7	1	53	-0.013	1.597	186	0.023	1.594	0.430	1.567	18-26	1	17	-0.106	1.552	197	0.038	1.380	0.468	1.364
	2	48	-0.009	0.873	261	0.054	0.830	0.359	0.814		2	7	-0.018	0.679	258	0.059	0.731	0.377	0.720
	3	48	-0.009	0.302	234	0.071	0.346	0.300	0.335		3	5	0.005	0.165	157	0.092	0.182	0.325	0.179
	4	47	-0.009	0.438	186	0.129	0.446	0.311	0.454		4	6	0.018	0.605	157	0.142	0.606	0.324	0.612
	5	48	-0.017	1.383	193	0.287	1.396	0.420	1.402		5	10	0.009	1.428	157	0.253	1.389	0.394	1.393

(Devam Ediyor)

Tablo 33 (Devam)

50 ve 160 Madde Üzerinden 50000 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde										160 Madde											
Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar						
Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler			Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		
						Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık							RMSE	Yanlılık	RMSE	Yanlılık	RMSE
8-10	1	21	-0.063	1.545	156	0.054	1.521	0.761	1.478	27-35	1	15	-0.117	1.312	288	0.062	1.452	0.803	1.425		
	2	25	-0.043	0.890	217	0.084	0.937	0.598	0.910	2	24	-0.039	0.683	150	0.130	0.694	0.702	0.678			
	3	22	-0.024	0.305	106	0.164	0.327	0.589	0.309	3	23	0.001	0.188	148	0.208	0.186	0.657	0.181			
	4	20	-0.035	0.429	102	0.278	0.434	0.606	0.447	4	25	0.015	0.592	147	0.325	0.600	0.676	0.608			
	5	21	-0.041	1.363	116	0.661	1.392	0.885	1.402	5	22	0.020	1.395	146	0.638	1.354	0.907	1.360			
11-13	1	8	-0.068	1.569	247	0.053	1.663	0.869	1.609	36-44	1	2	-0.136	1.201	70	0.056	1.620	0.902	1.585		
	2	10	-0.056	0.866	118	0.102	0.911	0.681	0.879	2	5	-0.035	0.700	52	0.127	0.700	0.699	0.682			
	3	12	-0.020	0.308	108	0.177	0.327	0.632	0.305	3	9	-0.011	0.179	53	0.207	0.184	0.651	0.179			
	4	14	-0.032	0.447	108	0.298	0.433	0.646	0.449	4	7	0.030	0.569	53	0.325	0.596	0.675	0.606			
	5	15	-0.053	1.403	126	0.747	1.331	0.999	1.345	5	5	0.038	1.261	63	0.825	1.412	1.092	1.420			
≥14	1	1	-0.152	1.389	211	0.090	1.645	1.196	1.577	≥= 45	1	0	—	—	203	0.112	1.458	1.233	1.418		
	2	3	-0.080	0.869	204	0.177	0.906	1.020	0.862	2	0	—	—	202	0.222	0.704	1.056	0.680			
	3	5	-0.027	0.315	203	0.304	0.336	0.964	0.306	3	0	—	—	201	0.355	0.189	1.008	0.182			
	4	5	-0.048	0.403	204	0.510	0.404	1.012	0.424	4	0	—	—	201	0.572	0.583	1.079	0.595			
	5	5	-0.070	1.296	165	1.052	1.099	1.414	1.117	5	0	—	—	192	1.241	1.251	1.627	1.261			

Tablo 34

50 ve 160 Madde Üzerinden 99990 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde									160 Madde										
Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar						
Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler			Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		
					Yanlılık	RMSE	Yanlılık	RMSE	Yanlılık						RMSE	Yanlılık	RMSE	Yanlılık	RMSE
0	33607	0.000	1.043	0	—	—	—	—	0	3626	0.000	0.991	0	—	—	—	—		
1	34170	0.000	1.043	23	0.016	0.221	0.063	0.219	1	11905	0.000	0.989	2	0.005	0.131	0.020	0.131		
2	17132	0.000	1.044	63	0.025	0.273	0.105	0.269	2	19325	0.000	0.992	10	0.005	0.140	0.024	0.139		
3	5634	-0.001	1.042	99	0.033	0.340	0.149	0.334	3	20719	0.000	0.992	22	0.008	0.153	0.032	0.152		
4	1387	-0.001	1.046	109	0.047	0.450	0.203	0.442	4	16574	0.000	0.989	34	0.010	0.167	0.041	0.165		
5	310	-0.004	1.052	895	0.104	0.992	0.348	0.983	5	10603	0.000	0.991	47	0.013	0.187	0.053	0.185		
6	100	-0.021	1.071	824	0.106	1.013	0.360	1.004	6	5609	0.000	0.989	54	0.015	0.209	0.066	0.207		
7	77	-0.031	1.084	431	0.106	1.002	0.378	0.991	7	2525	0.000	0.987	56	0.019	0.235	0.083	0.232		
8	81	-0.038	1.036	193	0.129	1.040	0.456	1.027	8	989	0.001	0.999	52	0.022	0.262	0.101	0.259		
9	76	-0.044	1.005	145	0.158	1.161	0.612	1.141	9	363	0.001	0.983	48	0.025	0.287	0.117	0.283		
10	68	-0.033	0.998	1032	0.236	1.071	0.735	1.051	10	129	0.001	0.998	47	0.026	0.311	0.134	0.307		
11	53	-0.042	0.997	884	0.231	1.142	0.759	1.118	11	42	0.006	1.005	35	0.028	0.341	0.153	0.336		
12	36	-0.046	0.961	388	0.239	1.247	0.803	1.218	12	10	0.001	1.124	32	0.033	0.383	0.177	0.378		
13	25	-0.053	0.910	130	0.356	1.367	0.919	1.338	13	1	0.002	1.180	30	0.037	0.422	0.202	0.416		
14	15	-0.045	0.909	46	0.706	1.451	1.197	1.433	14	1	-0.074	1.626	29	0.044	0.472	0.233	0.466		
≥15	19	-0.052	0.927	1938	0.398	0.982	1.109	0.954	15	2	-0.080	1.709	28	0.047	0.511	0.256	0.504		

(Devam Ediyor)

Tablo 34 (Devam)

50 ve 160 Madde Üzerinden 99990 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

160 Madde									
Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar				
Silme Sayısı	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		RMSE
					Yanlılık	RMSE	Yanlılık	RMSE	
16	1	-0.064	1.573	139	0.102	0.919	0.349	0.915	
17	2	-0.057	1.387	345	0.112	0.965	0.364	0.960	
18	3	-0.077	1.639	488	0.114	0.975	0.366	0.971	
19	3	-0.061	1.335	479	0.112	0.968	0.367	0.964	
20	6	-0.065	1.325	354	0.114	0.968	0.369	0.964	
21	7	-0.054	1.398	213	0.108	0.954	0.373	0.950	
22	8	-0.042	1.279	114	0.104	0.952	0.384	0.946	
23	12	-0.033	1.175	68	0.092	0.945	0.414	0.937	
24	14	-0.041	1.190	50	0.078	0.960	0.465	0.949	
25	16	-0.028	1.141	40	0.073	0.999	0.517	0.985	
26	22	-0.029	1.095	36	0.067	1.057	0.556	1.041	
27	21	-0.014	1.016	37	0.066	1.122	0.596	1.105	
28	27	-0.019	1.013	35	0.073	1.188	0.634	1.170	
29	27	-0.014	0.950	36	0.073	1.269	0.678	1.249	
30	30	-0.011	0.913	38	0.062	1.333	0.708	1.311	
31	23	-0.012	0.883	36	0.072	1.421	0.759	1.398	
32	25	-0.013	0.847	185	0.235	1.088	0.765	1.076	
33	25	-0.011	0.861	428	0.265	1.033	0.768	1.023	

(Devam Ediyor)

Tablo 34 (Devam)

50 ve 160 Madde Üzerinden 99990 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Silme Sayısına Göre Yanlılık ve Ortalama Hataların Karekökü

160 Madde									
Silme Sayısı	Tahrifatsız Test Alanlar				Tahrifatlı Test Alanlar				
	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		RMSE
					Yanlılık	RMSE	Yanlılık	RMSE	
34	19	-0.006	0.806	534	0.268	1.029	0.769	1.019	
35	16	-0.013	0.799	453	0.269	1.030	0.770	1.020	
36	14	-0.009	0.781	296	0.269	1.055	0.777	1.044	
37	13	-0.001	0.754	153	0.282	1.088	0.795	1.077	
38	9	-0.012	0.821	67	0.283	1.117	0.808	1.105	
39	7	0.003	0.616	27	0.371	1.180	0.886	1.169	
40	6	-0.013	0.648	11	0.613	1.360	1.119	1.349	
41	4	-0.020	0.684	6	1.007	1.488	1.457	1.482	
42	2	0.031	0.690	4	1.406	1.517	1.711	1.522	
43	2	0.012	0.664	5	1.661	1.503	1.947	1.510	
44	1	-0.023	0.547	4	1.803	1.534	2.080	1.542	
45	1	0.032	0.648	5	1.890	1.516	2.170	1.525	
46	0	—	—	6	1.825	1.516	2.093	1.527	
47	0	—	—	6	1.980	1.504	2.272	1.514	
>= 48	0	—	—	1975	0.485	0.938	1.188	0.926	

Tablo 35

50 ve 160 Madde Üzerinden 99990 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde										160 Madde									
Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar					Tahrifatsız Test Alanlar					Tahrifatlı Test Alanlar				
Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler		Silme Sayıları	Yetenek Düzeyi	Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
						Yanlılık	RMSE	Yanlılık	RMSE							Yanlılık	RMSE	Yanlılık	RMSE
0-1	1	13435	0.001	1.579	0	—	—	—	—	0-8	1	13435	0.001	1.579	0	—	—	—	—
	2	13465	0.000	0.857	0	—	—	—	—		2	13465	0.000	0.857	2	0.042	0.353	0.165	0.350
	3	13419	0.000	0.289	23	0.016	0.221	0.063	0.219		3	13419	0.000	0.289	276	0.015	0.210	0.066	0.208
	4	13716	0.000	0.471	0	—	—	—	—		4	13716	0.000	0.471	0	—	—	—	—
	5	13742	-0.001	1.381	0	—	—	—	—		5	13742	-0.001	1.381	0	—	—	—	—
2-4	1	4799	0.003	1.582	0	—	—	—	—	9-17	1	114	0.002	1.440	87	0.029	1.411	0.000	1.396
	2	4764	0.001	0.862	0	—	—	—	—		2	110	0.002	0.665	270	0.050	0.559	0.279	0.552
	3	4808	0.001	0.294	270	0.035	0.360	0.159	0.353		3	108	0.000	0.178	205	0.050	0.249	0.204	0.245
	4	4902	-0.002	0.468	0	—	—	—	—		4	110	0.000	0.615	86	0.141	0.612	0.322	0.616
	5	4879	-0.005	1.377	1	0.630	2.480	0.698	2.484		5	110	-0.001	1.423	85	0.262	1.420	0.400	1.423
5-7	1	105	-0.021	1.596	374	0.026	1.595	0.432	1.567	18-26	1	35	-0.100	1.521	382	0.038	1.361	0.462	1.345
	2	95	-0.009	0.869	551	0.055	0.822	0.360	0.806		2	14	-0.030	0.694	520	0.059	0.720	0.377	0.710
	3	93	-0.005	0.295	473	0.072	0.339	0.296	0.328		3	10	-0.010	0.174	315	0.093	0.180	0.324	0.178
	4	98	-0.010	0.463	373	0.131	0.454	0.312	0.462		4	11	0.018	0.639	313	0.141	0.613	0.323	0.618
	5	96	-0.014	1.369	379	0.274	1.363	0.407	1.370		5	20	0.012	1.522	313	0.259	1.413	0.400	1.417

(Devam Ediyor)

Tablo 35 (Devam)

50 ve 160 Madde Üzerinden 99990 Örneklem Büyüklüğü İçin Yetenek Kestirimlerinin Yetenek Düzeylerine ve Gruplandırılmış Silme Sayılarına Göre Yanlılık ve Ortalama Hataların Karekökü

50 Madde										160 Madde									
Silme Sayıları	Yetenek Düzeyi	Tahrifatsız Test Alanlar			Tahrifatlı Test Alanlar					Silme Sayıları	Yetenek Düzeyi	Tahrifatsız Test Alanlar			Tahrifatlı Test Alanlar				
		Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler				Test Alanlar	Yanlılık	RMSE	Test Alanlar	Silinmemiş Maddeler		Tüm Maddeler	
						Yanlılık	RMSE	Yanlılık	RMSE							Yanlılık	RMSE	Yanlılık	RMSE
8-10	1	42	-0.053	1.574	330	0.056	1.509	0.761	1.467	27-35	1	32	-0.107	1.301	585	0.062	1.422	0.794	1.396
	2	48	-0.035	0.872	406	0.086	0.928	0.595	0.901		2	45	-0.037	0.678	304	0.130	0.679	0.698	0.663
	3	48	-0.029	0.298	203	0.166	0.313	0.588	0.295		3	45	0.006	0.179	296	0.210	0.181	0.657	0.178
	4	46	-0.038	0.455	206	0.280	0.442	0.608	0.455		4	47	0.016	0.604	297	0.325	0.604	0.676	0.613
	5	42	-0.038	1.365	225	0.652	1.378	0.878	1.388		5	45	0.029	1.401	299	0.667	1.382	0.934	1.388
11-13	1	14	-0.081	1.492	475	0.054	1.656	0.863	1.603	36-44	1	3	-0.123	1.189	143	0.060	1.580	0.890	1.546
	2	21	-0.045	0.874	234	0.105	0.895	0.678	0.863		2	14	-0.052	0.646	102	0.132	0.684	0.698	0.666
	3	23	-0.036	0.308	223	0.180	0.317	0.631	0.295		3	16	0.006	0.178	105	0.208	0.182	0.650	0.178
	4	27	-0.023	0.448	215	0.300	0.445	0.648	0.462		4	15	0.020	0.609	102	0.326	0.608	0.674	0.618
	5	28	-0.059	1.337	254	0.743	1.323	0.996	1.336		5	10	0.029	1.315	122	0.889	1.442	1.153	1.450
≥14	1	2	-0.163	1.400	421	0.090	1.627	1.190	1.560	≥ 45	1	0	-0.132	1.196	403	0.116	1.438	1.229	1.399
	2	6	-0.050	0.841	408	0.179	0.897	1.019	0.854		2	0	-0.031	0.632	403	0.222	0.689	1.053	0.665
	3	7	-0.040	0.306	407	0.307	0.326	0.962	0.297		3	0	-0.075	0.193	403	0.359	0.185	1.008	0.178
	4	9	-0.019	0.423	407	0.518	0.415	1.019	0.436		4	1	0.022	0.482	403	0.572	0.583	1.079	0.595
	5	10	-0.066	1.345	342	1.047	1.097	1.409	1.116		5	0	0.189	1.379	380	1.256	1.259	1.640	1.269

BENZERLİK BİLDİRİMİ

“Test Tahrifatının Belirlenmesinde Silme Belirleme İndeksinin I. Tip Hata ve İstatistiksel Güç Açısından İncelenmesi” başlıklı tezimin ana bölümü (ön bölüm, kaynaklar ve ekler hariç) Turnitin İntihali Engelleme Programı aracılığıyla incelenmiş ve ilgili rapor danışmanım tarafından da kontrol edilmiştir. Kontrol sırasında (1) “Beş sözcükten daha az olan benzeşmeler” (2) “Kaynaklar” (3) “Doğrudan Alıntılar” dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıda sunulmuştur.

Rapor Tarihi	: 15.07.2020
Gönderim Numarası	: 1357771838
Sayfa Sayısı	: 123
Sözcük Sayısı	: 37227
Karakter Sayısı	: 243117
Benzerlik Oranı	: % 2
Savunma Tarihi	: 27.07.2020

Yukarıda belirtilen sonuçları gösteren Turnitin İntihali Engelleme Programı’na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu beyanım ekinde Enstitüye teslim ettiğimi, tezimin %10’dan fazla benzerlik oranı içerdiğinin belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı Soyadı: Esra KINAY ÇİÇEK

Tarih: 15.07.2020

İmza:



ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı : Esra KINAY ÇİÇEK
E-Posta Adresi : esrakinay@gmail.com

İş Deneyimi:

Unvan	Görev Yeri	Yıl
Ölçme ve Değerlendirme Uzmanı	Gazi Üniversitesi Vakfı Özel Okulları	2010 – 2014
Ölçme ve Değerlendirme Uzmanı	TED Ankara Koleji Vakfı Özel Okulları	2014 – 2019

Akademik Bilgiler

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	Fen Bilgisi Öğretmenliği	Hacettepe Üniversitesi	2004 – 2008
Yüksek Lisans	Ölçme ve Değerlendirme	Ankara Üniversitesi	2008 – 2012
Doktora	Ölçme ve Değerlendirme	Ankara Üniversitesi	2012 – 2020

Yayınlar:

Özdil Örs, S. ve Kınay, E. (2015). 5. Sınıf Seçmeli Ders Tercihlerinin Sıralama Yargıları Kanunıyla Ölçeklenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(2), 268-278.

Kınay, E. ve Tavşancıl, E. (2015, Mayıs). *Üniversite Giriş Sınavı Yordama Geçerliliği Çalışmalarının Meta Analizi*. Uluslararası Eğitim Kongresi: Gelecek İçin Eğitim Kongresi (ICEFIC 2015). Ankara Üniversitesi, Ankara.