

**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

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

**MADDE PARAMETRE SAPMASININ BİREYE
UYARLANMIŞ TESTLERDEN ELDE EDİLEN YETENEK
KESTİRİMLERİNE ETKİSİNİN FARKLI KOŞULLAR
ALTINDA İNCELENMESİ**

DOKTORA TEZİ

MERVE ŞAHİN KÜRŞAD

**ANKARA
ARALIK, 2020**



**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

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

**MADDE PARAMETRE SAPMASININ BİREYE
UYARLANMIŞ TESTLERDEN ELDE EDİLEN YETENEK
KESTİRİMLERİNE ETKİSİNİN FARKLI KOŞULLAR
ALTINDA İNCELENMESİ**

DOKTORA TEZİ

MERVE ŞAHİN KÜRŞAD

**DANIŞMANLAR: PROF. DR. ÖMAY ÇOKLUK BÖKEOĞLU
PROF. DR. R. NÜKHET ÇIKRIKÇI**

**ANKARA
ARALIK, 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.



Merve ŞAHİN KÜRŞAD

ÖZET

MADDE PARAMETRE SAPMASININ BİREYE UYARLANMIŞ TESTLERDEN ELDE EDİLEN YETENEK KESTİRİMLERİNE ETKİSİNİN FARKLI KOŞULLAR ALTINDA İNCELENMESİ

ŞAHİN KÜRŞAD, Merve

Doktora Tezi, Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanları: Prof.Dr.Ömay ÇOKLUK BÖKEOĞLU, Prof.Dr.R. Nükhet ÇIKRIKÇI

Aralık, 2020, xiii + 99 sayfa

Madde parametre sapması (MPS), maddelerin parametre değerlerinin bireylerin maddelere aşına olması, müfredat değişimi gibi çeşitli sebeplerle zaman içerisinde sistematik olarak farklılaşmasıdır. Bu durum bilgisayar ortamında bireye uyarlanmış testlerde (BOBUT) ortaya çıktığı durumda madde ve yetenek parametre kestirimlerinde hataya neden olmaktadır. Ancak MPS'nin BOBUT uygulamaları üzerinde, hangi koşullarda etkisi olduğu tam olarak bilinmemektedir. Bu doğrultuda bu çalışmanın amacı BOBUT uygulamalarında MPS'nin ölçme kesinliği, test bilgi fonksiyonu (TBF) ve test etkililiği üzerindeki etkisini incelemektir. Simülasyon çalışması olarak yürütülen bu çalışmada örneklem büyüklüğü (1000, 5000), MPS büyüklüğü (0.00 logit, 0.50 logit, 0.75 logit, 1.00 logit), MPS içeren madde yüzdesi (%0, %5, %10, %20), üç zaman noktası ve madde bankası büyüklüğü (200, 500, 1000) koşulları karşılaştırılmıştır. Belirtilen koşulların yetenek kestirimleri üzerindeki etkisini incelemek için ölçme kesinliği, test bilgi fonksiyonu ve test etkililiği değerleri hesaplanmış ve ölçme kesinliği ve test bilgi fonksiyonu açısından kestirimler arasında fark olup olmadığını incelemek için bağımsız örneklemeler için üç faktörlü ANOVA analizi yapılmıştır. Araştırma sonucuna göre MPS büyüklüğü, MPS yüzdesi ve MPS'li madde bankası ile yapılan ölçme sayısı arttıkça ölçme kesinliği, test bilgi fonksiyonu ve test etkililiği azalmaktadır. Yapılan bağımsız örneklemeler için üç faktörlü ANOVA analizleri de ölçme kesinliği ve test bilgi fonksiyonu açısından elde edilen farkların çoğunlukla istatistiksel olarak manidar olduğunu göstermiştir. Tüm MPS'li koşullar ölçme kesinliği, test bilgi fonksiyonu ve test etkililiğini olumsuz etkilese de örneklem büyüklüğü ve madde bankası büyüklüğünün bu faktörler üzerinde artan veya azalan yönde bir etkisinin olmadığını göstermiştir.

Anahtar Sözcükler: Madde parametre sapması, bilgisayar ortamında bireye uyarlanmış test, ölçme kesinliği, test bilgi fonksiyonu, test etkililiği

ABSTRACT

THE STUDY OF THE EFFECT OF ITEM PARAMETER DRIFT ON ABILITY ESTIMATION OBTAINED FROM ADAPTIVE TESTING UNDER DIFFERENT CONDITIONS.

ŞAHİN KÜRŞAD, Merve

Dissertation, Department of Educational Measurement and Evaluation

Supervisors: Prof.Dr. Ömay ÇOKLUK BÖKEOĞLU, Prof.Dr. R. Nükhet ÇIKRIKÇI

December 2020, xiii + 99 pages

Item parameter drift (IPD) is the differentiation of item parameters systematically due to different reasons such as curriculum change and overexposure of items. If IPD occurs in Computer Adaptive Testing (CAT), it causes errors in estimation of item and ability parameters. However, in which conditions IPD affect CAT applications is not known exactly. Accordingly, the purpose of this study is to examine the effect of IPD on measurement precision, test information function and test efficiency in CAT applications. This study was implemented as simulation a study and sample size (1000, 5000), magnitude of IPD (0.00 logit, 0.50 logit, 0.75 logit, 1.00 logit), percentage of IPD (%0, %5, %10, %20), three time points and item pool size (200, 500, 1000) factors were compared. To investigate the effect of these conditions on ability estimations measurement precision, test information function and test efficiency values were calculated. Moreover, to investigate the differentiation between values of measurement precision and test information function, multi-factor ANOVA for independent samples analysis were done. According to the results, as the magnitude of IPD, percentage of IPD and number of measurement increased, measurement precision, test information function and test efficiency decreased. Also, multi-factor ANOVA for independent samples analysis indicated that there were statistically significant differences in aspect of measurement precision and test information function. Despite all of IPD conditions affected the measurement precision, test information function and test efficiency negatively, there were no increasing or decreasing effect of sample size and size of item pool were observed.

Key Words: Item parameter drift, computer adaptive testing, measurement precision, test information function, test efficiency

ÖNSÖZ

Madde parametre sapması, maddelerin parametre değerlerinin; test aşinalığı, maddelerin çok fazla kullanımı, müfredat değişimi vb. nedenlerden dolayı zamanla farklılaşmasıdır. Madde parametre sapması, bilgisayar ortamında bireye uyarlanmış test uygulamalarında ortaya çıktığında, yetenek ve madde parametre kestirimlerinde hatalara neden olmaktadır. Bu durum da test puanlarından yapılan çıkarımların geçerliğini olumsuz yönde etkilemekte ve puanların yorumlanması ve programların değerlendirilmesinde hatalara neden olmaktadır. Bilgisayar ortamında bireye uyarlanmış test uygulamalarında, bu durumu etkileyen koşulların belirlenmesi, madde ve yetenek parametrelerinin en az hata ile kestirilmesi açısından önemlidir. Bu nedenle bu çalışma kapsamında bilgisayar ortamında bireye uyarlanmış test uygulamalarında, madde parametre sapmasının ölçme kesinliği, test bilgi fonksiyonu ve test etkililiği üzerindeki etkisi incelenmiştir.

Doktora sürecim boyunca desteğini esirgemeyen, her türlü sorumu sabırla cevaplayan, tezimin şekillenmesine yardımcı olan ve gelişmemde büyük katkısı olan değerli hocalarım, danışmanlarım Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU ve Prof. Dr. R. Nükhet ÇIKRIKÇI'ya; tez izleme komitemde yer alarak değerli görüşlerini benimle paylaşan ve tezime katkıda bulunan değerli hocalarım Prof. Dr. Duygu ANIL ve Doç. Dr. Deha DOĞAN'a; tez jürimde yer alan ve tezime katkıda bulunan Prof. Dr. Zekeriya NARTGÜN ve Doç. Dr. K.Zülfikar DENİZ'e; tez sürecinde analizlerde karşılaştığım problemleri sabırla çözen ve yardımını esirgemeyen değerli hocam Doç. Dr. Okan BULUT'a; tezimin her aşamasında manevi yönden desteklerini esirgemeyen, bana moral olan sevgili arkadaşlarım Doç. Dr. Seher YALÇIN, Dr. Öğr. Üyesi Münevver İLGÜN DİBEK, Arş. Gör. Dr. Hatice Çiğdem BULUT, ve Arş. Gör. Burcu TOPTAŞ'a; süreç boyunca bana en büyük desteği veren, omuzlarımdaki yükün çoğunu alan hayat arkadaşım Servet Meriç KÜRŞAD'a; bugünlere gelmemde bana sabırla destek olan ve her anımda yanımda olan canım anneme ve babama; hem hamilelik sürecimde hem de doğduktan sonra bu stresli süreci benimle yaşayan, yüzümü güldürerek stresimi alan canım kızım İdil'ime çok teşekkür ederim.

Merve ŞAHİN KÜRŞAD



*Hayat arkadaşım Meriç'e ve
Canım kızım İdil'ime...*

İÇİNDEKİLER

	Sayfa
ETİK İLKELERE UYGUNLUK BİLDİRİMİ.....	iii
ÖZET	iv
ABSTRACT	v
ÖNSÖZ	vi
İÇİNDEKİLER	viii
TABLolar DİZİNİ	x
ŞEKİLLER DİZİNİ	xii
KISALTMALAR.....	xiii
BÖLÜM 1.....	1
GİRİŞ	1
Problem	1
Amaç	16
Önem	17
Sınırlılıklar.....	18
BÖLÜM 2.....	19
YÖNTEM.....	19
Araştırma Modeli	19
Verilerin Üretilmesi ve Analizi	19
Simülatif Veri Üretiminde Dikkate Alınan Sabit Koşullar	20
Simülatif Veri Üretiminde Değişimlenen Koşullar.....	26
BÖLÜM 3.....	33
BULGULAR VE YORUMLAR	33
Örneklem Büyüklüğünün 1000 Olduğu Durumda Koşulların Karşılaştırılmasına İlişkin Bulgular	33
Örneklem Büyüklüğünün 5000 Olduğu Durumda Koşulların Karşılaştırılmasına İlişkin Bulgular	55
BÖLÜM 4.....	82
SONUÇLAR VE ÖNERİLER.....	82
Sonuçlar.....	82
Öneriler.....	84
KAYNAKLAR.....	87
EKLER.....	96
EK 1. Etik Kurul Muafiyet Onayı.....	97
BENZERLİK BİLDİRİMİ	98
ÖZGEÇMİŞ.....	99



TABLolar DİZİNİ

Tablo

Sayfa

Tablo 1 Simülatif Veri Üretiminde Dikkate Alınan Sabit ve Değişimlenen Koşullar ...	20
Tablo 2 Madde Bankasının $k=200$ Olduğu MPS'li Veri Setlerinde Madde Bankalarının Ortalama Madde Güçlük Değerleri.....	28
Tablo 3 Madde Bankasının $k=500$ Olduğu MPS'li Veri Setlerinde Madde Bankalarının Ortalama Madde Güçlük Değerleri.....	29
Tablo 4 Madde Bankasının $k=1000$ Olduğu MPS'li Veri Setlerinde Madde Bankalarının Ortalama Madde Güçlük Değerleri.....	29
Tablo 5 Madde Parametre Sapmasını Değerlendirme Kriterleri	30
Tablo 6 Örneklem $n=1000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Yanlılık Değerlerinin Karşılaştırılması	36
Tablo 7 Örneklem $n=1000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre OHKK Değerlerinin Karşılaştırılması	41
Tablo 8 Örneklem $n=1000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre MOF Değerlerinin Karşılaştırılması	45
Tablo 9 Örneklem $n=1000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre TBF Değerlerinin Karşılaştırılması	50
Tablo 10 Örneklem $n=1000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Test Etkililiği Değerleri	52
Tablo 11 Örneklem $n=5000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Yanlılık Değerlerinin Karşılaştırılması	58
Tablo 12 Örneklem $n=5000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre OHKK Değerlerinin Karşılaştırılması	63

Tablo 13 Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre MOF Değerlerinin Karşılaştırılması	67
Tablo 14 Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre TBF Değerlerinin Karşılaştırılması	72
Tablo 15 Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Test Etkililiği Değerleri	74



ŞEKİLLER DİZİNİ

Şekil	Sayfa
Şekil 1.a ve 1.b. $n= 1000$ ve $n= 5000$ Kişilik Verilerde Yetenek Dağılımları	21
Şekil 2.a, 2.b ve 2.c. $k= 200$, $k=500$ ve $k=1000$ Maddelik Madde Bankaları için Madde Bankası Bilgi Fonksiyonu Grafikleri	22
Şekil 3.a, 3.b ve 3.c. Örneklem $n=1000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre Yanlılık Değerlerinin Karşılaştırılmasına İlişkin Grafikler	34
Şekil 4.a, 4.b ve 4.c. Örneklem $n=1000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre OHKK Değerlerinin Karşılaştırılmasına İlişkin Grafikler	39
Şekil 5.a, 5.b ve 5.c. Örneklem $n=1000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre MOF Değerlerinin Karşılaştırılmasına İlişkin Grafikler	44
Şekil 6.a, 6.b ve 6.c. Örneklem $n=1000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre TBF Değerlerinin Karşılaştırılmasına İlişkin Grafikler	48
Şekil 7.a, 7.b ve 7.c. Örneklem $n=5000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre Yanlılık Değerlerinin Karşılaştırılmasına İlişkin Grafikler	56
Şekil 8.a, 8.b ve 8.c. Örneklem $n=5000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre OHKK Değerlerinin Karşılaştırılmasına İlişkin Grafikler	61
Şekil 9.a, 9.b ve 9.c. Örneklem $n=5000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre MOF Değerlerinin Karşılaştırılmasına İlişkin Grafikler	66
Şekil 10.a, 10.b ve 10.c. Örneklem $n=5000$ Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre TBF Değerlerinin Karşılaştırılmasına İlişkin Grafikler	71

KISALTMALAR

1PLM	Bir Parametrelili Lojistik Model
2PLM	İki Parametrelili Lojistik Model
3PLM	Üç Parametrelili Lojistik Model
BOBUT	Bilgisayar Ortamında Bireye Uyarlanmış Test
BSD	Beklenen Sonsal Dağılım
EYOY	En Yüksek Olabilirlik Yöntemi
EYSD	En Yüksek Sonsal Dağılım
MOF	Mutlak Ortalama Fark
MPS	Madde Parametre Sapması
MTK	Madde Tepki Kuramı
OHKK	Ortalama Hataların Karekökü
TBF	Test Bilgi Fonksiyonu

BÖLÜM 1

GİRİŞ

Bu bölümde araştırmanın problem durumu, amacı, önemi ve sınırlılıkları yer almaktadır.

Problem

Geleneksel kâğıt-kalem testlerine göre birçok avantajı olması ve bireylerin yetenek kestirimlerinde daha doğru ve güvenilir sonuçlar vermesi nedeniyle bilgisayar ortamında bireye uyarlanmış testler (BOBUT, Computer Adaptive Testing- CAT), son yıllarda sıklıkla tercih edilen uygulamalardan biri olmuştur. BOBUT uygulamalarında çeşitli bilgisayar algoritmaları kullanılarak madde güçlüğü ve bireylerin yetenek düzeyleri eşleştirilmekte ve böylece her bireyin, yeteneğine uygun maddelerle karşılaşması sağlanmaktadır (Reckase, 2011).

Madde Tepki Kuramı'na (MTK) dayalı BOBUT uygulamalarında her bir bireyin yeteneği, madde güçlüğü ile aynı ölçek üzerine yerleştirilmekte ve ilgili maddeyi %50 olasılıkla doğru yanıtlama durumları hesaplanmaktadır (Lord, 1980). Bu durum, bireylerin yeteneklerinin kestirilmesinde, kendi yetenek düzeylerine uygun maddelerin uygulanmasını sağlayarak, geleneksel kâğıt-kalem testlerine göre zaman ve maliyet açısından daha kullanışlı testler elde edilmesine olanak vermektedir (Wainer, 1993; Weiss ve Kingsbury, 1984). Öyle ki, geleneksel kâğıt-kalem uygulamalarına göre hem test uzunluğunda %50 oranında azalma sağlanarak daha verimli testler elde edilmekte, hem de en az bu testler kadar geçerli ve güvenilir sonuçlara ulaşılmaktadır (Çıkrıkçı-Demirtaşlı, 1999; Kaptan, 1993).

BOBUT uygulamalarından geçerli ve güvenilir sonuçlar elde edilmesinde en önemli faktör nitelikli maddelerden oluşan geniş bir madde bankasıdır; çünkü maddeler nitelikli olduğu ve bu özelliğini sürdürebildiği ölçüde doğru sonuçlar elde edilebilir (Bock, Muraki ve Pfeifferberger, 1988). Madde bankasının sürekliliğini koruması, hem

test güvenliği hem de madde parametrelerindeki değişimi gözlemlemek açısından önemlidir (Risk, 2015). Madde bankasının sürekliliğine ilişkin olarak Eğitimde ve Psikolojik Testlerde Standartlar (Standards for Educational and Psychological Testing) içerisinde yer alan bir öneride, test uygulayan kurumların madde bankalarındaki maddelere ilişkin periyodik kontroller yapması gerektiği, ancak böylelikle uzun yıllar ortak bir ölçek üzerinde değerlendirme yapabilmenin mümkün olacağı belirtilmiştir. Mesleki, yasal ve bilimsel alanlardaki gelişmeler, test içeriklerinin de güncellenmesi ihtiyacını gündeme getirmektedir. Bu değişimin sürekliliğine paralel olarak mesleki ve bilimsel alanlara ilişkin tanımlar değişmektedir. Bu doğrultuda test içeriklerinin de bu değişimi yansıtacak şekilde gözden geçirilmesi gerekmektedir (Standards; American Educational Research Association [AERA], American Psychological Association [APA] ve National Council on Measurement in Education [NCME], 2014). Yukarıda sözü edilen gerekçeler, madde bankalarının güncelliğine ilişkin kontrollerin düzenli olarak yapılmasına olan ihtiyacı ortaya koymaktadır.

Bir madde bankasında yer alan maddelerin uzun süre kullanılması, maddelerin niteliğine ilişkin sürekliliği sağlamayı güçleştirmektedir. Örneğin, bazı maddelerin uygulamalarda görülme sıklığı, diğer maddelerden daha yüksek olabilmektedir. Bu durum bireylerin ilgili maddelere aşına olmasına neden olmaktadır. Madde bankasının güvenliği sağlanmış olsa bile bireylerin maddelerle olan bu etkileşim sıklığı, güvenliği tehdit eden bir unsur olmakta ve madde parametrelerinin zamanla değişmesine neden olmaktadır. Bu değişim, madde parametre sapması (kayması) (MPS-Item Parameter Drift) olarak adlandırılır (Bock vd., 1988).

İlk olarak 1980'lerde alanyazında yer almaya başlayan MPS, farklı şekillerde tanımlanmaktadır. Bu tanımlardan bazıları; bazı maddelerin bir ya da daha fazla parametresinin zamanla değişimi (Goldstein, 1983); arka arkaya yapılan uygulamalarda madde parametrelerinin değişkenlik göstermesi (Bock vd., 1988) veya madde parametrelerinin zaman içerisinde sistematik olarak farklılaşması (Hatfield ve Nhouyvanisvong, 2005) şeklindedir. Bazı araştırmacılar da MPS'yi, değişen madde fonksiyonu (DMF) olarak tanımlamaktadır. Ancak iki kavram farklı durumları ifade etmektedir. DMF, farklı test uygulamalarına katılan gruplarda bazı maddelerin farklı performans göstermesi iken (Babcock ve Albano, 2012); MPS, madde parametrelerinde zaman içerisinde gerçekleşen değişimdir (McCoy, 2009). Bir diğer deyişle DMF, madde parametrelerinin gruplar arasında (cinsiyet, etnik gruplar vb. özelliklere göre)

farklılaşması iken; MPS madde parametrelerinin gruplar arasında farklılaşmasından çok, zaman içerisinde kullanım sıklığı ile ilişkili olarak farklılaşmasını ifade etmektedir.

Sapma durumunun parametre, ölçek veya ölçülen yapıda olmak üzere farklı türleri vardır. Buna bağlı olarak sapmanın farklı türleri vardır. Bunlardan biri parametre sapmasıdır (parameter drift). Parametre sapması, bir maddenin parametrelerinin zaman içerisinde sistematik olarak farklılaşması olarak tanımlanmaktadır. Ölçek sapması (scale drift), çoklu test uygulamalarında rastgele hatanın birikmesi (Livingston, 2004) veya zamanla yapı veya içeriğin farklılaşmasından kaynaklı olarak ölçme ölçeğinin farklılaşmasıdır (Babcock ve Albano, 2012). Yapı sapması (construct shift) ise farklı test formları için yapıdaki farklılaşma olarak tanımlanmaktadır (Martineau, 2004).

MPS, nitelikli madde yazıldığı veya madde bankasının güvenliğinin sağlandığı durumda bile karşılaşılabilecek bir problemdir. Bu problemin ortaya çıkmasının farklı nedenleri vardır ve etkisi pozitif (zor), negatif (kolay) veya her iki yönde olabilmektedir. Bir madde zamanla zorlaşıyorsa (pozitif yönde sapma gösteriyorsa), bunun temel nedenleri; tarihi ve kültürel değişimler, madde kalibrasyonunun doğru yapılmaması, maddenin ölçekteki yerinin (item location) yanlış hesaplanması, bilgi, yetenek ve öğretim etkinliklerindeki değişimlerdir. Bir madde zamanla kolaylaşıyorsa (negatif yönde sapma gösteriyorsa), bunun nedenleri de; maddelerin çok fazla kullanılması, madde kalibrasyonunun doğru yapılmaması, politika veya müfredattaki değişimler, kopya veya güvenlik açığı, bilgi, yetenek ve öğretim etkinliklerindeki değişimler, maddenin ölçekteki yerinin yanlış hesaplanması, tarihi ve kültürel değişimlerdir (Li, 2008; Stahl ve Muckle, 2007). Her iki yönde sapma görülmesi ise madde bankasındaki bir kısım maddenin zor, bir kısım maddenin ise kolay yönde sapma göstermesi anlamına gelmektedir. MPS'nin yalnızca zor veya kolay yönde olması, her iki yönde olmasına göre yetenek kestirimlerini daha olumsuz etkilemektedir (Babcock ve Albano, 2012). Yukarıda sözü edilen nedenlerle MPS, bireylerin puanlarının yorumlanmasında ya da programların değerlendirilmesinde hatalara yol açmaktadır (Kolen ve Brennan, 1995).

Madde parametre sapmasının bahsi geçen nedenlerine bağlı olarak bazı olumsuz sonuçlar ortaya çıkmaktadır. Bunlar içerisinde en önemlisi, MTK'nın temel varsayımlarından biri olan değişmezlik varsayımının ihlal edilmesidir. MPS görülmesi durumunda ölçme sonuçlarına hata karışmakta ve ölçülmesi gereken asıl yapının dışına çıkılmaktadır. Ölçülmesi amaçlanan yapıyla ilgisi olmayan değişkenlerin ölçme sonuçlarına karışması da geçerliği düşürmektedir (McCoy, 2009). Değişmezliğin sağlandığı durumda puanlar arasındaki fark, bireylerarası yetenek farklılıklarını veya

zamana bağılı olarak bireylerin gelişimini doğru bir şekilde yansıtmaktadır. Ancak MPS'nin olduğu durumda, madde parametrelerinde değişmezlik gerektiren, test eşitleme, test geliştirme/paralel test geliştirme, BOBUT gibi uygulamalarda bazı problemlere neden olmaktadır (Li, 2008). Örneğin, MPS olması test eşitlemede eşitleme hatasını arttırmakta, BOBUT uygulamalarında da öntest maddelerinde olması halinde madde kalibrasyonunda hatalara neden olmaktadır (Meng, Steinkamp ve Matthews-Lopez, 2010).

MPS'nin ortaya çıkardığı bir diğer önemli sorun, özellikle bir bireyin puanının kesme puanına çok yakın olması, iki bireyin puanının birbirine çok yakın olması veya norm dayanaklı değerlendirmelerde görülmektedir. Bu durum, bireylerin yetenek kestirimlerinde sapmaya ve geçti-kaldı kararlarının yanlış verilmesine neden olmaktadır (Rupp ve Zumbo, 2006). MPS'nin bir başka olumsuz sonucu da birikimli etkisi ile ilgilidir. Test için yeterli yanıtlama süresinin verilmemesi, bireylerin test sonunda yer alan bazı maddelere erişememesine ve buna bağılı olarak test sonunda yer alan maddelerin, olduğundan daha güç görünmesine neden olmaktadır. Bu durum, ilerleyen yıllarda önceki test maddeleri ile yapılan ölçeklemeyi etkilemektedir. Bu döngü yıllar boyunca devam ettiğinde de ölçme ölçeğinin kaymasına/sapmasına (measurement scale to drift) neden olmaktadır (Wise ve Kingsbury, 2000). Ölçeğin uzun yıllar tutarlılığını koruyamaması ise, puanların ve farklı test uygulamalarının karşılaştırılabilirliğinde hatalara neden olmaktadır (Guo ve Wang, 2003).

İlgili alanyazında yapılan çalışmalar, MTK modellerinin MPS'nin yukarıda söz edilen etkilerini belli bir dereceye kadar tolare edebildiğini göstermektedir. Rupp ve Zumbo (2003; 2004) orta düzeydeki MPS'nin MTK modelleriyle yapılan yetenek kestirimlerini önemli düzeyde etkilemediğini belirtmişlerdir. Bu durum, MPS'nin daha çok kâğıt-kalem uygulamalarında iki zaman noktası gibi sınırlı bir kapsamda değerlendirilmiş olmasından kaynaklanmaktadır (Rupp ve Zumbo, 2003a, 2003b; Wells, Subkoviak ve Serlin, 2002;). Oysa ki MPS'nin etkilerinin ortaya çıkması için uzun bir süreç gerekmektedir (Kim ve Cohen, 1992). İlgili alanyazında yapılan çalışmalarda, MPS'nin belirlenebilmesi için ilgili test maddelerinin ikiden fazla kez, farklı zamanlarda uygulanması gerektiği belirtilmektedir (Babcock ve Albano, 2012; Chan, Drasgow ve Sawin, 1999; Deng ve Melican, 2009; Kim ve Cohen, 1992).

MPS'nin bir diğer etkisi de BOBUT uygulamalarında görülmektedir. Bu etki hem madde hem de yetenek parametre kestirimlerinde hataya neden olmasından dolayı iki yönlü olarak düşünülmektedir: Madde parametreleri açısından, eski madde parametre

kestirimleri, yeni test maddelerini ölçeklemek amacıyla kullanıldığında madde parametrelerinin kalibrasyonunda hataya neden olmaktadır. Yeni test maddeleri, BOBUT madde bankasıyla kalibre edilirken, MPS'li maddelerden dolayı madde bankasındaki madde parametreleri sabit kalmayabilir. Diğer etkisi ise, gerçek madde parametrelerinden sapma, öntest maddelerinin yanlış kalibre edilmesine ve zamanla tüm ölçeğin sapmasına neden olabilir. Bu durum da bireylerin yetenek kestirimlerinde hatalara neden olur. BOBUT'un başarılı bir şekilde uygulanabilmesi, madde bankasının bütünlüğüne (integrity of item pool) ve madde parametrelerinin sürekliliğine bağlıdır (Deng ve Melican, 2010).

BOBUT uygulamalarında, özellikle madde bankasındaki maddelerin çok fazla ifşa edilmesinin ciddi düzeyde MPS'ye neden olduğu belirtilmektedir (Bock vd., 1988). Bu nedenle madde bankasındaki madde parametrelerinin sürekli kontrol edilmesi ve madde bankalarının düzenli olarak güncellenmesi gerekmektedir. Bu durum da BOBUT uygulamalarında MPS çalışmalarının yapılmasını gerektirmektedir. Alanyazında MPS'nin yetenek kestirimine etkisini gösteren çalışma sayısı çok azdır. Bu çalışmalardan bir kısmı sabit uzunluktaki testler (Babcock ve Albano, 2012; Demirus ve Uysal, 2016; Lee ve Geisinger, 2019; Wells vd., 2002; Wollack, Sung ve Kang, 2005), bir kısmı da BOBUT uygulamalarındaki testler için yapılmıştır (Aksu ve Dünya, 2017; Guo ve Wang, 2003; Han ve Guo 2011; Hagge, Woo ve Dickison, 2011; McCoy, 2009; Risk, 2015; Wells vd., 2002; Zhang, 2014). Bu durumun incelenmesi özellikle geniş ölçekli test uygulamaları, yeterliğin test edildiği durumlar ve seçme ve yerleştirme kararlarının alındığı test uygulamalarında önemlidir. Bu sınavlarda MPS'nin görülmesi, bireylerin yetenek kestirimlerinde hatalara neden olmakta, testin geçerliğini olumsuz etkilemekte, bireyler hakkında güvenilir olmayan kararlar alınmasına neden olabilmekte ve yapılan test uygulamalarının adilliğini tartışmaya açık hale getirmektedir (Risk, 2015).

Sabit Uzunluktaki Test Uygulamalarında Madde Parametre Sapmasının İncelendiği Araştırmalar

Yukarıda belirtildiği üzere MPS'nin olduğu durumlar, MTK varsayımlarının ihlaline, madde kalibrasyonu ve yetenek kestirimlerinde hatalara neden olarak geçerliği olumsuz etkilemektedir. MPS'nin ortaya çıkardığı bir diğer sorun ise bireye uyarlanmış testlerde yetenek kestirimlerinde görülebilir. Bu nedenle bu çalışma kapsamında MPS'nin BOBUT uygulamalarında yetenek kestirimleri, test bilgi fonksiyonu ve test etkililiği

üzerindeki etkisi incelenmektedir. MPS'nin incelendiği çalışmaların birçoğu test uzunluğunun sabit olduğu geleneksel test uygulamaları için yapılmıştır. Bu çalışmalarda farklı MTK modelleri, MPS koşulları ve sınavlar kullanılarak MPS'nin etkileri incelenmiştir. İlgili çalışmalar bu kısımda kronolojik olarak özetlenmiştir.

Bock ve diğerleri (1988) MPS olduğu durumda MTK ölçeğinin sürekliliğini koruması üzerine bir çalışma yapmışlardır. Çalışmada Üniversite Fizik Başarı Testinin (College Board Physics Achievement Test) mekanik alt testinde yer alan 29 madde analiz edilmiştir. Madde parametrelerini ve bu parametrelerin 10 yıllık bir süreç içerisindeki değişimini, üç parametrelili zaman bağımlı model (3PL time-dependent model) ile tahmin etmişlerdir. Örneklem büyüklüğünün 2380 ile 3634 arasında değiştiği çalışmada beş zaman noktasında ölçümler alınmıştır. Çalışma sonucuna göre MPS'nin büyük örneklemelerde zaman içerisinde sistematik olarak ortaya çıktığı bulgusuna ulaşılmıştır. Yapılan ANOVA analizlerine göre her iki test için de MPS'nin madde güçlük parametresini, madde ayıricılık parametresinden daha çok etkilediği belirtilmiştir. Testte yer alan bazı fizik maddelerinin madde güçlük değerlerinin zamanla farklılaştığını ve bu farklılığın istatistiksel olarak manidar olduğunu bulmuşlardır. Araştırmacılar bu durumun nedeninin on yıllık süreçte fizik öğretiminde meydana gelen değişimler olarak açıklamışlardır.

Chan ve diğerleri (1999) tarafından yapılan çalışmada Silahlı Hizmetler Mesleki Yeterlik Bataryasının (Armed Services Vocational Aptitude Battery-ASVAB) psikometrik tutarlılığı incelenmiştir. ASVAB, Amerikan ordusuna personel seçme amacıyla, yılda bir kez uygulanan bir yetenek testidir. Bireylerin geleceğe yönelik akademik ve mesleki başarısını kestirmede kullanılmakta ve dokuz alt testten oluşmaktadır. Bireylerin geleceğe yönelik başarısını da kestirmesinden dolayı, farklı alt boyutlarda maddelerin zamana bağlı değişimleri incelenmiştir. 16 yıl boyunca beş zaman noktasında ölçüm alınmış ve değişen madde fonksiyonu ve değişen test fonksiyonu ile ilgili analizler yapılmıştır. Çalışma kapsamında birey sayısının 2600 ile 12.300 arasında değiştiği veri setleri kullanılmıştır. 200 maddenin analiz edildiği uygulamada madde ve test karakteristik eğrileri incelenerek, maddelerin zamana karşı tutarlılığı incelenmiştir. 200 maddeden 25 (%12.5) tanesinin güçlük değerinde önemli değişiklikler olduğu belirlenmiştir. Güçlük düzeyinde değişen maddeler incelendiğinde; genel yetenek ile ilgili maddelerin zamanın etkisine daha dayanıklı olduğu, anlam çıkarmaya dayalı maddelerin ise zaman etkisinden daha olumsuz etkilendiği belirlenmiştir. Bu bulgu, özellikle anlam çıkarmaya dayalı test maddelerinin zamanın etkisine karşı daha hassas

olduğunu ve bu nedenle madde bankasının sürekli kontrol edilmesi gerektiğini ortaya koymuştur. Değişen test fonksiyonu ile ilgili olarak test karakteristik eğrileri incelendiğinde ise ortalama test puanları arasında büyük farklar olduğu, ancak bu farklara ilişkin etki büyüklüklerinin düşük olduğu belirlenmiştir. Bu değişim de ASVAB alt testlerinin özelliklerinden çok, sınava giren bireylerin özellikleri ile ilişkili olarak değerlendirilmiştir.

Wells ve diğerleri (2002) MPS'nin yetenek kestirimleri üzerindeki etkisini 2 parametrelili lojistik model (2PLM) ile incelemiştir. Simülatif verilerin kullanıldığı çalışmada; örneklem büyüklüğü (300 ve 1000), test uzunluğu (40 ve 80 madde), MPS yüzdesi (%5, %10, %15 ve %20), MPS türü (-a parametresi sapması, -b parametresi sapması ve -ab parametresi sapması), MPS büyüklüğü (0.40 ve 0.50 logit) ve iki zaman noktası koşulları ile çalışılmıştır. MPS'nin değerlendirilmesinde ortalama hataların karekökü (OHKK) kullanılmıştır. MPS, yetenek kestirimlerini en fazla 1000 kişilik örneklem büyüklüğü, -a ve -ab parametre sapması ve MPS'nin %20 olduğu koşulda etkilemiştir. Araştırmacılar, belirlenen koşullar altında MPS'nin yetenek kestirimlerine etkisini düşük düzeyde bulsa da, yetenek kestirimlerini en çok etkileyen iki faktörün örneklem büyüklüğü ve MPS yüzdesi olduğunu belirtmişlerdir.

Demars (2004), çoklu test uygulamalarında MPS'yi belirlemede zaman bağımlı MTK modellerinden (time-dependent IRT model) KPC (Kim, Park, Cohen procedure), CUSUM (Cumulative Sum) ve BILOG-MG (Zimowski, Muraki, Mislevy ve Bock, 2003) modellerini karşılaştırmıştır. Çalışmada üç parametrelili lojistik model (3PLM) kullanılmıştır. 100 maddelik madde bankasında 10 maddenin madde güçlüğü 0.25, 0.50 ve 1.00 logit, madde ayırt ediciliği de 0.05 ve 0.10 logit büyüklüğünde MPS gösterecek şekilde simüle edilmiştir. Veriler beş zaman noktasında toplanmış ve üç, dört ve beşinci zaman noktalarında MPS değerlendirilmiştir. MPS koşulları doğrusal, düzgün olmayan ve ani değişim gösterecek şekilde düzenlenmiştir. Tüm koşullar için KPC ve BILOG-MG yöntemlerinin CUSUM yöntemine göre madde ayırıcılık ve madde güçlük parametrelerindeki MPS'yi belirlemede daha iyi performans gösterdiği saptanmıştır.

Wollack ve diğerleri (2005) yaptıkları çalışmada MPS'nin boylamsal etkisini araştırmışlardır. Araştırmada her yıl uygulanan ve yaklaşık 55 maddesi olan Almanca Yerleştirme Testi (German Placement Test) kullanılmıştır. Bu teste ilişkin yedi zaman noktasında elde edilen 750-1000 kişiye ait veriler kullanılmıştır. Araştırmada 10 farklı ölçekleme modelinde MPS'nin etkisi araştırılmıştır. MPS'nin yetenek kestirimlerine etkisini değerlendirmede modellere ilişkin korelasyonlar, ortalama farklar ve ortalama

farkların karekökü (root mean square difference) değerleri kullanılmış ve 3PLM'ye göre değerlendirmeler yapılmıştır. Araştırmanın bulguları, farklı ölçekleme modellerindeki MPS'nin yetenek kestirimleri üzerinde etkisi olduğunu ortaya koymuş, ancak modeller arasındaki fark büyük olsa da, pratikte bu farkların düşük olduğu belirtilmiştir. Bir başka deyişle, MPS gösteren maddelerin sayısının ve MPS miktarının düşük olduğu belirtilmiştir. Bunun yanı sıra gözlenen ve beklenen yetenek değerleri arasındaki farkın düşük çıkması, MTK modellerinin MPS'ye yeterince dayanıklı olduğunu ve MPS olduğunda bile güvenilir bir şekilde yetenek kestirimi yapılabileceğini göstermiştir.

Giordano, Subhiyah ve Hess (2005) öğrencilerin sınıf ortamı dışında yaptıkları (take-home) sınavlarda kopya çekme davranışı olabileceğinden, maddelerin görülme sıklığının MPS bağlamında testin geçerliğine etkisini incelemişlerdir. Çalışma kapsamında sertifika yenilemek için yılda iki kere sınıf ortamı dışında alınan ulusal bir sınavın verileri kullanılmıştır. Bu sınavda bireylerin, cevapları birbirleriyle tartışabilecekleri veya maddeleri kaydedebileceklerinden yola çıkılarak, bu durumun madde güçlük parametrelerini nasıl etkileyebileceği araştırılmıştır. Üç aylık periyotlarda her uygulamada 300 yeni maddenin kullanıldığı bu sınavdan 60 madde seçilmiştir. Seçilen 60 madde, arka arkaya yapılan üç uygulamada 1018 bireye uygulanmış ve madde parametrelerindeki değişim incelenmiştir. Bu bağlamda hem MPS'nin etkisi hem de maddelerin görülme sıklığının geçerliğe etkisi incelenmiştir. Madde parametre kestirimi Rasch model ile yapılmış ve MPS tüm maddeler için kontrol edilmiştir. Araştırma sonunda 60 maddeden 12'sinin (%20) manidar düzeyde MPS gösterdiği bulunmuştur. MPS gösteren maddelerden altı tanesi kolay, altı tanesi zor yönde MPS göstermiştir. Bu kadar az maddede zor yönde MPS gözlenmesi, bireylerin maddeleri sürekli görmesi durumunda bile, bunu kendi lehlerine kullanamadıkları şeklinde yorumlanmıştır. Ayrıca araştırmacılar bu çalışma kapsamında katılımcı sayısının az olması ve elde edilen madde istatistiklerine göre maddelerin tekrarlı kullanılmasının geçerliği etkilediğine dair kesin bir şey söylemenin mümkün olmadığını belirtmişlerdir.

Li (2008) yaptığı bir çalışmada, boyutluluğun MPS üzerindeki etkisini incelemiştir. Ortak maddelerde MPS görülmesi durumunda bunun tek boyutlu MTK modellerinde parametre kestirimlerini ne düzeyde etkilediği ve bu durumun çeşitli boyutluluk koşullarında parametre değişmezliğini ne düzeyde etkilediği araştırılmıştır. Bu bağlamda İngilizce Yeterlik Belgesi Sınavına (Examination for the Certificate of Proficiency in English- ECPE) giren 70.000 kişiye ait veriler kullanılmıştır. Veriler, altı yıllık bir süreç içerisinde üç zaman noktasında elde edilmiştir. 70.000 kişilik veriden her

uygulama için 2000 birey rastgele seçilmiştir. Madde bankasından da farklı kombinasyonlarda maddeler seçilerek bir, iki ve üç boyutlu testler oluşturularak güçlük ve ayırt edicilik değerlerindeki sapmalar incelenmiştir. Analizler 3PLM'ye göre gerçekleştirilmiştir. Elde edilen sonuçlar, madde güçlük parametresinin tek ve çok boyutlu yapılar altında değişmezliğini koruyabildiğini, ancak ayırt edicilik parametresinin değişmezlik özelliğini koruyamadığını göstermiştir. Madde ayırt ediciliğinin değişmezlik özelliği, boyut sayısı azaldıkça bozulmuştur. Bu bulgu, MPS'yi belirlemede boyut sayısının etkili bir faktör olduğunu ortaya koymuştur.

Babcock ve Albano (2012) tarafından yapılan çalışmada MPS olması durumunda Rasch ölçeğinin tutarlılığı incelenmiştir. Simülasyon verilerin kullanıldığı çalışmada örneklem büyüklüğü 500, madde sayısı 220, MPS büyüklüğü (0.00, 0.05, 0.10, 0.20 ve 0.50), MPS yönü (kolay, zor ve her iki yönde), MPS yüzdesi (%0, %1, %5, %10) ve iki zaman noktası koşulları ile çalışılmıştır. Geçme oranı, sınıflama oranı ve birey uyum istatistiklerine ilişkin OHKK ve yanlılık değerleri hesaplanmış ve bunlara ilişkin ANOVA analizleri yapılmıştır. Araştırma sonucuna göre %10 oranında MPS olmasının yetenek kestirimlerini önemli düzeyde etkilediği, MPS miktarı arttıkça hatanın arttığı ve MPS'nin her iki yönde gerçekleşmesi durumunda yanlılığın sıfıra yaklaştığı görülmüştür. Temel düzey veri seti ile karşılaştırıldığında, MPS'nin zor yönde olması durumunda yanlış sınıflama oranının arttığı, kolay yönde olması durumunda daha az yanlış sınıflama yapıldığı, her iki yönde olması durumunda ise sınıflama doğruluğu açısından önemli bir farkın ortaya çıkmadığı belirlenmiştir. Geçme oranları incelendiğinde, kolay yönde sapma olması durumunda bireylerin yetenek kestirimlerinde şişmeler olduğu ve öğrenciler hakkında yanlış geçti kararı alma olasılığının arttığı sonucuna ulaşılmıştır. Buna ek olarak, test uygulamaları arasındaki zaman ve MPS miktarı arttıkça, birey uyum istatistiklerinde çok düşük düzeyde değişimler olduğu belirlenmiştir. Araştırma sonucunda genel olarak, örtük değişkendeki değişimler düşük düzeyde olduğu sürece Rasch modelin bu değişimlere dayanıklı olduğu ve özellikle değişim iki yönlü olduğunda, madde ve birey parametrelerine ilişkin doğru çıktılar verdiği belirtilmiştir. Ancak büyük miktardaki değişimlerin ölçeğin uzun süre kullanımını olumsuz yönde etkileyeceği uyarısında da bulunulmuştur.

Demirus ve Uysal (2016) MPS'nin test eşitleme üzerindeki etkisini Monte Carlo simülasyon çalışması ile incelemiştir. MPS'nin eşitleme üzerindeki etkisi ortalama-ortalama, ortalama-sigma, Haebara ve Stocking-Lord yöntemleri kullanılarak ortaya konmuştur. Araştırmada MPS'nin sadece arttığı ve sadece azaldığı tek yönlü koşul ile

aynı zamanda hem arttığı hem azaldığı çok yönlü koşul simülasyon koşulu olarak ele alınmıştır. MPS'nin olmadığı koşul ile MPS'nin olduğu durumda elde edilen eşitleme hataları (RMSD) karşılaştırılmıştır. Simülasyonda madde güçlük parametresinde 0.50 logit düzeyinde MPS oluşturulmuş ve MTK modellerinden 3PLM kullanılmıştır. Araştırmanın sonucuna göre ortak maddelerin madde güçlük parametresinin azalan yönde MPS gösterdiği durumda en büyük eşitleme hatası Stocking-Lord yönteminde, en düşük hata ise ortalama-ortalama yönteminde elde edilmiştir. Madde güçlük parametresinin artan yönde MPS gösterdiği durumda ise en büyük eşitleme hatası ortalama-sigma yönteminde, en düşük eşitleme hatası da Stocking-Lord yönteminde elde edilmiştir. Madde güçlük parametresinin ortak maddelerde hem artan hem azalan yönde MPS gösterdiği durumda ise en yüksek eşitleme hatası ortalama-sigma yönteminde, en düşük hata da ortalama-ortalama yönteminden elde edilmiştir.

Lee ve Geisinger (2019) MPS'nin etkilerini TIMSS 2011 sınavından aldıkları üç farklı uzunluktaki ölçek üzerinde incelemişler ve MPS'li maddelere ne yapılması gerektiği konusunda en uygun yolu belirlemeyi amaçlamışlardır. Bu amaç doğrultusunda ölçek uzunluğu (3, 6 ve 9 maddelik ölçekler), MPS gösteren madde oranı (3 maddelik ölçek için %34 büyük; 6 maddelik ölçek için %12-17 düşük ve %34 büyük; 9 maddelik ölçek için de %12-17 düşük, %22 orta, %34 büyük), MPS'nin yönü (negatif, pozitif ve her iki yönde), MPS'nin büyüklüğü (0.10 logit düşük, 0.20 logit yüksek) ve MPS hakkında verilecek karar (MPS gösteren maddenin tutulması, MPS gösteren maddenin düzeltilmesi, MPS gösteren maddenin çıkartılması) faktörlerini değişimlemiştir. Araştırmanın sonucuna göre madde sayısı az olan ölçeklerde yetenek kestirimi ve bireylerin sınıflandırılmasında MPS'nin önemli düzeyde etkisi olduğu belirtilmiştir. Özellikle yetenek tahminlerinin uç noktalarda olduğu durumda, sınıflama doğruluğunun azaldığı bulgulanmıştır. Bu doğrultuda özellikle kısa ölçeklerde MPS gösteren maddelerin revize edilmesinin, MPS'nin etkilerini azaltmak bakımından önemli olduğu belirtilmiştir.

Sabit uzunluktaki test uygulamalarında MPS'nin incelendiği çalışmalarda MPS'nin yetenek kestirimlerini, dolayısıyla ölçme kesinliği (Babcock ve Albano, 2012; Bock vd., 1988; Chen vd., 1999; Giordano vd., 2005; Wells vd., 2002) ve test eşitlemeye etkisinin (Demirus ve Uysal, 2016) önemli düzeyde olduğu, bu bulguların yanı sıra özellikle madde sayısının kısa olduğu likert tipi ölçeklerde MPS'nin etkilerinin daha çok olduğunu (Lee ve Geisinger, 2019) belirten çalışmalar mevcuttur. Bunun yanında MTK ölçeğinin MPS'ye dayanıklı olmasından dolayı yetenek kestirimlerinin önemli düzeyde

etkilenmediğini gösteren çalışmalar da literatürde yer almaktadır (Li, 2008; Wollack vd., 2005). İlgili çalışmalarda farklı bulguların elde edilme nedeni farklı sınavlar, farklı örneklem büyüklüğü, madde bankası büyüklüğü, MPS koşulları ve MTK modelleri ile çalışılmış olması olabilir.

Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulamalarında Madde Parametre Sapmasının İncelendiği Araştırmalar

MPS'nin yetenek ve madde parametrelerine etkileri, MPS olduğu durumda MTK modellerinin sürekliliği ve MPS belirleme yöntemleri gibi çalışmalar genellikle sabit test koşulu altında araştırılmıştır. Ancak bu yöntem ve teknikler BOBUT uygulamaları için uygun değildir. Bu nedenle özellikle BOBUT uygulamalarında MPS'nin araştırıldığı çalışmalar da yapılmıştır. Burada, alanyazında BOBUT ve MPS'nin birlikte incelendiği çalışmalar kronolojik olarak özetlenmiştir.

Guo ve Wang (2003) bir test programı için puanların karşılaştırılabilir olmasının önemini vurgulamışlar ve puanların karşılaştırılmasında ölçek tutarlılığının önemli bir faktör olduğunu belirterek ölçek kaymasının (scale drift) BOBUT uygulamasına etkisini incelemişlerdir. Çalışma hem gerçek hem simülatif verilerle yürütülmüş olup, yetenek kestirimlerindeki yanlışlık ve test puanlarındaki değişim hesaplanmıştır. Gerçek veri ile yapılan uygulamada bir BOBUT uygulamasından 31 madde, 20 ay ara ile iki zaman noktasında uygulanmış ve kalibre edilmiştir. Simülatif veri ile yapılan uygulama için gerçek uygulamadan elde edilen yetenek dağılımları kullanılarak veriler üretilmiş ve madde güçlük parametresinin kolay yönde değiştiği durum ele alınmıştır. Değerlendirmede yanlışlık değerleri, test ve madde karakteristik eğrileri karşılaştırılmıştır. Araştırma sonucunda düşük miktarda yanlışlık gözlemlendiği ve bunun pratik anlamda önemli olmadığı belirtilmiştir. Bunun yanında ölçek kaymasının test puanlarını etkilediği ancak iki zaman noktası arasındaki bu değişimin çok düşük düzeyde olduğu belirlenmiştir. Araştırmacılar, test puanları arasındaki değişim düşük olsa da test karakteristik eğrilerindeki değişimler, iki ölçüm arasındaki ölçeğin birbirinden çok farklı olması durumunda, ölçme kesinliğinin olumsuz yönde etkileneceğini belirtmişlerdir. Bunun yanı sıra bu durumun BOBUT uygulamalarıyla elde edilen yetenek kestirimlerinin, büyük sapmalara karşı hassas olduğunu belirtmişlerdir.

Deng ve Melican (2009), BOBUT uygulamalarında MPS'yi çoklu zaman noktalarında çalışmışlardır. Çalışma kapsamında okuma, yazma ve matematik becerisini

ölçen ve uyarlamalı test olan ACCUPLACER® testi değerlendirilmiştir. Testi 2004-2007 yılları arasında alan 800.000'in üzerinde bireyin verileri kullanılmış ve maddelerin görülme sıklığında herhangi bir müdahalede bulunulmamıştır. MPS'yi araştırmak için veri seti her yıl kalibre edilmiştir. Sadece 200 ve üzeri birey tarafından yanıtlanan maddeler kalibrasyon için kullanılmış olup, toplam 223 madde kalibre edilmiş ve MPS durumları incelenmiştir. Testin dört zaman noktası için ölçümü alınmış, 3PLM ile analiz edilerek ve a, b ve c parametrelerindeki MPS incelenmiştir. Değerlendirmede madde ve test karakteristik eğrileri karşılaştırılmıştır. Dört yıllık süreç içerisinde çok az sayıda maddede MPS olduğu belirlenmiş, ancak hiçbir madde sık görülmesinden kaynaklı MPS göstermemiştir. Araştırmacılar bunun nedeninin, testin kritik kararların alınmasında kullanılmayan (low-stake test) bir test olmasından kaynaklı olduğunu belirtmişlerdir.

Jiang, Tay ve Drasgow (2009) geleneksel testler ve BOBUT uygulamalarında MPS'yi etkileyen faktörlerden biri olan maddelerin hatırlanma durumunu (memorizing) çalışmışlardır. Çalışma simülatif verilerle yürütülmüş olup, veriler 1000 kişi ve 435 maddelik madde bankası şeklinde üretilmiştir. Madde bankasından rastgele maddeler seçilerek 30 ve 60 maddelik testler oluşturulmuştur. 30 maddelik test için 5 madde, 60 maddelik test için 10 madde öğrenciler tarafından hatırlanabilecek maddeler olarak belirlenmiştir. Testler, paralel formlar şeklinde oluşturulmuş ve hem çoklu test formlarının hem test uzunluğunun maddelerin hatırlanma durumu üzerindeki etkisi araştırılmıştır. BOBUT uygulaması için sabit test uzunluğu durdurma kuralı ve altı farklı madde seçim yöntemi ile çalışılmıştır. Analizler 3PLM'ye göre gerçekleştirilmiş, değerlendirme kriteri olarak yanlışlık değerleri karşılaştırılmıştır. Çalışma sonucuna göre test uygulanma sayısının sabit olduğu durumda 60 maddelik uzun testin, 30 maddelik kısa teste göre maddelerin hatırlanma olasılığına karşı daha dayanıklı olduğu bulunmuştur. Geleneksel testlerde daha çok madde içeren testin ve paralel formların test güvenliğini arttırdığı belirtilmiştir. BOBUT uygulamalarında da her bireyin farklı maddeleri alması ve bazı madde seçim yöntemlerinin (a-tabakalama ve en yüksek madde bilgisi) maddelerin hatırlanma oranını düşürmesi bakımından daha iyi performans gösterdiği belirtilmiştir. Sonuç olarak BOBUT uygulamalarının, kopya veya maddelerin hatırlanma durumuna geleneksel test uygulamalarına göre daha dayanıklı olduğu vurgulanmıştır.

McCoy (2009) BOBUT uygulamalarında, madde bankasında MPS'li madde olması durumunda, bu durumun yetenek ve madde parametre kestirimlerine etkisini Rasch modelin uzantısı olan Boylamsal Rasch Model kullanarak incelemiştir. Çalışma kapsamında Amerikan Klinik Patoloji Derneği- Tıbbi Tekniker (American Society for

Clinical Pathology- Medical Technologist ASCP- MT) sınavının 1993-2003 yılları arasına ait verileri kullanılmıştır. Çalışma kapsamında 2555 kişi ve yedi alt ölçekten oluşan 1270 maddelik madde bankası kullanılmıştır. Araştırma sonucuna göre Boylamsal Rasch Modelin, Rasch modele göre daha doğru yetenek ve madde parametre kestirimi yaptığı sonucuna ulaşılmıştır. Özellikle zaman faktörünün de modele eklendiği durumda, Boylamsal Rasch Modelin daha iyi model-uyum indeksleri verdiği görülmüştür. Boylamsal Rasch model ile önemli düzeyde MPS gösteren maddelerin bir başka deyişle madde güçlük parametresi değişmezliği varsayımının sağlanmadığı durumların belirlenebileceği vurgusu yapılmıştır. Boylamsal Rasch Model ile hem MPS gösteren maddeler belirlenebilmekte hem de MPS gösteren maddeler kontrol değişkeni olarak ele alınarak yetenek kestirimi yapılmaktadır.

Abad, Olea, Aguado, Ponsoda ve Barrada (2010) MPS'yi, BOBUT uygulaması olan İspanyolca Konuşanların İngilizce Düzeyi (English level of Spanish speakers (eCAT)) sınavında 3PLM kullanarak araştırmışlardır. Çalışma kapsamında 7254 kişi ve 3224 maddelik madde bankası kullanılmıştır. Test birçok zaman noktasında uygulanarak ve madde bankasına yeni eklenen maddelerle eski maddeleri kalibre ederek değişen madde fonksiyonuna ilişkin analizler yapılmıştır. Yeni eklenen maddelerle yapılan kalibrasyon sonucu a ve b parametrelerindeki değişimden, yetenek kestirimlerinin orta düzeyde etkilendiği, çok az maddenin ise a ve c parametrelerinde önemli düzeyde MPS oluştuğu gözlenmiştir. Bu nedenle de madde bankasının bu tür değişimlere karşı kontrol edilmesi gerektiği vurgusu yapılmıştır.

Hagge ve diğerleri (2011) BOBUT uygulamalarında farklı uzunluktaki testlerde, madde güçlük parametresinde MPS olmasının bireylerin yetenek kestirimlerine etkisini Rasch model kapsamında incelemişlerdir. Çalışmada bireylerin yetenek kestirimlerinin MPS'nin etkisine ne kadar dayanıklı olduğu ve MPS'nin geçti-kaldı kararlarını ne ölçüde etkilediği araştırılmıştır. Çalışma kapsamında iki zaman noktasında 18.004 ve 52.765 kişinin katıldığı geniş ölçekli iki uygulamanın verileri kullanılmıştır. Madde bankasındaki maddeler rastgele seçilerek MPS durumu oluşturulmuştur. MPS koşulları; maddelerin %5, %10 ve %20 oranında, 0.50, 0.75 ve 1.00 logit büyüklüğünde ve MPS yönünün kolay, zor ve her iki yönde olduğu şeklinde oluşturulmuştur. Araştırma sonucuna göre MPS büyüklüğü ve MPS gösteren madde yüzdesi arttıkça yetenek kestirimleri arası fark, MPS olmayan yetenek kestirimine göre artmaktadır. En büyük fark hem yetenek kestirimi hem de geçti-kaldı kararları için maddelerin %20'sinde 1.00 logit büyüklüğünde MPS olan durum için elde edilmiştir. MPS yüzdesi ve büyüklüğü arttıkça yetenek kestirimleri

açısından fark artsa da, madde bankası büyükse, yetenek kestirimlerinin bu duruma daha dayanıklı olabileceği vurgusu yapılmıştır.

Han ve Guo (2011) pratik yapma ve müfredat değişiminden kaynaklı MPS'yi BOBUT bağlamında çalışmışlardır. Araştırmada MPS'nin madde kalibrasyonu ve yetenek kestirimlerine etkisi incelenmiştir. Çalışmada hem gerçek hem simülatif veri kullanılmıştır. Gerçek uygulamalarda MPS'nin tüm bireylerde değil, testi alan bazı bireylerde görülebileceğinden yola çıkılarak, bazı bireylerin aldığı maddelerde MPS görülmesi durumu incelenmiştir. Mezun Yönetimi Kabul Testi'nin (Graduate Management Admission Test- GMAT) madde bankasından 1000 madde seçilmiş ve bu maddelerden de 100 tanesi ön test maddesi olarak belirlenerek 3PLM'ye göre kalibre edilmiştir. Çalışmada belirli bir grup bireyin aldığı maddelerin %10, %20, %30, %40 ve %50'sinde MPS oluşturulmuş, diğerlerinin aldığı maddelerde ise oluşturulmamıştır. MPS gösteren maddelerin, madde güçlük parametresinde 0.50 logit büyüklüğünde MPS olacak şekilde düzenlenmiştir. MPS'nin olmadığı koşul temel düzey veri seti olarak değerlendirilmiştir. Bu sınava katılan 50.000 kişi için iki zaman noktasında ölçüm alınmıştır. Çalışma sonucuna göre MPS'nin madde kalibrasyonu ve yetenek kestirimleri üzerinde etkisinin yüksek düzeyde olduğu, ancak bu etkinin istatistiksel olarak manidar olmadığı saptanmıştır. MPS'nin etkisinin iki zaman noktasında çalışılmasından dolayı bu etkinin manidar olmadığı, pratik yapma ve müfredattan kaynaklı MPS'nin etkisi daha uzun vadede görülebileceğinden, bu durumun uzun süreli etkisinin incelenmesi gerektiği belirtilmiştir.

Zhang (2014), BOBUT uygulamasında madde bankasında gizliliği ifşa olmuş bir grup maddeyi bir sıralı izleme süreci (sequential monitoring procedure) ile belirlemeye yönelik bir çalışma gerçekleştirmiştir. 400 maddeden oluşan madde bankasının kullanıldığı geniş ölçekli bir sınavın verileri kullanılmış, normal dağılım gösteren 10.000 kişilik veri simüle edilmiş ve iki farklı simülasyon çalışması yapılmıştır. Sonuçların değerlendirilmesinde Tip-I ve Tip-II hata değerleri ele alınmıştır. Araştırmanın sonucuna göre sıralı izleme sürecinin MPS'li maddeleri belirlemede Tip-I hatayı kontrol altında tutabilirken, Tip-II hatayı çok düşük düzeyde kontrol altına alabildiği belirlenmiştir.

Risk (2015) BOBUT uygulamasında MPS'nin yetenek kestirimlerine etkisini çeşitli simülatif koşullar altında incelemiştir. Çalışmada Rasch model kullanılmış olup MPS'nin ölçme kesinliği ve test etkililiği üzerindeki etkisi incelenmiştir. Çalışma 500 kişilik örneklem ile yürütülmüş ve değişimleme yapılan koşullar; madde bankası büyüklüğü (300, 500, 1000), MPS büyüklüğü (0.50, 0.75, 1.00), MPS gösteren madde

sayısı (50, 75, 100), MPS yönü (%75 zor, %25 kolay yönde olacak şekilde iki yönlü) olarak belirlenmiştir. Çalışma sonucunda ortaya çıkan en önemli bulgu; MPS gösteren madde sayısından çok MPS büyüklüğünün ölçme kesinliği üzerinde daha büyük etkisi olduğudur. Ancak tüm koşullardan elde edilen bulgular genel olarak değerlendirildiğinde, temel düzey veri seti ile MPS oluşturan koşullar arasında ihmal edilir düzeyde farklar ortaya çıktığı sonucuna ulaşılmıştır.

Aksu Dünya (2017) BOBUT ortamında MPS'nin yetenek kestirimleri ve sınıflama doğruluğuna etkisini, MPS'nin alt grupları etkilediği koşul altında incelemiştir. BOBUT uygulamasında MPS'den etkilenen birey yüzdesi (%20, %30, %40, %50), madde bankasındaki MPS'li maddelerin içerik alanı yüzdesi (%20, %40, %60) ve madde bankasındaki MPS'li maddelerin hedef aldığı grup (orta ve düşük yetenek düzeyine sahip bireyleri hedef alan maddeler) değişkenleri simüle edilmiştir. Yanlılık, ortalama hataların karekökü, maksimum ortalama fark ve korelasyon değerleri değerlendirme kriteri olarak ele alınmıştır. Araştırmanın bulgularına göre belirli bir grup birey MPS'li maddelere maruz kaldığı durumda sınıflama doğruluğu önemli düzeyde etkilenirken, ortalama yetenek tahminleri MPS'den daha az etkilenmiştir.

BOBUT uygulamalarında MPS'nin araştırıldığı çalışmalar incelendiğinde çoğunlukla MPS'nin madde ve yetenek parametreleri üzerindeki etkisi üzerine odaklanıldığı görülmektedir. Bu çalışmaların bir kısmında MPS'nin BOBUT aracılığı ile elde edilen yetenek kestirimlerine önemli düzeyde etkisi olduğu belirlenirken (Abad vd., 2010; Hagge vd., 2011; Risk, 2015), yetenek kestirimleri üzerindeki etkisinin düşük ve önemli olmayan düzeyde olduğunu belirten çalışmalar da bulunmaktadır (Aksu Dünya, 2017; Deng ve Melican, 2009; Guo ve Wang, 2003; Han ve Guo, 2011; Jiang vd., 2009; McCoy, 2009). MPS'nin yetenek kestirimleri üzerinde düşük düzeyde etkisinin olduğu belirlenen çalışmalarda bunun nedeninin ya iki zaman noktası gibi az bir noktada ölçüm alınmasından ya da MTK modellerinin MPS koşullarına dayanıklı olmasından kaynaklı olduğu belirtilmiştir.

Hem BOBUT hem sabit uzunluktaki koşullar için yapılan araştırmalar genel olarak gözden geçirildiğinde, MPS çalışmalarının daha çok maksimum performansın ölçüldüğü testlerde çalışıldığı görülmektedir. Bu araştırmalarda çoğunlukla MPS'nin yetenek parametresi, madde parametreleri ve madde kalibrasyonuna etkisi ve MPS belirleme yöntemleri incelenmiştir. Bunun yanı sıra MPS bağlamında maddelerin görülme sıklığı ve kopyanın etkileri de incelenmiştir. Belirtilen çalışmalar genellikle sabit uzunluklu (fixed-length) veya sabit maddeli (fixed-item) testler için yapılmıştır. Ancak

bu çalışmalarda kullanılan süreçlerin çoğu BOBUT uygulamaları için uygun değildir. Bunun yanında bu çalışmalarda genellikle örneklem büyüklüğü ve/veya madde bankası büyüklüğü sabit tutulmuş olup MPS'nin sadece yüzdesi, büyüklüğü, yönü gibi faktörler ya tek tek ya da ikili kombinasyonlar şeklinde incelenmiştir. Özellikle BOBUT uygulamaları ile yapılan çalışmalarda MPS'nin etkisi çoğunlukla iki zaman noktası için incelenmiştir. Ancak MPS'nin etkilerinin görülebilmesi için ikiden fazla noktada ölçüm almak gerekmektedir. Ulaşılabilen alanyazında örneklem büyüklüğü, madde bankası büyüklüğü ve MPS'nin farklı koşullarını ele alarak bu durumun yetenek kestirimleri, test bilgi fonksiyonu (TBF) ve test etkililiğine etkisini inceleyen bir çalışmaya rastlanmamıştır. Bu nedenle MPS'nin BOBUT uygulamalarında yukarıda belirtilen faktörler üzerindeki etkisi tam olarak bilinmemektedir. Bu nedenle bu çalışmada, BOBUT uygulamalarında madde parametre sapmasının ölçme kesinliği, test bilgi fonksiyonu ve test etkililiği üzerindeki etkisi araştırılmıştır.

Amaç

Bu çalışmanın genel amacı BOBUT uygulamalarında madde parametre sapması gösteren maddelerin, ölçme kesinliği, test bilgi fonksiyonu ve test etkililiği üzerindeki etkisini incelemektir. Bu genel amaç doğrultusunda aşağıdaki araştırma sorularına yanıt aranmıştır;

1. Örneklem büyüklüğünün 1000, MPS büyüklüğünün 0.00, 0.50, 0.75, 1.00 logit; MPS gösteren madde oranının %0, %5, %10, %20; madde bankası büyüklüğünün 200, 500, 1000 olduğu ve üç zaman noktasında ölçüm alındığı durumda, BOBUT uygulamalarında ölçme kesinliği, test bilgi fonksiyonu ve test etkililiği değerleri nasıl değişmektedir?

2. Örneklem büyüklüğünün 5000, MPS büyüklüğünün 0.00, 0.50, 0.75, 1.00 logit; MPS gösteren madde oranının %0, %5, %10, %20; madde bankası büyüklüğünün 200, 500, 1000 olduğu ve üç zaman noktasında ölçüm alındığı durumda, BOBUT uygulamalarında ölçme kesinliği, test bilgi fonksiyonu ve test etkililiği değerleri nasıl değişmektedir?

Önem

Büyük madde bankalarının kullanıldığı ve özellikle test alanlar hakkında önemli kararların verildiği test uygulamalarında MPS görülmesi, yapı ile ilişkili olmayan değişkenlerin ölçme sonuçlarına karışmasına neden olarak geçerliği düşürür. Bu durum, maddelerin daha kolay/zor görünmesi ve/veya ayırıcılığının düşmesine neden olur (Donoghue ve Isham, 1998). Bunun yanı sıra bazı kavramlar teknolojik, kültürel, sosyal değişimlerden dolayı zamanla anlamlarını kaybetmekte veya anlamları değişmektedir (Roznowski, 1989). Bu tür nedenler, maddelerin veya testlerin “*kullanım ömrü (shelf life)*” kavramını gündeme getirmektedir. Bu durum, madde veya testlerin zamanla kullanışlılığını kaybettiğini göstermektedir. Özellikle seçme amaçlı kullanılan testlerin psikometrik özelliklerinin incelenmesi, testin geliştirilmesi ve standart test olarak kullanımının devam etmesi açısından önemlidir. Örneğin, 25 maddelik bir testte, maddelerin zamanla kolaylaşması test puanlarında 1-2 puanlık kaymalara yol açabilmekte ve seçme kararlarını önemli düzeyde etkileyebilmektedir. Bu durum, özellikle de test alanın puanının kesme puanına yakın olması durumunda daha da önem kazanacaktır.

Madde parametre sapması, bireylerin yetenek kestirimleri ve geçti-kaldı kararlarının verilmesinin doğruluğunu etkileyen bir durumdur. MPS’nin yetenek kestirimleri ve test etkililiği üzerindeki etkisinin incelenmesi, sınavların bir bütün olarak güvenli bir şekilde birleştirilmesi ve test puanlarından yapılacak çıkarımların geçerliği açısından önemlidir. BOBUT’un kullanımı yaygınlaşmakla birlikte, bu uygulamalarda MPS olması, yetenek kestirimlerinin doğruluğunu ve test puanlarından yapılan çıkarımların geçerliğini olumsuz yönde etkiler. Bu sebeple BOBUT uygulamalarında, bu durumu etkileyen koşulların belirlenmesi, madde ve yetenek parametrelerinin en az hatalı şekilde kestirilmesi açısından önemlidir.

Madde parametre sapması içeren maddeler, farklı test uygulamalarına katılan gruplar arasında farklı performans gösterebilir. Bu durum, BOBUT uygulamaları için önemli bir problemdir; çünkü MTK’nın temel varsayımı olan “aynı yetenek düzeyindeki bireylerin, bir maddeyi doğru cevaplama olasılıkları da aynıdır” şeklinde ifade edilen değişmezlik varsayımının ihlaline neden olmaktadır (Babcock ve Albano, 2012). Bu durum özellikle puanlara ilişkin ölçme kesinliğini ve puanların yorumlanmasında geçerliği olumsuz yönde etkilemektedir (Risk, 2015). Bu nedenle, özellikle BOBUT

uygulamalarında madde bankasına ilişkin MPS çalışmalarının yapılması, yapı geçerliğine tehdit oluşturan bu durumun önüne geçilmesi açısından önemlidir (Wainer vd., 2010).

Bu çalışmanın, BOBUT uygulamalarında çeşitli koşullardaki MPS'nin, yetenek kestirimleri üzerindeki etkisinin incelenmesi bakımından alanyazına katkı sağlaması beklenmektedir. Bu çalışmanın sonuçları, gelecek BOBUT uygulamaları için madde güçlük parametresindeki sapmaların yönü, miktarı ve büyüklüğünün, yetenek kestirimleri, TBF ve test etkililiğini ne denli etkilediğinin görülmesi açısından önemlidir. Bu doğrultuda araştırmanın bulgularının, ölçme ve değerlendirme alanında hizmet veren kurum ve kuruluşlara BOBUT'un organizasyonu, madde bankasının sürdürülebilirliği ve güncellenmesi konusunda psikometrik bilgi sağlaması beklenmektedir.

BOBUT uygulamalarında MPS'nin yetenek ve madde parametreleri kestirimine etkisini gösteren çalışma sayısı çok azdır (Aksu Dünya, 2017; Chan vd., 1999; Han ve Guo, 2011; Risk, 2015; Wells vd., 2002). Belirtilen çalışmaların genelinde örneklem büyüklüğü ve/veya madde bankası büyüklüğü sabit tutulmuş olup genellikle iki zaman noktasında ölçüm alınan değerler karşılaştırılmıştır. Bunun yanında MPS'nin BOBUT uygulamalarında parametre kestirimlerindeki etkisi de tam olarak bilinmemektedir. MPS'nin olması ölçmenin doğruluğu ve güvenilirliğine bir tehdittir ve bireylerin yanlış sınıflanmasına neden olur ve bu durum da sınavların geçerliğini olumsuz etkiler (Risk, 2015). Bu çalışma, BOBUT uygulamalarında, testin psikometrik özelliklerini etkileyebilecek olası MPS koşullarının belirlenmesi açısından önemlidir.

Sınırlılıklar

Araştırma Rasch model kapsamında üretilmiş veriler ile sınırlıdır. Yalnızca bu model ile sınırlı tutulma gerekçesi; araştırma kapsamında örneklem büyüklüğü, madde bankası büyüklüğü, ölçüm alınan zaman noktası ve madde parametre sapmasına ilişkin birçok değişken koşul ile çalışılması ve bu nedenle de analizlerin güçleşmesi ve harcanan sürenin çok fazla olmasıdır. Analiz süresinin daha fazla uzamaması için de yalnızca madde güçlük parametresinde sapma olan durum ile sınırlandırılmıştır.

BÖLÜM 2

YÖNTEM

Bu bölümde araştırma araştırma modeli ile verilerin üretilmesi ve analizi ile ilgili bilgiler yer almaktadır.

Araştırma Modeli

Bu çalışma simülatif (yapay) verilerle yürütülmüştür. Simülasyon çalışmaları, çeşitli hipotezleri test etmek amacıyla gerçek durumlara uygun olacak şekilde, bilgisayar ortamında veri üretim ve analiz süreçlerini içerir. Bilgisayar yoğunluklu süreçlerin kullanıldığı bu çalışmalar, gerçek yaşam durumlarında yer alan görece daha karmaşık süreçlerin olduğu, uygulama sıkıntılarının yaşandığı veya duruma uygun gerçek verinin bulunmadığı durumlarda sıklıkla tercih edilmektedir (Burton, Altman, Royston ve Holder, 2006; Ranganathan ve Foster, 2003).

BOBUT uygulamalarının MTK'ya dayalı olması, uygulama zorluğu, geniş madde bankası ve geniş örneklem gerektirmesi gibi nedenlerle alanyazında yer alan çalışmaların birçoğunda simülatif verilerin kullanıldığı görülmektedir (Barrada, Olea, Ponsoda ve Abad 2010; Kalender, 2011; McDonald, 2002; Patton, Cheng, Yuan ve Diao, 2013; Scullard, 2007; Wang, Kuo, Tsai ve Liao, 2012). Bu araştırma da MTK'ya dayalı BOBUT uygulamasında belirli MPS durumlarının kontrollü koşullar altında incelenmesi nedeniyle simülatif araştırma niteliğindedir.

Verilerin Üretilmesi ve Analizi

Bu araştırmanın verilerinin üretilmesi ve analizinde R programlama dili kullanılmıştır. Veriler R Studio 3.3.2'nin catIrt paketi ile üretilmiş ve analizleri gerçekleştirilmiştir (Nydick, 2015). Veri üretiminde madde ve yetenek parametreleri geniş ölçekli sınavlar ve BOBUT uygulamalarının özellikleri dikkate alınarak düzenlenmiştir. Veri üretiminde ilk üretilen veriler temel düzey yani referans olarak

alınmış (MPS'nin olmadığı koşul) ve MPS'li veri setleri, temel düzey veri seti ile karşılaştırılmıştır.

Madde parametre sapmasına ilişkin koşulların oluşturulmasında MPS büyüklüğü ve MPS yüzdesi ele alınmıştır. Bunun yanında MPS'nin etkisinin daha iyi görülebilmesi için ikiden fazla zaman noktasına ya da ikiden fazla noktada ölçüm alınmasına ihtiyaç olduğundan, veriler üç zaman noktası için düzenlenmiş ve MPS'nin boylamsal etkisinin daha doğru bir şekilde belirlenmesine çalışılmıştır (Babcock ve Albano, 2012; Han ve Guo, 2011; Kim ve Cohen, 1992). Simülatif veri üretiminde dikkate alınan sabit ve değişen koşullara Tablo 1'de yer verilmiştir;

Tablo 1

Simülatif Veri Üretiminde Dikkate Alınan Sabit ve Değişimlenen Koşullar

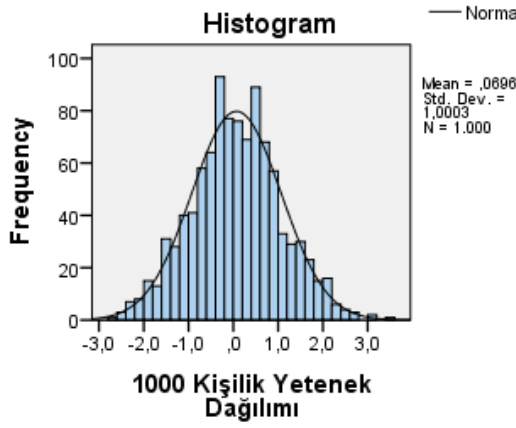
Sabit Koşullar	Değişimlenen Koşullar
1. Yetenek parametreleri dağılımı	1. Örneklem büyüklüğü
2. MTK modeli ve madde parametreleri dağılımı	a. 1000 kişi
3. MPS'nin yönü ve türü	b. 5000 kişi
4. BOBUT Koşulları	2. MPS büyüklüğü
• Yetenek kestirim yöntemi	a. 0.00 logit
• Başlama kuralı	b. 0.50 logit
• Madde seçim yöntemi	c. 0.75 logit
• Durdurma kuralı	d. 1.00 logit
	3. MPS içeren madde yüzdesi
	a. %0
	b. %5
	c. %10
	d. %20
	4. Üç zaman (ölçüm) noktası
	5. Madde bankası büyüklüğü
	a. 200 madde
	b. 500 madde
	c. 1000 madde

Aşağıda öncelikle simülatif veri üretiminde dikkate alınan sabit koşullar ve ardından da değişimlenen koşullar ile ilgili açıklamalara, gerekçeleri ile yer verilmiştir.

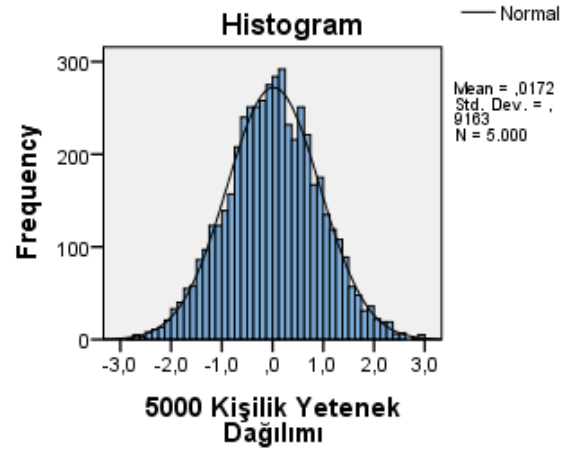
Simülatif Veri Üretiminde Dikkate Alınan Sabit Koşullar

Simülatif veri üretiminde dikkate alınan sabit koşullar aşağıda kısaca açıklanmaktadır.

Yetenek Parametreleri Dağılımı: Veri üretiminde yetenek parametrelerinin dağılımı, ortalaması sıfır, standart sapması bir olan normal dağılımdan üretilmiştir. Normal dağılım seçilmesinin nedeni, yetenek kestirim yöntemi olarak Bayes yöntemlerinden biri seçildiği durumda yetenek parametre dağılımlarının, dağılımı bilinen bir evrenden seçilmesi veya normal dağılım özelliğine sahip olması gerekliliğindendir (Gershon, 2005; Risk, 2015). Bu araştırma kapsamında BOBUT'ta yetenek kestirimi için Bayes yöntemlerinden Beklenen Sonsal Dağılım (BSD) kullanıldığı için, yetenek parametreleri dağılımı normal dağılıma göre üretilmiştir. 1000 ve 5000 kişilik örneklemelere ilişkin yetenek dağılımlarına Şekil 1'de yer verilmiştir.



Şekil 1.a. n=1000



Şekil 1.b. n=5000

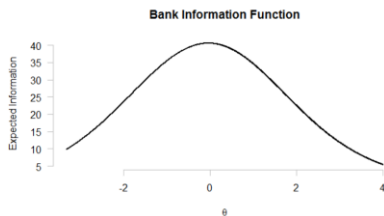
Şekil 1.a ve 1.b. n= 1000 ve n= 5000 Kişilik Verilerde Yetenek Dağılımları

MTK Modeli ve Madde Parametreleri Dağılımı: Uluslararası alanda yapılan TIMSS (Trends in International Mathematics and Science Study), PISA (The Programme for International Student Assessment) gibi geniş ölçekli sınavların analizinde Rasch ölçme modeli kullanılmaktadır (Schulz ve Frallion, 2009). Geniş ölçekli sınavların özellikleri dikkate alınarak yapılan MPS ile ilgili araştırmalarda analizler çoğunlukla Rasch model kullanılarak yapılmıştır (Babcock ve Albano, 2012; Bergstrom, Stahl ve Netzky, 2001; Hagge vd., 2011; Jones ve Smith, 2006; Kingsbury ve Wise, 2011; McCoy, 2009; Meyers, Miller ve Way, 2009; Witt, Stahl, Bergstrom ve Muckle, 2003). Bu çalışma kapsamında hem geniş ölçekli test uygulamalarında Rasch modelin tercih edilmesi hem de sadece madde güçlük parametresinde MPS olan durumun yetenek kestirimlerine

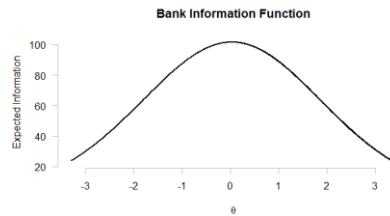
etkisinin incelenmesinden dolayı Rasch model tercih edilmiştir. Rasch (1960) modele göre bir maddeye doğru cevap verme olasılığı şu şekilde hesaplanmaktadır;

$$P(X_{ni}=1|\theta_n, \delta_i) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}$$

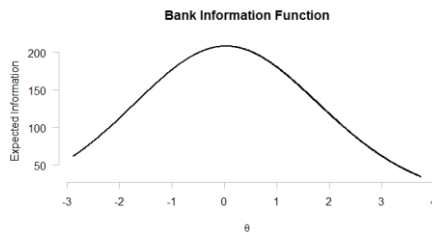
Formülde P, n. bireyin i maddesine doğru cevap verme olasılığını; δ_i , i maddesine ilişkin madde güçlük değerini; θ_n ise n. bireyin yetenek düzeyini göstermektedir. Formülde görüldüğü üzere Rasch model kullanıldığı durumda sadece madde güçlük parametresinde sapma olan durum incelenebilmektedir. Madde güçlük parametresinde sapma olması ise maddenin zamanla, orijinal madde parametresine göre daha kolay ya da daha zor olması anlamına gelir (Risk, 2015). İlgili alanyazında yapılan çalışmalar ve BOBUT uygulamalarının özellikleri ele alındığında (Filho, Machado ve Damaiso, 2014; Svetina vd., 2013), 200, 500 ve 1000 maddelik madde bankaları için madde güçlük parametresi dağılımı, ortalaması sıfır, standart sapması bir olan normal dağılımdan üretilmiştir. 200, 500 ve 1000 maddelik madde bankalarına ilişkin, madde bankası bilgi fonksiyonlarına Şekil 2.a, 2.b ve 2.c’de yer verilmiştir.



Şekil 2.a. k=200



Şekil 2.b. k=500



Şekil 2.c. k=1000

Şekil 2.a, 2.b ve 2.c. k= 200, k= 500 ve k= 1000 Maddelik Madde Bankaları için Madde Bankası Bilgi Fonksiyonu Grafikleri

MPS’nin Yönü ve Türü: MPS’nin yönü, parametre sapmasının hangi yönde olduğunu göstermektedir. Sapmanın yönü kolay (negatif), zor (pozitif) veya her iki yönde

(hem pozitif hem negatif yönde) de olabilir. İlgili alanyazında yapılan araştırmalara göre sapmanın yönü genellikle “kolay” veya “her iki yönde” olmaktadır (Babcock ve Albano, 2012; Hagge vd., 2011; Stahl ve Muckle, 2007). Madde parametreleri içerisinde madde güçlük parametresinin, kolay yönde sapma gösterdiği durumla daha sık karşılaşılmaktadır (Babcock ve Albano, 2012; Hagge vd., 2011; Risk, 2015; Stahl ve Muckle, 2007). Bunun yanı sıra maddelerin görülme sıklığı ve/veya kopya gibi faktörlerle sıklıkla karşılaşılmakla birlikte, bu faktörlerin daha çok kolay yönde sapmaya neden olduğu ve tek yönlü sapmanın da iki yönlü sapmaya göre yetenek kestirimlerini daha olumsuz etkilediği belirtilmektedir (Risk, 2015; Wells vd., 2012). Bu nedenle çalışma kapsamında kolay yönde sapma olan durum ele alınmıştır. MPS’nin türü açısından da madde ayırıcılık parametresinin sapma gösterdiği durumların hem belirlenmesinin zor olması hem de MPS’nin madde güçlük parametresini madde ayırıcılığından daha çok etkilemesi nedeniyle, madde güçlük parametresinde sapma olan durumların daha önemli olduğu belirtilmektedir (Bock vd., 1988; Donoghue ve Isham, 1998; Song ve Arce-Ferrer, 2009). Bununla birlikte araştırmada Rasch model kullanıldığı için yalnızca madde güçlük parametresinde sapma olan durum ele alınmıştır.

BOBUT Koşulları: Veriler üretilerek MPS koşullarının oluşturulmasının ardından, BOBUT uygulaması çalıştırılmıştır. BOBUT uygulaması için tercih edilen yetenek kestirim yöntemi, başlama kuralı, madde seçim yöntemi ve durdurma kuralı ile ilgili açıklamalara aşağıda yer verilmiştir;

- 1. Yetenek Kestirim Yöntemi:** BOBUT uygulamalarında sıklıkla kullanılan yetenek kestirim yöntemleri En Yüksek Olabilirlik Yöntemi (EYOY), Beklenen Sonsal Dağılım (BSD) ve En Yüksek Sonsal Dağılım (EYSD) yöntemleridir. BSD ve EYSD yöntemleri, Bayes kestirim yöntemleridir. Hangi yetenek kestirim yönteminin daha iyi sonuç verdiği BOBUT uygulamasının diğer koşullarına göre değişmektedir (Linden ve Glass, 2002). Bu yetenek kestirim yöntemlerinin birbirleri ile karşılaştırıldığı bazı araştırmalar mevcuttur (Kezer, 2013; Kingsbury ve Zara, 2009; Wang vd., 2012). Araştırmaların büyük bir kısmında BSD yetenek kestirim yönteminin diğer iki yöntemle göre daha iyi sonuçlar verdiği (Eroğlu, 2013; Kezer, 2013; Keller, 2000; Kingsbury ve Zara, 2009; Wang vd., 2012), EYOY yöntemine göre daha düşük standart hata (Wang, 1997) ve yanlılık değeri verdiği belirtilmiştir (Eroğlu, 2013). EYOY yöntemi, BSD ve EYSD yöntemlerine göre daha fazla madde ile yetenek kestirimi yaptığı için etkili bir yöntem olarak belirtilmemiştir (Kezer, 2013). Ayrıca EYOY yöntemi, tümü doğru ya da tümü yanlış yanıt örüntülerinde ve

büyük örneklemelerde tutarlı sonuçlar vermemesi nedeniyle problem oluşturmaktadır. Bunun yanında BSD yöntemi, EYOY ve EYSD yöntemleri gibi iteratif bir yöntem olmamasından dolayı kolay ve hızlı hesaplama yapmaktadır (Embretson ve Reise, 2000). Ayrıca Sulak (2013) tarafından yapılan araştırmada farklı madde seçim yöntemlerine ait standart hata değerleri, BSD yetenek kestirimi kullanıldığı durumda düşük değerler verdiği için BSD yetenek kestirim yönteminin kullanılması önerilmiştir. Bu nedenlerle bu çalışma kapsamında, BOBUT uygulamasında yetenek kestirim yöntemi olarak BSD yöntemi kullanılmıştır.

2. **Başlama Kuralı:** Başlama kuralı, teste başlamak için ilk maddenin seçim ölçütünü kapsar. İlk maddenin seçimi için çeşitli başlangıç kuralları vardır ve ilk maddenin seçimi, yetenek kestirim yönteminin EYOY veya Bayes yaklaşımı olmasına göre değişmektedir. EYOY kullanıldığı durumda seçilecek ilk madde θ değeri 0.00'a karşılık gelen veya ortalama güçlük derecesine sahip olan maddedir. Bayes yaklaşımı kullanıldığında ise, bireylerin gerçek yetenekleri kestirilmeden önce uygulanan ön testten başlangıç θ düzeyi kestirilir. Böylece uygulanacak ilk madde, başlangıç θ düzeyinde en fazla bilgi veren madde olacaktır (Eroğlu, 2013; Mills ve Stocking, 1996; Kezer, 2013; Segall, 2004). Bu doğrultuda, bu çalışma kapsamında yetenek kestirim yöntemi olarak BAYES yöntemlerinden BSD kullanıldığı için başlangıç kuralı olarak bireylerin ön testten aldıkları puanlara göre önsel θ dağılımları kullanılmıştır.
3. **Madde Seçim Yöntemi:** BOBUT uygulamalarında kullanılan temel madde seçim yöntemleri; Maksimum Fisher Bilgisi (MFB), Kullbak-Leibler Bilgisi, Aralık Bilgisi Ölçütü, Olabilirlik Ağırlıklı Bilgi Ölçütü, a-tabakalama, Aşamalı Maksimum Bilgi Oranı, En Uygun b Değeridir (Sulak, 2013). Bu yöntemler içerisinde en sık kullanılan yöntem MFB yöntemidir (Van der Linden ve Glas, 2010; Wainer, 2000; Weiss ve Kingsbury, 1984). Bu yöntemlerin performanslarının karşılaştırıldığı bazı araştırmalarda, -a tabakalama yönteminin yüksek ayırcılığa sahip olan maddeleri seçme sorununun önüne geçtiği ve MFB yönteminden daha iyi performans gösterdiği (Chang ve Ying; 1999; Deng, Ansley ve Chang 2010; Sulak, 2013), Kullbak-Leibler yönteminin de yetenek kestiriminde MFB yöntemine göre daha iyi sonuçlar verdiği belirtilmektedir (Barrada vd., 2010; Chen, Ankenmann ve Chang, 2000; Eggen, 1999; Linda 1996; Veldkamp ve van der Linden, 2006; Yao, 2013). Belirtilen çalışmalar incelendiğinde, -a tabakalama ve Kullbak-Leibler madde seçim yöntemlerinin performanslarının, diğer yöntemlere göre daha iyi olduğu görülmektedir. -a

tabakalama yönteminde madde bankasındaki maddeler, madde ayıricılıklarına göre tabakalandırıldıktan sonra bireylere uygulanmakta ve yüksek ayıricılığa sahip maddeler, uygulamanın son aşamasına bırakılmaktadır (Chang, Qian ve Ying, 2001). Ancak bu çalışma kapsamında Rasch model temele alınarak analizler yapıldığı için tüm maddelerin ayıricılık değerleri sabit olduğundan, -a tabakalama yöntemi uygun değildir. Bu nedenle Kullbak-Leibler madde seçim yöntemi tercih edilmiştir. Yöntem, en son yapılan yetenek kestirimine en yakın noktaya göre Kullbak-Leibler yöntemini kullanarak madde seçimi yapmaktadır.

4. **Durdurma Kuralı:** Durdurma kuralı, BOBUT uygulamasının sonlandırılma ölçütünü ifade eder. Farklı sonlandırma kuralları olmasına karşın en sık kullanılan iki yöntem sabit uzunluk ve değişen uzunluk kuralıdır. Sabit uzunluk kuralında, özellikle testin kapsam geçerliğini sağlayabilmek için test, önceden belirlenen madde sayısına ulaştığında sonlandırılırken; değişen uzunluk kuralında, bireylerin yanıt örüntülerine göre sonlanmaktadır (Babcock ve Weiss, 2009; Risk, 2015; Segall, 2004). Standart Hata, Minimum Bilgi ve Teta Değişimi olmak üzere farklı sonlandırma kuralları vardır. Yapılan araştırmalarda değişen uzunluk sonlandırma kuralının, sabit uzunluk sonlandırma kuralına göre daha iyi ölçme kalitesi verdiği (Babcock ve Weiss, 2012; Chang ve Ansley, 2003; Eroğlu, 2013; Yi, Wang, ve Ban, 2001) ve testin ekonomik olmasını sağladığı belirtilmektedir (Eroğlu, 2013). Bu çalışma kapsamında ise durdurma kuralı olarak en düşük sayıda madde kuralı (the minimum number of items rule) ve standart hata yöntemi uygulanmıştır. İlgili alanyazında uygulanan en düşük madde sayısının 10 maddeden az olması durumunda daha yüksek hata ve yanlışlık değerleri elde edildiği (Babcock ve Weiss 2012; Eroğlu, 2013) ve uygulanan en düşük madde sayısının az olması durumunda normal dağılım özelliğinin bozulduğu belirtilmiştir (Blais ve Raiche, 2002) . Bu nedenle bu çalışmada da en düşük sayıda madde kuralı için minimum 10 madde tercih edilmiştir. Standart hata sonlandırma kuralında ise [-3,00; +3,00] yetenek aralığında, standart hatanın 0.40'a eşit veya daha düşük olmasının, ölçme kesinliği açısından uygun olduğu belirtilmiştir (Babcock ve Weiss, 2009; Blaise ve Raiche, 2002). Dolayısıyla bu araştırma kapsamında standart hatanın 0.40'tan düşük olduğu sonlandırma kuralı ele alınmıştır.

Simülatif Veri Üretiminde Değişimlenen Koşullar

Simülatif veri üretiminde değişimleme yapılan koşullar aşağıda kısaca açıklanmaktadır.

Örneklem Büyüklüğü: MTK'da parametre kestirimlerinin doğru bir şekilde yapılabilmesi ve parametrelerin birbirlerinden iyi bir şekilde ayrılabilmesi için geniş bir örneklem gereklidir. Genellikle örneklem büyüklüğünün 1000 veya daha fazla olduğu durumda, kaliteli madde parametreleri elde edilmektedir. Madde parametreleri açısından doğru kestirimler ve ayrımlar yapabilmek için de örneklem büyüklüğünün en az 1000-2000 civarında olması gerekmektedir (Rudner ve Guo, 2011; Stahl ve Muckle, 2007). Bunun yanı sıra Şahin (2012) tarafından çeşitli örneklem büyüklüğü ve test uzunluğunun MTK modelleri açısından (1PLM, 2PLM, 3PLM) karşılaştırıldığı çalışmada, örneklem büyüklüğünün 5000 olduğu durumda en düşük standart hata değerinin elde edildiği belirtilmiştir. Ayrıca örneklem büyüklüğü arttıkça, ölçme ve sınıflama doğruluğunun da arttığını belirten çalışmalar mevcuttur (Rudner ve Guo, 2011). Bu bağlamda bu çalışma kapsamında örneklem büyüklüğünün 1000 olduğu durum küçük örneklem ve 5000 olduğu durum da büyük örneklem olarak ele alınmıştır.

MPS Büyüklüğü: MPS büyüklüğü, yetenek kestirimini etkileyen önemli faktörlerden biridir (Risk, 2015). MPS büyüklüğü testin tek boyutluluğunu olumsuz etkilemekte ve değişmezliğin bozulmasına neden olmaktadır (Li, 2008). MPS'nin çalışıldığı tüm durumlar içerisinde, ölçme kesinliğini en çok etkileyen durumun MPS'nin büyüklüğü olduğu belirtilmiştir. Özellikle 0.50 logit ve üzerinde MPS olan maddelerin parametre kestirimlerine etkisinin daha önemli olduğu belirtilmektedir (Donoghue ve Isham, 1998; Han ve Wells, 2007; Wollack vd., 2005). 0.50 logit'in altında MPS olan maddeler kolay belirlenemediği ve pratikte önemli olmadığı için kullanılmaya devam edildiğinden (Han ve Guo, 2011), bu araştırma kapsamında her maddenin madde güçlük parametresinde 0.00, 0.50, 0.75 ve 1.00 logit büyüklüğünde MPS oluşturularak, bu durumun etkisi incelenmiştir. Bunun yanı sıra MPS gösteren maddelerin kolay yönde sapma göstermesi durumu ele alındığından, MPS büyüklükleri temel düzey veri setindeki madde güçlüklerinden, 0.00, 0.50, 0.75 ve 1.00 logit birimlerinde değerler çıkartılarak oluşturulmuştur.

MPS İçeren Madde Yüzdesi: Madde bankasında MPS içeren madde yüzdesi, yetenek kestirimlerini etkileyen bir diğer koşuldur. Özellikle MPS içeren maddelerin

yüzdesi arttıkça, yetenek kestirimleri olumsuz etkilenmektedir (Hagge vd., 2011; Huang ve Shyu, 2003; Wells vd., 2002). İlgili alanyazında gerek gerçek, gerek simülatif ortamda yapılan çalışmalar incelendiğinde, MPS içeren madde yüzdesinin genelde minimum %5, maksimum %20-%25 oranında olduğu görülmektedir (Hagge vd., 2011; Stahl vd., 2002; Song ve ArceFerrer, 2009; Wells vd., 2002). Bu doğrultuda bu çalışma kapsamında MPS içeren madde yüzdesi %0, %5, %10 ve %20 olarak ele alınmıştır.

Üç Zaman (Ölçüm) Noktası: MPS'nin etkilerinin ortaya çıkması için uzun bir süreç gerekmektedir ve MPS'nin etkilerini iyi bir şekilde görebilmek için ikiden fazla zaman noktası veya ölçüme ihtiyaç olduğu belirtilmektedir (Babcock ve Albano, 2012; Chan vd., 1999; Deng ve Melican, 2009; Kim ve Cohen, 1992). Bu nedenle bu çalışma kapsamında üç zaman noktasında ölçümler alınmıştır. Hem MPS'nin olmadığı 0.00 logit'lik veri setleri hem de MPS'li veri setleri için MPS'nin etkisinin kademeli olarak arttırıldığı üç adet veri seti birinci, ikinci ve üçüncü zaman noktası olarak ele alınmıştır. MPS'nin etkisinin kademeli olarak nasıl arttırıldığına ilişkin bilgi, madde bankası büyüklüğü başlığı altında açıklanmıştır.

Madde Bankası Büyüklüğü: MTK'ya dayalı BOBUT uygulamalarında, kullanım amacına göre her yetenek düzeyine hitap eden kaliteli maddelerin olduğu geniş bir madde bankası gereklidir (Flaughner, 2000). Alanyazında BOBUT ortamında madde bankası büyüklüğü ile ilgili yapılmış olan bazı çalışmalar doğrultusunda (Han ve Guo, 2011; Risk, 2015; Veldkamp ve Linden, 2006; Wise ve Kingsbury, 2000), bu araştırmada kullanılacak madde bankası büyüklükleri sırasıyla 200 küçük, 500 orta, 1000 büyük madde bankası olacak şekilde belirlenmiştir. Madde bankası büyüklüğü olarak bunların seçilme nedeni şu şekilde belirtilebilir;

Madde bankasındaki madde sayısı 200: BOBUT uygulamaları için minimum madde bankası büyüklüğünün 100 olduğu durumda, kâğıt-kalem testiyle benzer doğrulukta yetenek kestirimi yapılabileceği belirtilse de (Urry, 1977) MPS'nin BOBUT ortamında çalışıldığı araştırmalarda da genellikle minimum 200 veya 300 maddelik madde bankası ile çalışıldığı duruma rastlanmıştır (Risk, 2015; Lunz, Stahl ve Bergstrom, 1993). Çünkü minimum 200 veya 300 maddelik banka ile çalışıldığı durumda maddelerin görülme sıklığının düşük düzeyde olduğu belirtilmektedir (Lunz vd., 1993). Bu durum dikkate alınarak minimum madde bankası büyüklüğü 200 olarak ele alınmıştır.

Madde bankasındaki madde sayısı 500: İlgili alanyazındaki çalışmalar incelendiğinde, 500 maddeden oluşan madde bankası büyüklüğü, BOBUT uygulamaları için genelde orta büyüklükte olarak ele alınmış ve birçok sertifika programı için ideal

madde bankası büyüklüğü olarak belirtilmiştir (Risk, 2015). Bu doğrultuda madde bankası büyüklüğünün 500 olduğu durum, orta büyüklükte madde bankası büyüklüğü olarak belirlenmiştir.

Madde bankasındaki madde sayısı 1000: Son yıllarda gerçekleştirilen BOBUT uygulamalarında kullanılan madde bankası büyüklüğü 1000 madde ve üzeridir. Madde bankasının büyük olması, maddelerin sürekli olarak bireyler tarafından görülmesinin ve madde aşinalığının önüne geçebilir (Wise ve Kingsbury, 2000). Madde bankası niteliği ve test uzunluğu arttıkça, test güvenilirliği, ölçme ve sınıflandırma hassasiyeti artış göstermektedir (Jodoin, 2003). Bu nedenle madde bankası büyüklüğünün 1000 olduğu durum, büyük madde bankası olarak belirlenmiştir.

200, 500 ve 1000 maddelik madde bankalarının MPS miktarı, MPS yüzdesi ve üç zaman noktasına göre düzenlenmiş hallerine ilişkin $-b$ parametre değerlerine Tablo 2, 3 ve 4'te yer verilmiştir.

Tablo 2

Madde Bankasının $k=200$ Olduğu MPS'li Veri Setlerinde Madde Bankalarının Ortalama Madde Güçlük Değerleri

MPS Büüklüğü	MPS Yüzdesi	1.zaman noktası	2.zaman noktası	3.zaman noktası	Ortalama -b değerleri
0.00 logit	%0	-0.03	-0.03	-0.03	-0.03
	%0	-0.03	-0.03	-0.03	-0.03
0.50 logit	%5	-0.05	-0.08	-0.11	-0.08
	%10	-0.08	-0.13	-0.18	-0.13
	%20	-0.13	-0.23	-0.33	-0.23
0.75 logit	%0	-0.03	-0.03	-0.03	-0.03
	%5	-0.06	-0.10	-0.14	-0.10
	%10	-0.10	-0.18	-0.26	-0.18
	%20	-0.18	-0.33	-0.48	-0.33
1.00 logit	%0	-0.03	-0.03	-0.03	-0.03
	%5	-0.08	-0.13	-0.18	-0.13
	%10	-0.13	-0.23	-0.33	-0.23
	%20	-0.23	-0.43	-0.63	-0.43

Tablo 2 incelendiğinde 200 maddelik madde bankasının temel düzey veri setine ilişkin madde güçlük ortalaması -0.03'tür. Oluşturulan MPS'li veri setlerinde, üç zaman noktası için madde güçlük parametrelerinin ortalaması ise -0.08 ile -0.43 arasındadır. MPS'li madde yüzdesi ve MPS büyüklüğü arttıkça, madde bankasına ilişkin madde güçlük değerlerinin ortalamasının kolay yönde değiştiği görülmektedir.

Tablo 3

Madde Bankasının k= 500 Olduğu MPS'li Veri Setlerinde Madde Bankalarının Ortalama Madde Güçlük Değerleri

MPS Büyükliği	MPS Yüzdesi	1.zaman noktası	2.zaman noktası	3.zaman noktası	Ortalama -b değerleri
0.00 logit	%0	0.02	0.02	0.02	0.02
	%0	0.02	0.02	0.02	0.02
0.50 logit	%5	0.00	-0.02	-0.05	-0.02
	%10	-0.02	-0.07	-0.12	-0.07
	%20	-0.07	-0.17	-0.27	-0.17
	%0	0.02	0.02	0.02	0.02
0.75 logit	%5	-0.01	-0.05	-0.08	-0.05
	%10	-0.05	-0.12	-0.20	-0.12
	%20	-0.12	-0.27	-0.42	-0.27
	%0	0.02	0.02	0.02	0.02
1.00 logit	%5	-0.02	-0.07	-0.12	-0.07
	%10	-0.07	-0.17	-0.27	-0.17
	%20	-0.17	-0.37	-0.57	-0.37

Tablo 3 incelendiğinde 500 maddelik madde bankasının temel düzey veri setine ilişkin madde güçlük ortalaması 0.02'dir. Oluşturulan MPS'li veri setlerinde, üç zaman noktası için madde güçlük parametrelerinin ortalaması ise -0.02 ile -0.37 arasındadır. MPS'li madde yüzdesi ve MPS büyüklüğü arttıkça, madde bankasına ilişkin madde güçlük değerlerinin ortalamasının kolay yönde değiştiği görülmektedir.

Tablo 4

Madde Bankasının k= 1000 Olduğu MPS'li Veri Setlerinde Madde Bankalarının Ortalama Madde Güçlük Değerleri

MPS Büyükliği	MPS Yüzdesi	1.zaman noktası	2.zaman noktası	3.zaman noktası	Ortalama -b değerleri
0.00 logit	%0	0.05	0.05	0.05	0.05
	%0	0.05	0.05	0.05	0.05
0.50 logit	%5	0.03	0.00	-0.02	0.00
	%10	0.00	-0.04	-0.09	-0.04
	%20	-0.04	-0.14	-0.24	-0.14
	%0	0.05	0.05	0.05	0.05
0.75 logit	%5	0.02	-0.02	-0.05	-0.02
	%10	-0.02	-0.09	-0.16	-0.09
	%20	-0.09	-0.24	-0.39	-0.24
	%0	0.05	0.05	0.05	0.05
1.00 logit	%5	0.00	-0.04	-0.09	-0.04
	%10	-0.04	-0.14	-0.24	-0.14
	%20	-0.14	-0.34	-0.54	-0.34

Tablo 4 incelendiğinde 1000 maddelik madde bankasının temel düzey veri setine ilişkin madde güçlük ortalaması 0.05'tir. Oluşturulan MPS'li veri setlerinde, üç zaman noktası için madde güçlük parametrelerinin ortalaması ise 0.00 ile -0.34 arasındadır. MPS'li madde yüzdesi ve MPS büyüklüğü arttıkça, madde bankasına ilişkin madde güçlük değerlerinin ortalamasının kolay yönde değiştiği görülmektedir. Simülatif veri üretiminde toplamda 2 (örneklem büyüklüğü) $\times$ 3 (madde bankası büyüklüğü) $\times$ 4 (MPS büyüklüğü) $\times$ 4 (MPS içeren madde yüzdesi) $\times$ 3 (zaman noktası) olmak üzere 288 durum ele alınmıştır. Her bir koşul için toplam 100 replikasyon yapılmış ve 28800 analiz gerçekleştirilmiştir.

Belirtilen koşulların yetenek kestirimleri üzerindeki etkisini incelemek için ölçme kesinliği, test bilgi fonksiyonu (TBF) ve test etkililiği değerleri hesaplanmıştır. Hesaplamalar için gerekli formüllere Tablo 5'te yer verilmiştir.

Tablo 5

Madde Parametre Sapmasını Değerlendirme Kriterleri

Ölçüt	Açıklama	Formül
Ölçme kesinliği	Yanlılık	Gerçek yeteneğin, kestirilen yetenekten sistematik olarak sapmasıdır.
	OHKK	Ortalama hataların kareköküdür.
	MOF	Maksimum ortalama farktır. Gerçek ve kestirilen yetenek arasındaki ortalama yanlılığı verir.
Test Bilgi Fonksiyonu	TBF	Test bilgi fonksiyonudur. Bireylerin ilgili testten aldıkları maddelerin bilgi fonksiyonu toplamına eşittir. Bu değer yetenek kestirimlerine ilişkin standart hata değerlerinden yola çıkarak hesaplanmaktadır.
	Test uzunluğu	Her bireye uygulanan ortalama madde sayısı
Test etkililiği	Maddelerin ortalama uygulanma oranı	Bir maddenin bireylere toplam uygulanma oranının, birey sayısına bölünmesi ile elde edilir.
	X^2	Ki-kare. Gözlenen ve beklenen madde kullanım sıklığı arasındaki farktır.

θ_i : Bireylerin yetenek düzeyi, $\hat{\theta}_i$: Bireylerin kestirilen yetenek düzeyi, n : Toplam birey sayısı, $I(\theta)$: Madde bilgi fonksiyonu, K_i : i. bireye uygulanan toplam madde sayısı, m_k : k maddesinin tüm bireylere uygulanma sayısı, K_{oj} : j maddesinin uygulanma sayısı/Kişi sayısı, $\bar{K}_{oj}$: Test uzunluğu/Madde bankası büyüklüğü

Tablo 5'te ölçme kesinliği, TBF ve test etkililiğini değerlendirmede kullanılan formüller yer almaktadır. Ölçme kesinliği için yanlılık, ortalama hataların karekökü (OHKK, Root Mean Square Error) ve mutlak ortalama fark (MOF, Absolute Average Difference) değerleri hesaplanmıştır. Çünkü bu istatistikler, ölçme kesinliğine ilişkin doğru bilgiler vermesinden kaynaklı olarak simülasyon çalışmalarında sıklıkla kullanılan istatistiklerdir (Guyer ve Thompson, 2011; Lee ve Dodd, 2012; Risk, 2015; Wang ve Chen, 2005). Ölçme kesinliğine ilişkin bu faktörler bireylerin gerçek yetenekleri, kestirilen yetenekleri ve örneklem büyüklüğü dikkate alınarak hesaplanmaktadır. Ölçme kesinliğine ilişkin değerler 100 replikasyon için hesaplandıktan sonra, elde edilen değerler arasında istatistiksel olarak manidar fark olup olmadığını incelemek amacıyla bağımsız örneklemeler için üç faktörlü ANOVA analizi yapılmıştır. Analizde MPS büyüklüğü (0.00 logit, 0.50 logit, 0.75 logit, 1.00 logit), MPS yüzdesi (%0, %5, %10, %20) ve MPS'li madde bankası ile yapılan ölçümler (3 ölçüm) bağımsız değişkenler; yanlılık, OHKK ve MOF değerleri de ayrı ayrı bağımlı değişkenler olarak ele alınmıştır. ANOVA analizi ile birlikte eta kare (η^2) etki değeri de raporlaştırılmıştır. Etki değerini yorumlamada .01 küçük, .09 orta, .25 geniş etki kabul edilmiştir (Cohen, 1988, s.413).

Test bilgi fonksiyonu testin verdiği bilgi miktarını göstermektedir. Bireylerin BOBUT uygulaması sonunda elde ettikleri yetenek kestirimlerinin güvenilirliği ile bu kestirime ilişkin ölçmenin standart hatası arasında ters bir ilişki vardır. Test bilgi fonksiyonu da bu ilişkiden yola çıkarak hesaplanmaktadır (van der Linden, 1998). Test bilgi fonksiyonu, her bireyin yetenek kestirimlerine ilişkin ölçmenin standart hatası değeri elde edilerek hesaplanmıştır. Bu değer her bir koşula ilişkin 100 replikasyon için hesaplanmış ve elde edilen değerler arasında istatistiksel olarak manidar fark olup olmadığını incelemek amacıyla bağımsız örneklemeler için faktöriyel ANOVA analizi yapılmıştır. Analizde MPS büyüklüğü (0.00 logit, 0.50 logit, 0.75 logit, 1.00 logit), MPS yüzdesi (%0, %5, %10, %20) ve MPS'li madde bankası ile yapılan ölçümler (3 ölçüm) bağımsız değişkenler; TBF de bağımlı değişken olarak ele alınmıştır. ANOVA analizi ile birlikte eta kare (η^2) etki değeri de raporlaştırılmıştır. Etki değerini yorumlamada .01 küçük, .09 orta, .25 geniş etki kabul edilmiştir (Cohen, 1988, s.413).

Test etkililiği için de test uzunluğu, maddelerin ortalama uygulanma oranı (Average item exposure rate) ve X^2 değerleri hesaplanmıştır. Bu değerler de BOBUT uygulamalarında, test etkililiğini değerlendirmede doğru bilgiler vermesinden kaynaklı sıklıkla kullanılan istatistiklerdir (Lee ve Dodd, 2012; Moyer, Galindo ve Dodd, 2012). Test uzunluğunu hesaplamada her bireyin aldığı ortalama madde sayısı dikkate alınmıştır.

Maddelerin ortalama uygulanma oranı ise bir maddenin bireylere toplam uygulanma sayısının, toplam birey sayısına oranı ile hesaplanmıştır. X^2 istatistiği de gözlenen ve beklenen madde kullanım sıklığını göstermektedir. Hesaplanmasında bir maddenin uygulanma sayısı, örneklem büyüklüğü, test uzunluğu ve madde bankası büyüklüğü faktörleri etkilidir ve madde bankasının kullanımının etkililiğine ilişkin bilgi verir. Oluşturulan MPS koşulları içerisinde düşük X^2 değerine sahip olan koşul, madde bankasındaki maddelerin birçoğunu eşit düzeyde kullanan koşul anlamına gelmektedir (Chang ve Ying, 1999). Belirtilen test etkililiği değerleri, her bir koşulda 100 replikasyon için hesaplanarak ortalaması alınmıştır.

Oluşturulan tüm koşullar içerisinde MPS'nin etkisini hesaplamada, ilk oluşturulan veri seti temel düzey veri seti olarak ele alınmıştır. Bu veri seti üzerinden MPS koşulları oluşturulduktan sonra MPS'li veri setleri ile temel düzey veri seti karşılaştırılarak yorumlar yapılmıştır.

BÖLÜM 3

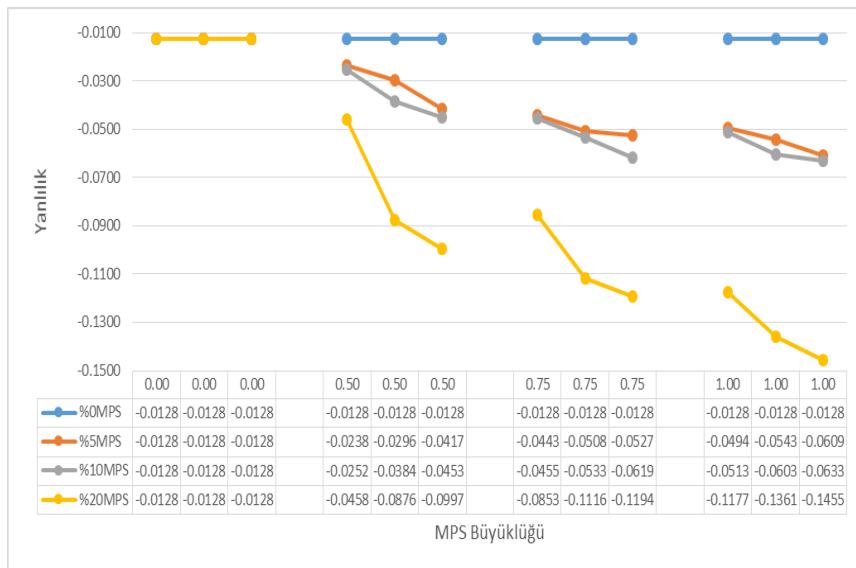
BULGULAR VE YORUMLAR

Bu bölümde, araştırma kapsamında yapılan analizler sonucunda elde edilen bulgulara ve yorumlara yer verilmiştir. Bu bölümde öncelikle 1000 kişilik veriden elde edilen bulgu ve yorumlar, ardından 5000 kişilik veriden elde edilen bulgu ve yorumlara yer verilmiştir.

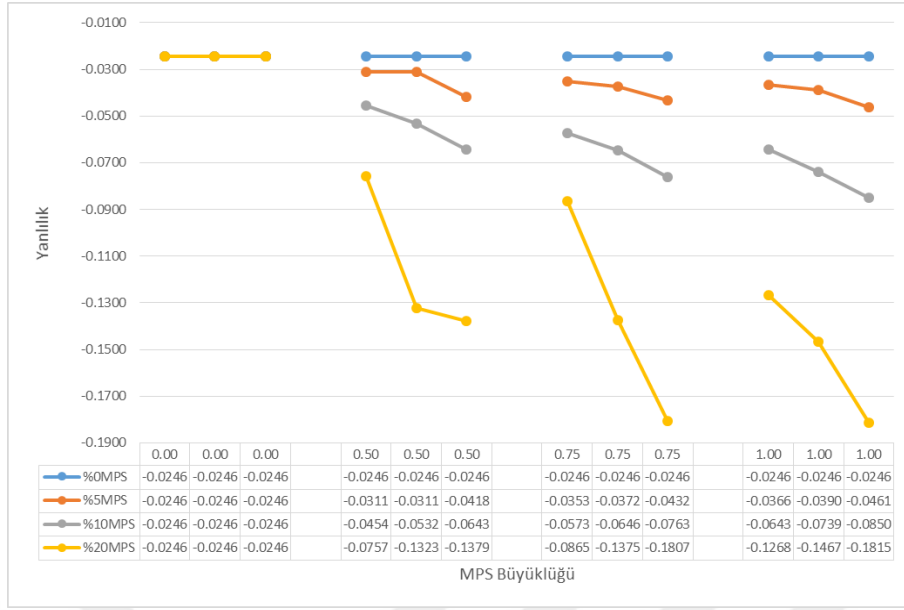
Örneklem Büyüklüğünün 1000 Olduğu Durumda Koşulların Karşılaştırılmasına İlişkin Bulgular

1000 kişilik örneklem büyüklüğü ve 200, 500 ve 1000 maddelik madde bankaları için bulgulara aşağıda yer verilmiştir.

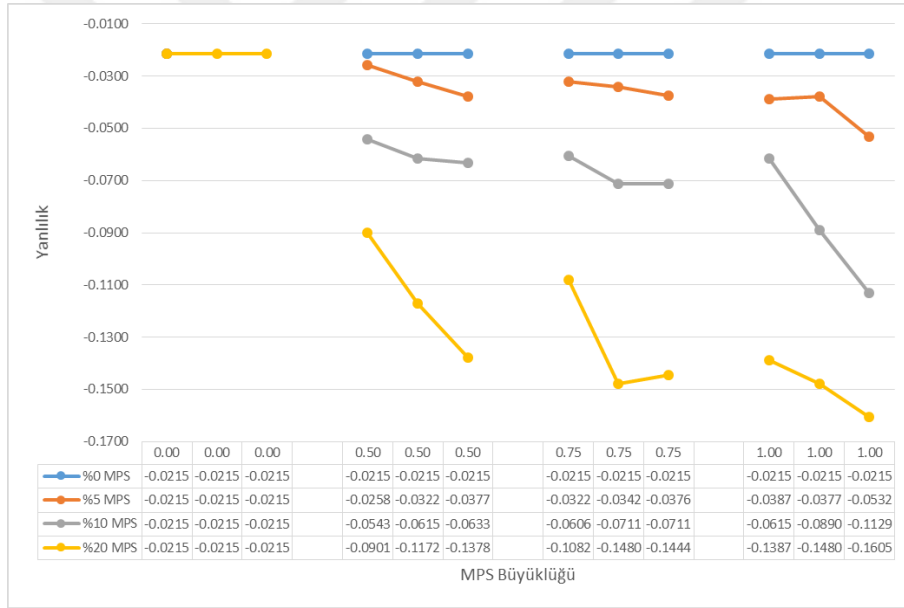
Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı ölçme kesinliğine ilişkin birinci ölçüt (bağımlı değişken), yeterlik kestirimlerine ait yanlılık değerleridir. Yanlılık değerlerine ilişkin bulgular şekil 3.a, 3.b ve 3.c’de yer almaktadır. Bulguların üç madde bankası için de paralellik göstermesinden dolayı, üç şekile ilişkin yorumlar şekillerin ardından toplu olarak sunulmuştur.



Şekil 3.a. Örneklem n=1000 olduğunda 200 Maddelik Banka için Yanlılık Değerleri



Şekil 3.b. Örneklem n=1000 olduğunda 500 Maddelik Banka için Yanlılık Değerleri



Şekil 3.c. Örneklem n=1000 olduğunda 1000 Maddelik Banka için Yanlılık Değerleri

Şekil 3.a, 3.b ve 3.c. Örneklem n=1000 olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre Yanlılık Değerlerinin Karşılaştırılmasına İlişkin Grafikler

Yanlılığa ilişkin değerler incelendiğinde; 1000 kişilik örneklemde 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç farklı zaman noktasında elde edilen yetenek kestirimlerine ait *yanlılık* değerlerinin negatif yönde artma eğiliminde olduğu

görülmektedir. Yanlılık değerlerine ilişkin en düşük değer, MPS olmadığı için temel düzey veri setinde elde edilmiştir. Bir diğer ifadeyle MPS miktarı, MPS yüzdesi ve MPS'li madde bankası ile yapılan ölçüm sayısı arttıkça daha yanlı yetenek kestirimleri elde edilmekte, dolayısıyla ölçme kesinliği de düşmektedir. Bu eğilim madde bankası büyüklüğü artsa da değişkenlik göstermemiştir. Bir diğer ifadeyle, BOBUT uygulamalı testlerde MPS'nin miktarı ve bankadaki MPS'li madde yüzdesi arttığı sürece, madde bankasının büyümesi, yeterli kestirimlerinin yanlı olmasını engellememektedir.

Elde edilen negatif yanlılık değerlerinin anlamı; bireylerin kestirilen yetenek değerlerinin, gerçek yetenek değerlerinden düşük olmasıdır. Madde bankasındaki bir kısım maddenin kolay yönde MPS göstermesinden dolayı, bireylerin kestirilen yetenek değerlerinin, gerçek yeteneklerinden yüksek olması bir başka deyişle yanlılık değerlerinin pozitif yönde artmış olması beklenirdi. Beklenenin tersi bir sonuç elde edilmesinin bir nedeni, sadece tek yönde gerçekleşen (kolay yönde) MPS durumunun ele alınmış olmasından kaynaklı olabilir. Çünkü MPS her iki yönde de gerçekleşiyor olsaydı yanlılığa ilişkin kestirimler sıfıra yakın çıkardı (Aksu Dünya, 2017; Wei, 2013). Bu durumun diğer bir nedeni de bireylerin yetenek düzeyindeki maddelerle karşılaşmalarına rağmen bunları yanlış yanıtlamaları olabilir. Benzer bulguya ilgili alanyazında yapılan bazı çalışmalarda da rastlanmaktadır (Chen, 2013; Risk, 2015; Rupp ve Zumbo, 2003b). Chen (2013), MPS'nin yetenek kestirimlerine etkisini incelediği çalışmasında madde güçlük parametresinin 0.40 logit büyüklüğünde kolay yönde değiştiği durumda, yetenek kestirimlerinin, dolayısıyla da yetenek kestirimlerine ilişkin yanlılığın düşük düzeyde de olsa negatif yönde değiştiğini belirtmiştir. Bu durum; aynı test formunun birden çok kez kullanılması sonucu öğrenciler bazı maddelere aşina olmakta ve bu aşinalık, maddeleri kolaylaştırmakta ve MPS riskini arttırmaktadır şeklinde yorumlanabilir. Bu durum yetenek kestirimlerini de olumsuz etkilemekte ve daha yanlı sonuçlar elde edilmesine neden olmaktadır (Jiang, Drasgow ve Tay, 2009). Ancak Guo ve Wang (2003) tarafından BOBUT kapsamında yapılan ölçek kaymasının test puanlarına etkisinin incelendiği çalışmada, MPS'li madde bankası olduğu durumda yetenek kestirimlerine ilişkin yanlılık değerlerinin çok fazla etkilenmediği belirtilmiştir. Bu durumun nedeni, belirtilen çalışmada iki zaman noktasında ölçüm alınmıştır. Ancak, Babcock ve Albano (2012) yeterliklere ilişkin iki ölçüm almanın MPS hakkında net çıkarımlar yapabilmek için yeterli olmadığını belirtmiştir. Bir başka ifade ile MPS'nin etkilerinin ortaya çıkarılabilmesi için en az üç noktada ölçüm almak gerekmektedir.

Yukarıda verilen yanlılık değerleri açısından elde edilen farkların istatistiksel olarak manidar olup olmadığını incelemek için bağımsız örneklemeler için uygulanan üç faktörlü ANOVA sonuçları Tablo 6’da verilmiştir. Tabloda yer alan MPS büyüklüğü, madde bankasında MPS’ye sahip maddelerin sapma miktarını (0.00 logit, 0.50 logit, 0.75 logit, 1.00 logit), MPS yüzdesi madde bankasındaki MPS’li madde yüzdesini (%0, %5, %10, %20) ve ölçüm faktörü de MPS’li maddelere sahip madde bankası ile yapılan ölçüm sayısını temsil etmektedir (3 ölçüm).

Tablo 6

Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Yanlılık Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
200	MPS Büyüklüğü	0.515	3	0.172	5366.84*	0.15
	MPS Yüzdesi	2.031	3	0.677	21165.16*	0.58
	Ölçüm	0.235	2	0.118	2448.95*	0.06
	MPS Büyüklüğü*MPS Yüzdesi	0.113	9	0.013	1177.58*	0.03
	MPS Büyüklüğü*Ölçüm	0.016	6	0.003	166.74*	0.01
	MPS Yüzdesi*Ölçüm	0.071	6	0.012	739.89*	0.02
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.011	18	0.001	114.63*	0.01
	Hata	0.456	4752	0.000		
	Toplam	3.448	4799			
500	MPS Büyüklüğü	0.196	3	0.065	2352.00*	0.03
	MPS Yüzdesi	4.414	3	1.471	52968.00*	0.73
	Ölçüm	0.493	2	0.247	5916.00*	0.08
	MPS Büyüklüğü*MPS Yüzdesi	0.070	9	0.008	840.00*	0.01
	MPS Büyüklüğü*Ölçüm	0.022	6	0.004	264.00*	0.01
	MPS Yüzdesi*Ölçüm	0.333	6	0.056	3996.00*	0.05
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.060	18	0.003	720.00*	0.01
	Hata	0.396	4752	0.000		
	Toplam	5.984	4799			
1000	MPS Büyüklüğü	0.271	3	0.090	2893.91*	0.05
	MPS Yüzdesi	4.232	3	1.411	45192.05*	0.78
	Ölçüm	0.248	2	0.124	2648.31*	0.04

(Devam ediyor)

Tablo 6 (Devam)

Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Yanlılık Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
1000	MPS Büyüklüğü*MPS Yüzdesi	0.047	9	0.005	501.90*	0.01
	MPS Büyüklüğü*Ölçüm	0.024	6	0.004	256.29*	0.01
	MPS Yüzdesi*Ölçüm	0.058	6	0.010	619.36*	0.01
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.075	18	0.004	800.90*	0.01
	Hata	0.445	4752	0.000		
	Toplam	5.400	4799			

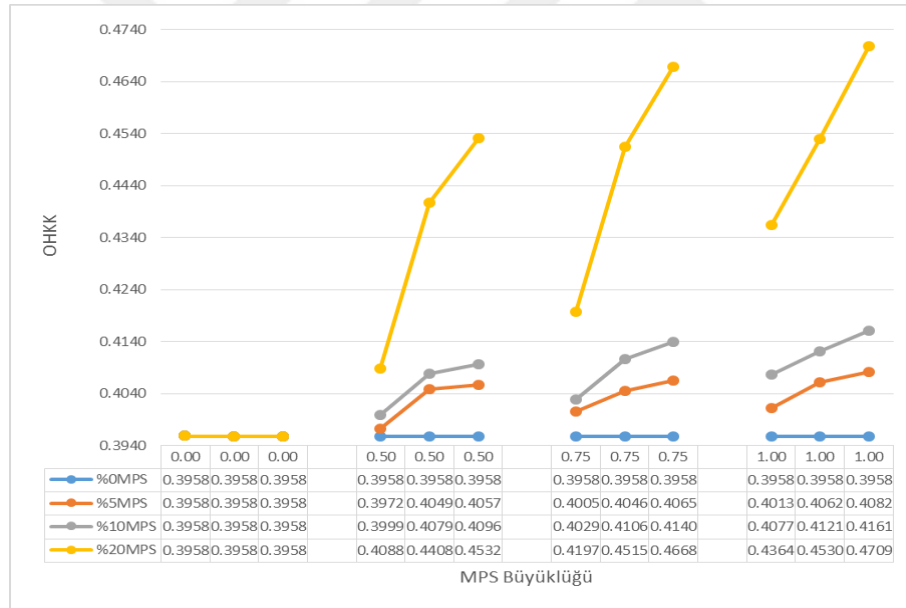
*p<.05

Yanlılığa ilişkin olarak bağımsız örneklemeler için üç faktörlü ANOVA sonuçları genel olarak incelendiğinde 1000 kişilik örneklemde 200, 500 ve 1000 maddelik madde bankaları için MPS büyüklüğü, MPS yüzdesi, MPS’li madde bankası ile yapılan ölçüm sayısının hem temel etki, hem ikili etkileşim hem de üçlü etkileşim etkilerinin yanlılık üzerinde istatistiksel olarak manidar etkisi olduğu görülmüştür. Bunlar genel olarak düşük etki büyüklüklerine sahiptir (Cohen, 1988, s. 413). Yapılan Post-Hoc test sonuçları da tüm faktörlerin tüm düzeyleri açısından fark olduğunu ortaya koymuştur. Yeterlik kestirimlerinin yanlılığı üzerinde en fazla etkisi olan faktör; MPS yüzdesidir. Bir başka ifade ile MPS’ye ilişkin belirtilen faktörler yanlılığı ve dolayısıyla ölçme kesinliğini olumsuz yönde etkilemektedir. Bu doğrultuda BOBUT uygulamalarında MPS’li madde bankasının uzun süre kullanımından kaynaklı, bazı maddelerin kolay yönde MPS gösterdiği durumda yetenek kestirimleri önemli düzeyde etkilenmekte ve daha yanlı sonuçlar elde edilmektedir. Benzer bulguya Rasch modeli kullanmış olan Aksu Dünya (2017), Babcock ve Albano (2012) ve 3PLM’yi kullanmış olan Abad ve diğerlerinin (2010) çalışmalarında da rastlanmaktadır.

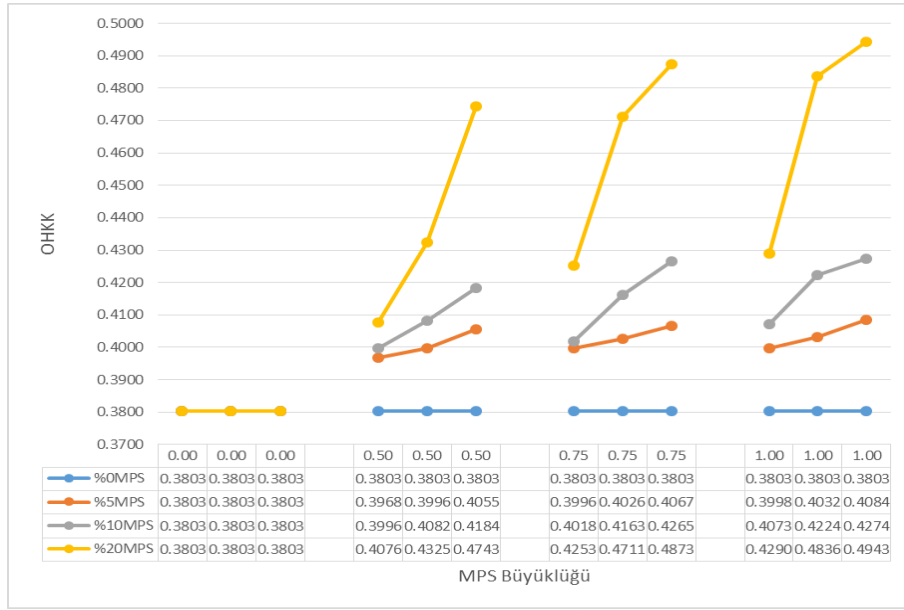
Babcock ve Albano (2012) yaptıkları çalışmada istatistiksel olarak manidar farkın yanında, yanlılık üzerinde en fazla etkisi olan faktörün MPS’nin yönü (madde güçlüğüünün artması daha etkin) ve MPS yüzdesi* MPS yönünün ortak etkisi olduğunu belirtmişlerdir. Bu çalışma kapsamında MPS’nin sadece kolay yönde değiştiği durum ele

alındığı için, MPS'nin yönü faktörünün etkisi incelenmemiş ancak en etkin faktörün MPS yüzdesi olduğu bulgusuna ulaşılmıştır. Chen (2013) ise MPS gösteren madde sayısının yanlılık üzerindeki etkisinin çok düşük düzeyde olduğunu belirtmiştir. Guo ve Wang (2010) ise BOBUT ortamında yanlılık değerlerini karşılaştırdıkları çalışmalarında yanlılık açısından anlamlı fark elde edilmediğini belirtmişlerdir. Bu durumun nedeni olarak da iki zaman noktası gibi az sayıda ölçüm alınmasını belirtmişlerdir. Çünkü MPS'nin etkisinin net bir şekilde görülebilmesi için ikiden fazla zaman noktasına ihtiyaç vardır (Chan vd., 1999; Deng ve Melican, 2009).

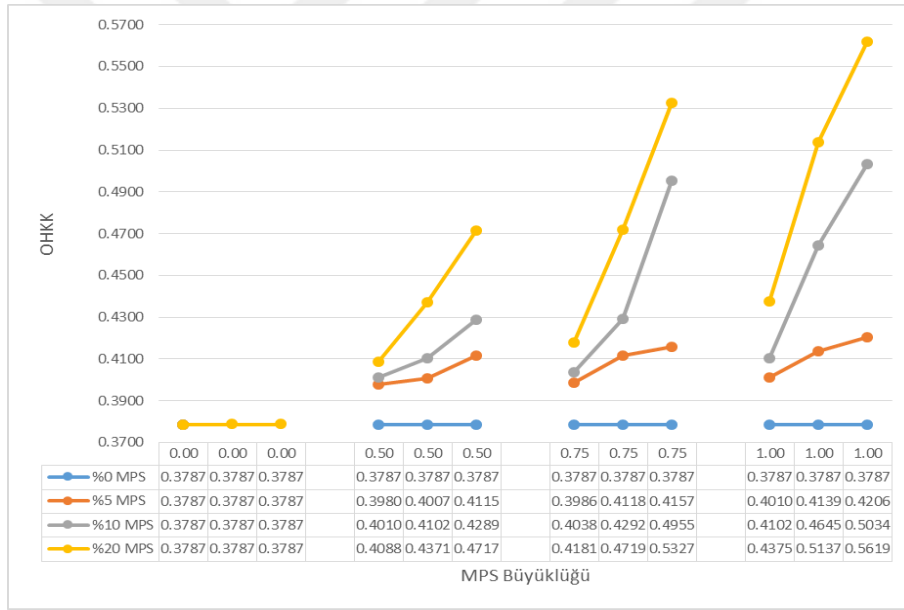
Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı ölçme kesinliğine ilişkin ikinci ölçüt (bağımlı değişken), yeterlik kestirimlerine ait OHKK değerleridir. OHKK değerlerine ilişkin bulgular şekil 4.a, 4.b ve 4.c'de yer almaktadır. Bulguların üç madde bankası için de paralellik göstermesinden dolayı, üç şekile ilişkin yorumlar şekillerin ardından toplu olarak sunulmuştur.



Şekil 4.a. Örneklem n=1000 olduğunda 200 Maddelik Banka için OHKK Değerleri



Şekil 4.b. Örneklem n=1000 olduğunda 500 Maddelik Banka için OHKK Değerleri



Şekil 4.c. Örneklem n=1000 olduğunda 1000 Maddelik Banka için OHKK Değerleri

Şekil 4.a, 4.b ve 4.c. Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre OHKK Değerlerinin Karşılaştırılmasına İlişkin Grafikler

Ortalama Hataların Karekökü'ne ilişkin değerler incelendiğinde; 1000 kişilik yanıtlayıcı örnekleminde 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç farklı zaman noktasında elde edilen yetenek kestirimlerine ait OHKK değerlerinin artma

eğiliminde olduğu görülmektedir. Buna göre, MPS miktarı, MPS yüzdesi ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça daha hatalı yetenek kestirimleri elde edilmekte, dolayısıyla ölçme kesinliği düşmektedir. OHKK açısından en yüksek değerler MPS büyüklüğünün 1.00, MPS yüzdesinin %20 olduğu ve üçüncü ölçümün alındığı noktada elde edilmiştir. OHKK değerleri, gerçek ve kestirilen yetenek değerleri arasındaki farkı (hatayı) göstermektedir. OHKK değeri sıfıra ne kadar yakınsa, bu durum BOBUT uygulaması sonunda elde edilen yeterlik kestirimi, gerçek yeterlik değerine o kadar yakın kestirim yapmış anlamına gelmektedir. Bu eğilim madde bankası büyüklüğü artsa da değişkenlik göstermemektedir. Benzer bulguya alanyazında yapılan bazı çalışmalarda da rastlanmaktadır (Aksu Dünya, 2017; Babcock ve Albano, 2012; Chen, 2013; Risk, 2015; Wells vd., 2002). Wells ve diğerleri (2012), yaptıkları çalışmada örneklem büyüklüğü arttıkça OHKK değerlerinin azaldığı ve dolayısıyla daha az hatalı kestirimler elde edildiğini ancak aynı örneklem büyüklüğü içerisinde, MPS miktarı arttıkça daha yüksek OHKK değerleri elde edildiğini ve ölçme kesinliğinin azaldığını belirtmişlerdir. Aksu Dünya (2017) da en düşük OHKK değerinin temel düzey veri seti için elde edildiğini belirtirken, özellikle MPS’li madde yüzdesi arttıkça yeteneklere ilişkin daha hatalı kestirimler yapıldığını belirtmişlerdir. OHKK değerlerine ilişkin en düşük değerler ise MPS olmadığı için temel düzey veri setinde elde edilmiştir. Ancak, temel düzey veri setleri için madde bankası büyüklüğü arttıkça OHKK değerleri azalmıştır. Bir başka deyişle madde bankası büyüdükçe temel düzey veri setinde, yetenek kestirimlerine ilişkin daha az hatalı sonuçlar elde edilmiştir.

Ortalama Hataların Karekökü değerleri düşük düzeyde artış göstermiş olsa da, zamanla bu hataların birikmesi sonucu ölçek kayması gözlenebilir. Bu durum da puanların yorumlanmasına ilişkin geçerliği olumsuz yönde etkileyen bir faktördür (Guo ve Wang, 2003). Özellikle, test maddelerinin görece kolay olduğu test durumlarında, zor yönde MPS gözlemlenmesi, yetenek kestirimlerine ilişkin OHKK değerlerini daha olumsuz etkilemektedir (Babcock ve Albano, 2012). Bu çalışma kapsamında, görece düşük miktarlarda OHKK değişiminin gözlenmesinin nedenlerinden biri yalnızca kolay yönde değişen MPS koşullarının ele alınmış olmasından kaynaklı olabilir.

Yukarıda verilen OHKK değerleri açısından elde edilen farkların istatistiksel olarak manidar olup olmadığını incelemek için bağımsız örneklem için yapılan üç faktörlü ANOVA sonuçları Tablo 7’de verilmiştir.

Tablo 7.

Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre OHKK Değerlerinin Karşılaştırılması

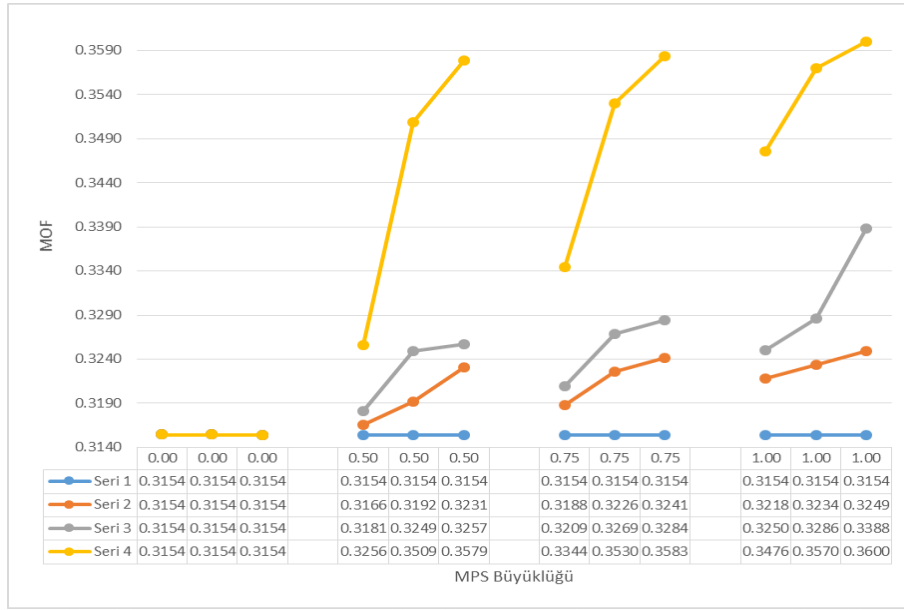
Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
200	MPS Büyüklüğü	0.039	3	0.013	652.56*	0.02
	MPS Yüzdesi	0.884	3	0.295	14791.44*	0.57
	Ölçüm	0.180	2	0.090	3011.83*	0.11
	MPS Büyüklüğü*MPS Yüzdesi	0.023	9	0.003	384.85*	0.01
	MPS Büyüklüğü*Ölçüm	0.005	6	0.001	83.66*	0.01
	MPS Yüzdesi*Ölçüm	0.115	6	0.019	1924.23*	0.07
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.005	18	0.000	83.66*	0.00
	Hata	0.284	4752	0.000		
	Toplam	1.535	4799			
500	MPS Büyüklüğü	0.104	3	0.035	1752.51*	0.03
	MPS Yüzdesi	1.431	3	0.477	24113.87*	0.54
	Ölçüm	0.449	2	0.225	7566.13*	0.16
	MPS Büyüklüğü*MPS Yüzdesi	0.068	9	0.008	1145.87*	0.02
	MPS Büyüklüğü*Ölçüm	0.016	6	0.003	269.62*	0.00
	MPS Yüzdesi*Ölçüm	0.274	6	0.046	4617.19*	0.10
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.020	18	0.001	337.02*	0.00
	Hata	0.282	4752	0.000		
	Toplam	2.644	4799			
1000	MPS Büyüklüğü	0.718	3	0.239	9425.24*	0.12
	MPS Yüzdesi	1.880	3	0.627	24678.90*	0.32
	Ölçüm	1.777	2	0.889	23326.81*	0.30
	MPS Büyüklüğü*MPS Yüzdesi	0.253	9	0.028	3321.15*	0.04
	MPS Büyüklüğü*Ölçüm	0.205	6	0.034	2691.05*	0.03
	MPS Yüzdesi*Ölçüm	0.552	6	0.092	7246.14*	0.09
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.089	18	0.005	1168.31*	0.01
	Hata	0.362	4752	0.000		
	Toplam	5.836	4799			

*p<.05

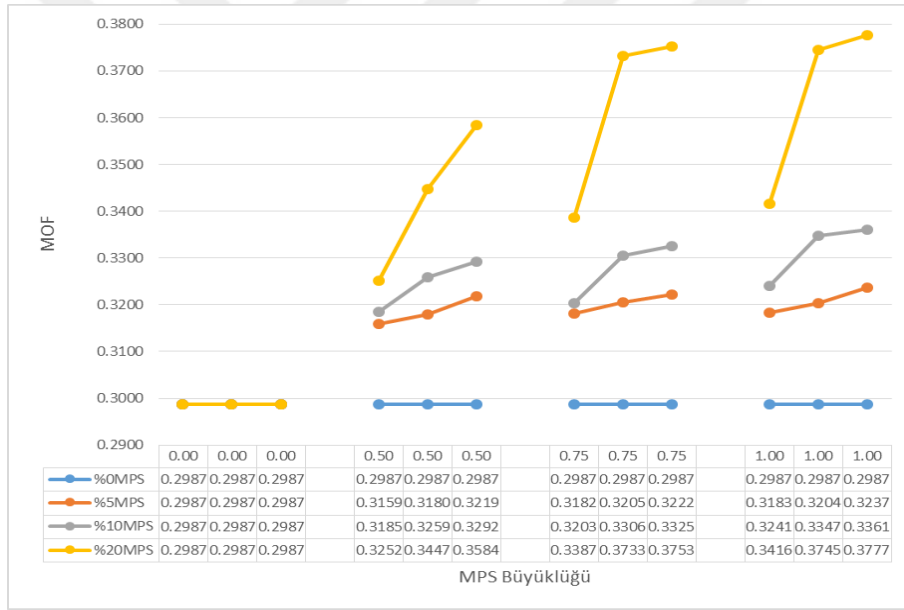
Ortalama Hataların Karekökü değerlerine ilişkin bağımsız örneklem için üç faktörlü ANOVA sonuçları genel olarak incelendiğinde 1000 kişilik yanıtlayıcı örnekleminde, 200, 500 ve 1000 maddelik madde bankaları için MPS büyüklüğü, MPS yüzdesi, MPS’li madde bankası ile yapılan ölçüm sayısının hem temel etki, hem ikili etkileşim hem de üçlü etkileşim etkilerinin OHKK üzerinde istatistiksel olarak manidar etkisi olduğu görülmüştür. Bu etkiler çoğunlukla düşük ve yüksek düzeydedir (Cohen, 1988, s.413). Yapılan Post-Hoc test sonuçları da tüm faktörlerin tüm düzeyleri açısından fark olduğunu ortaya koymuştur. Yeterlik kestirimlerine ilişkin OHKK üzerinde en fazla etkisi olan faktör; MPS yüzdesidir. Bir başka ifade ile MPS’ye ilişkin belirtilen faktörler OHKK’yı yani hata değerlerini ve dolayısıyla ölçme kesinliğini olumsuz yönde etkilemektedir. Bu doğrultu BOBUT uygulamalarında MPS’li madde bankasının uzun süre kullanımından kaynaklı, bazı maddelerin madde güçlük parametrelerinin kolay yönde MPS gösterdiği durumda yetenek kestirimleri önemli düzeyde etkilenmekte ve daha hatalı sonuçlar elde edilmektedir. Bu bulgu Risk’in (2015) çalışmasıyla da paralellik göstermektedir. Benzer bulguya Babcock ve Albano’nun (2012) çalışmasında da rastlanmış olmakla birlikte, bu çalışmada OHKK değerleri üzerinde en fazla etkisi olan faktörün MPS büyüklüğü olduğu belirtilmiştir.

Alanyazında MPS’nin OHKK üzerinde istatistiksel olarak manidar etkisinin olmadığını belirten çalışmalar da mevcuttur (Guo ve Wang, 2003; Wells vd., 2002). Örneğin, 2PLM’yi kullanan Wells ve diğerleri (2002) örneklem büyüklüğü ve MPS’li madde sayısının yetenek kestirimlerindeki hatayı daha çok etkilediğini, ancak bu etkinin istatistiksel olarak manidar olmadığını belirtmiştir. Ancak, belirtilen çalışmalarda simüle edilen örneklem büyüklüğü, madde bankası büyüklüğü ve MPS koşulları değişkenlik göstermektedir. Bu doğrultuda, BOBUT uygulamalarında madde bankasında MPS’li madde bulunması yetenek kestirimlerine ilişkin OHKK değerlerini olumsuz yönde etkilerken hangi faktörün bu değeri daha olumsuz etkilediği ve/veya etkilemediği, kullanılan MTK modeli, örneklem büyüklüğü, madde bankası büyüklüğü veya MPS koşuluna göre değişkenlik gösterebilir.

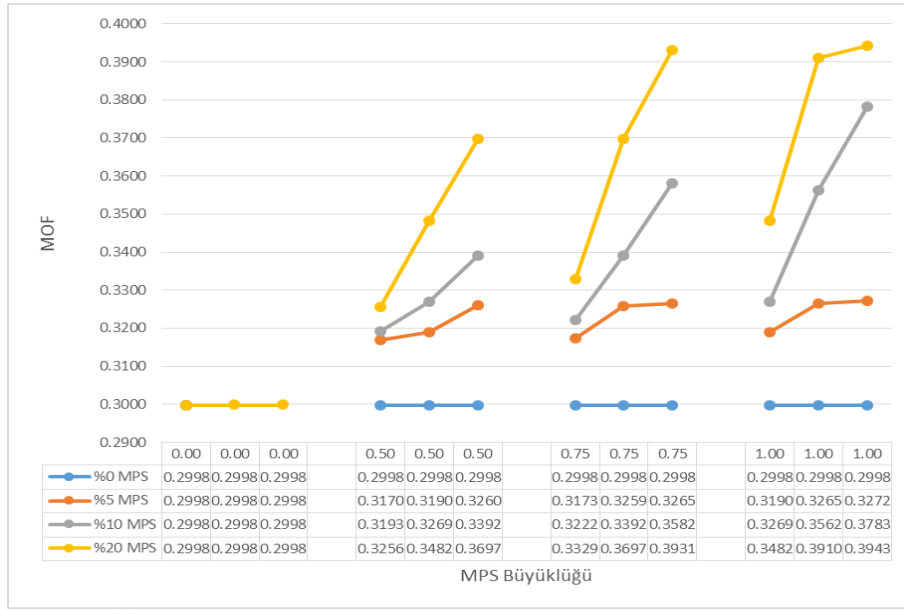
Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı ölçme kesinliğine ilişkin üçüncü ölçüt (bağımlı değişken), yeterlik kestirimlerine ait MOF değerleridir. MOF değerlerine ilişkin bulgular şekil 5.a, 5.b ve 5.c’de yer almaktadır. Bulguların üç madde bankası için de paralellik göstermesinden dolayı, üç şekile ilişkin yorumlar şekillerin ardından toplu olarak sunulmuştur.



Şekil 5.a. Örneklem n=1000 olduğunda 200 Maddelik Banka için MOF Değerleri



Şekil 5.b. Örneklem n=1000 olduğunda 500 Maddelik Banka için MOF Değerleri



Şekil 5.c. Örneklem n=1000 olduğunda 1000 Maddelik Banka için MOF Değerleri

Şekil 5.a, 5.b ve 5.c. Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre MOF Değerlerinin Karşılaştırılmasına İlişkin Grafikler

Maksimum Ortalama Fark değerine ilişkin grafikler incelendiğinde; 1000 kişilik yanıtlayıcı örnekleminde 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç farklı zaman noktasında elde edilen yetenek kestirimlerine ait *MOF* değerlerinin artma eğiliminde olduğu görülmektedir. *MOF* açısından en yüksek değerler MPS büyüklüğünün 1.00, MPS yüzdesinin %20 olduğu ve üçüncü ölçümün alındığı noktada elde edilmiştir. *MOF* değerleri, gerçek ve kestirilen yetenek değerleri arasındaki ortalama yanlılığı göstermektedir. Sıfıra yakın *MOF* değerleri, daha doğru yetenek kestirimi yapıldığı anlamına gelmektedir. Bu doğrultuda Şekil 5.a, 5.b ve 5.c'de görüldüğü üzere, MPS miktarı, MPS yüzdesi ve MPS'li madde bankası ile yapılan ölçüm sayısı arttıkça daha yanlı yetenek kestirimleri elde edilmekte ve dolayısıyla da ölçme kesinliği düşmektedir. Bu eğilim madde bankası büyüklüğü artsa da değişkenlik göstermemektedir. *MOF* değerlerindeki bu değişim sadece MPS'nin olmadığı temel düzey veri seti için ortaya çıkmamıştır.

Park, Li ve Xing (2016) MPS'nin etkisini gerçek ve simülasyon veri üzerinde inceledikleri araştırmalarında, madde güçlük ve ayırıcılık değerlerinde MPS olması durumunda madde parametrelerinin değişmesine bağlı olarak yetenek kestirimlerinin

olumsuz etkilendiğini, dolayısıyla MOF değerlerinin arttığı sonucuna ulaşmışlardır. Benzer bulguya Risk'in (2015) çalışmasında da rastlanmaktadır. Chen (2013), yaptığı çalışmada madde bankasında madde güçlük parametresi kolay yönde değişim gösteren madde sayısı arttıkça, ortalama yanlılığı ifade eden bu değer arttığını belirtmiştir. Wells ve diğerleri (2002) ise madde bankasındaki %20 oranındaki maddenin madde güçlük parametresindeki sapmanın, madde ayırıcılık parametresindeki sapmaya göre yetenek kestirimlerine ilişkin MOF değerlerini düşük düzeyde de olsa etkilediğini ancak bu etkinin önemli düzeyde olmadığını belirtmişlerdir. Ancak hem güçlük hem de ayırıcılık parametresinde aynı anda MPS görülmesi durumunda, MOF değerlerinin daha çok etkilendiği sonucuna ulaşılmıştır.

Yukarıda verilen MOF değerleri açısından elde edilen farkların istatistiksel olarak manidar olup olmadığını incelemek için bağımsız örneklemeler için yapılan üç faktörlü ANOVA sonuçları Tablo 8'de verilmiştir.

Tablo 8

Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre MOF Değerlerinin Karşılaştırılması

Madde Bankası	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
200	MPS Büyüklüğü	0.034	3	0.011	878.09*	0.04
	MPS Yüzdesi	0.303	3	0.101	7825.30*	0.43
	Ölçüm	0.071	2	0.036	1833.65*	0.10
	MPS Büyüklüğü*MPS Yüzdesi	0.019	9	0.002	490.70*	0.02
	MPS Büyüklüğü*Ölçüm	0.014	6	0.002	361.57*	0.02
	MPS Yüzdesi*Ölçüm	0.028	6	0.005	723.13*	0.04
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.040	18	0.002	1033.04*	0.05
	Hata	0.184	4752	0.000		
	Toplam	0.693	4799			
	MPS Büyüklüğü	0.054	3	0.018	20.76*	0.00
500	MPS Yüzdesi	0.670	3	0.223	257.59*	0.05
	Ölçüm	0.150	2	0.075	57.67*	0.01
	MPS Büyüklüğü*MPS Yüzdesi	0.041	9	0.005	15.76	-

(Devam ediyor)

Tablo 8 (Devam)

Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre MOF Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
500	MPS Büyüklüğü*Ölçüm	0.004	6	0.001	1.54	-
	MPS Yüzdesi*Ölçüm	0.090	6	0.015	34.60*	0.01
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.004	18	0.000	1.54	-
	Hata	12.360	4752	0.003		
	Toplam	13.373	4799			
1000	MPS Büyüklüğü	0.174	3	0.058	51.42*	0.01
	MPS Yüzdesi	0.758	3	0.253	224.01*	0.04
	Ölçüm	0.457	2	0.229	135.05*	0.02
	MPS Büyüklüğü*MPS Yüzdesi	0.061	9	0.007	18.03	-
	MPS Büyüklüğü*Ölçüm	0.023	6	0.004	6.80	-
	MPS Yüzdesi*Ölçüm	0.141	6	0.024	41.67*	0.01
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.026	18	0.001	7.68	-
	Hata	16.080	4752	0.003		
	Toplam	17.720	4799			

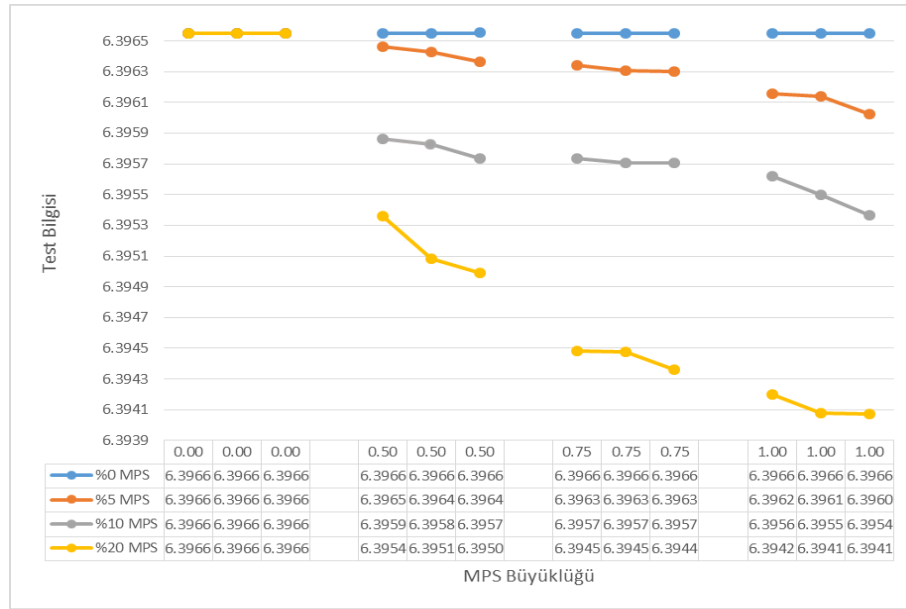
*p<.05

Maksimum Ortalama Fark değerlerine ilişkin bağımsız örneklem için üç faktörlü ANOVA sonuçları genel olarak incelendiğinde 1000 kişilik yanıtlayıcı örnekleme ve 200 maddelik madde bankası için MPS büyüklüğü, MPS yüzdesi, MPS’li madde bankası ile yapılan ölçüm sayısının hem temel etki, hem ikili etkileşim hem de üçlü etkileşim etkilerinin MOF üzerinde istatistiksel olarak manidar etkisi olduğu görülmüştür. Bunlar genel olarak düşük etki büyüklüklerine sahiptir (Cohen, 1988, s.413). 1000 kişilik yanıtlayıcı örnekleminde 500 ve 1000 maddelik madde bankaları için de faktörlerin temel etki düzeyinde ve MPS yüzdesi*Ölçüm şeklinde iki faktörün etkileşim etkisinin MOF üzerinde manidar etkisi olduğu görülse de genel olarak düşük etki büyüklüğüne sahiptirler (Cohen, 1988, s.413). Yapılan Post-Hoc test sonuçları da anlamlı fark çıkan tüm faktörlerin tüm düzeyleri açısından fark olduğunu ortaya koymuştur. Yeterlik kestirimlerine ilişkin MOF üzerinde en fazla etkisi olan faktör; MPS yüzdesidir. Bu doğrultu BOBUT uygulamalarında MPS’li madde bankasının uzun süre

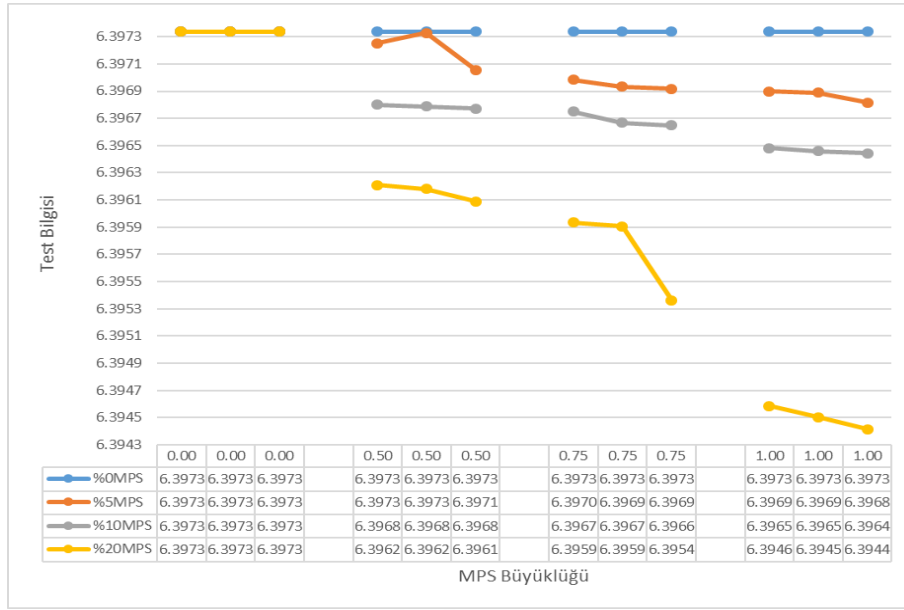
kullanımından kaynaklı, bazı maddelerin madde güçlük parametresinin kolay yönde MPS gösterdiği durumda yetenek kestirimleri önemli düzeyde etkilenmekte ve ortalama yanlışlık açısından daha yanlış sonuçlar elde edilmektedir. Bu durum da ölçme kesinliğini düşürmektedir.

Risk'in (2015) çalışmasında MPS'li koşullar altında MOF değerleri açısından istatistiksel olarak manidar sonuçlar elde edilmese de madde bankasındaki MPS'li madde sayısından çok MPS büyüklüğünün, MOF değerlerini daha çok değiştirdiği belirtilmiştir. Öyle ki en yüksek MOF değeri MPS büyüklüğünün 1.00 olduğu durumda elde edilmiştir. İlgili alanyazında MOF değerlerinin incelendiği bazı çalışmalarda örneklem büyüklüğü, MPS büyüklüğü, MPS'li madde sayısı gibi farklı faktörler açısından manidar sonuçlar elde edilmemesinin nedeninin, MPS koşullarından çok simüle edilen örneklem ve/veya madde bankasının yapaylığından kaynaklı olduğu belirtilmektedir (Risk, 2015; Wells vd., 2002).

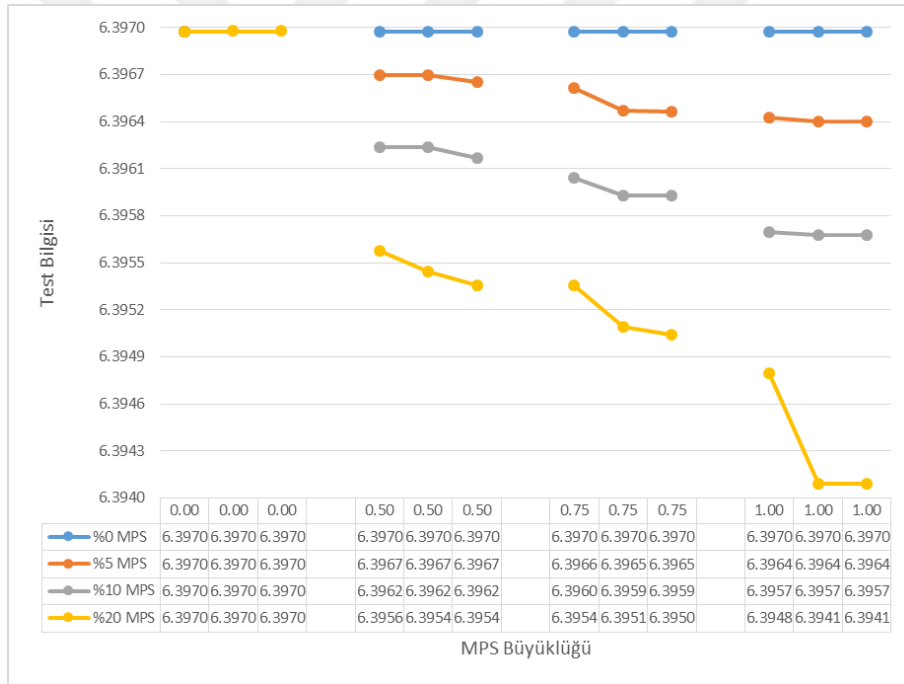
Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı dördüncü ölçüt (bağımlı değişken) TBF değerleridir. TBF değerlerine ilişkin bulgular şekil 6.a, 6.b ve 6.c'de yer almaktadır. Bulguların üç madde bankası için de paralellik göstermesinden dolayı, üç şekile ilişkin yorumlar şekillerin ardından toplu olarak sunulmuştur.



Şekil 6.a. Örneklem n=1000 olduğunda 200 Maddelik Banka için TBF Değerleri



Şekil 6.b. Örneklem n=1000 olduğunda 500 Maddelik Banka için TBF Değerleri



Şekil 6.c. Örneklem n=1000 olduğunda 1000 Maddelik Banka için TBF Değerleri

Şekil 6.a, 6.b ve 6.c. Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre TBF Değerlerinin Karşılaştırılmasına İlişkin Grafikler

Test Bilgi Fonksiyonuna ilişkin değerler incelendiğinde; 1000 kişilik yanıtlayıcı örnekleminde, 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç farklı zaman

noktasında elde edilen yetenek kestirimlerine ait *TBF* değerlerinin azalma eğiliminde olduğu görülmektedir. *TBF* açısından en düşük değerler *MPS* büyüklüğünün 1.00, *MPS* yüzdesinin %20 olduğu ve üçüncü ölçümün alındığı noktada elde edilmiştir. *TBF* değerleri standart hata ile ters orantılıdır; standart hata değeri ne kadar düşükse, testin verdiği bilgi miktarı da o kadar artacaktır (Anıl, 2005). Bu doğrultuda, *MPS* miktarı, *MPS* yüzdesi ve *MPS*'li madde bankası ile yapılan ölçüm sayısı arttıkça testin verdiği bilgi miktarı düşmektedir. Bu eğilim madde bankası büyüklüğü artsa da değişkenlik göstermemektedir. *TBF* değerlerindeki bu değişim sadece, *MPS* olmadığı için temel düzey veri seti için ortaya çıkmamıştır. Alanyazında, *MPS* olduğu durumda *TBF* değerlerinin düşük düzeyde de olsa değişme eğiliminde olduğunu gösteren çalışmalar mevcuttur (Chan vd., 1999; Deng ve Melican, 2009; Guo ve Wang, 2003; 2005).

Chan ve diğerleri (1999), 16 yıllık süreç içerisinde zamana karşı *ASVAB*'ın psikometrik tutarlılığını inceledikleri çalışmalarında, 200 madde ve sekiz alt boyuttan oluşan testin, madde ve test karakteristik eğrilerini karşılaştırmışlardır. *MPS*'nin test karakteristik eğrileri üzerinde düşük düzeyde de olsa farklara yol açtığını belirtmişlerdir. Guo ve Wang (2005) ise geniş ölçekli bir *BOBUT* uygulamasında iki zaman noktası için ölçek kaymasını inceledikleri çalışmalarında test karakteristik eğrileri arasında önemsenmeyecek düzeyde farklılıklar belirlemişlerdir. Ancak çalışmada hem iki zaman noktası için ölçüm alınmış hem de sınava katılan 169.111 kişi arasından rastgele 10 tane 1000 kişilik örneklem çekerek, bu 10 replikasyon üzerinden çalışıldığından, araştırmacılar çalışma sonucunda 100 replikasyon üzerinden çalışarak sonuçların değerlendirilmesi gerektiğini vurgulamışlardır. Bu çalışmada *TBF* açısından fark gözlenmemesinin nedeni hem replikasyon sayısı hem de iki zaman noktası gibi az bir noktada ölçüm alınması olabilir.

Yukarıda verilen *TBF* değerleri açısından elde edilen farkların istatistiksel olarak manidar olup olmadığını incelemek için bağımsız örneklem için yapılan üç faktörlü ANOVA sonuçları Tablo 9'da verilmiştir.

Tablo 9

Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre TBF Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
200	MPS Büyüklüğü	19.342	3	6.447	79578.51*	0.35
	MPS Yüzdesi	8.060	3	2.687	33161.14*	0.14
	Ölçüm	1.075	2	0.538	4422.86*	0.02
	MPS Büyüklüğü*MPS Yüzdesi	16.112	9	1.790	66289.37*	0.30
	MPS Büyüklüğü*Ölçüm	2.148	6	0.358	8837.49*	0.03
	MPS Yüzdesi*Ölçüm	2.148	6	0.358	8837.49*	0.04
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	4.294	18	0.239	17666.74*	0.08
	Hata	1.155	4752	0.000		
	Toplam	54.334	4799			
500	MPS Büyüklüğü	0.000	3	0.000	0.00*	0.00
	MPS Yüzdesi	0.001	3	0.000	316.80*	0.02
	Ölçüm	0.012	2	0.006	3801.60	-
	MPS Büyüklüğü*MPS Yüzdesi	0.000	9	0.000	0.00*	0.00
	MPS Büyüklüğü*Ölçüm	0.003	6	0.000	950.40	-
	MPS Yüzdesi*Ölçüm	0.005	6	0.000	1584.00	-
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.008	18	0.000	2534.40	-
	Hata	0.015	4752	0.000		
	Toplam	0.045	4799			
1000	MPS Büyüklüğü	0.000	3	0.000	0.00*	0.00
	MPS Yüzdesi	0.001	3	0.001	279.53*	0.00
	Ölçüm	0.018	2	0.009	5031.53	-
	MPS Büyüklüğü*MPS Yüzdesi	0.068	9	0.007	19008.00*	0.00
	MPS Büyüklüğü*Ölçüm	0.003	6	0.000	838.59	-
	MPS Yüzdesi*Ölçüm	0.014	6	0.002	3913.41	-
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.008	18	0.000	2236.24	-
	Hata	0.017	4752	0.000		
	Toplam	0.130	4799			

*p<.05

Test Bilgi Fonksiyonu değerlerine ilişkin bağımsız örneklemeler için üç faktörlü ANOVA sonuçları genel olarak incelendiğinde 1000 kişilik yanıtlayıcı örnekleminde, 200 maddelik madde bankası için MPS büyüklüğü, MPS yüzdesi, MPS’li madde bankası ile yapılan ölçüm sayısının hem temel etki, hem ikili etkileşim hem de üçlü etkileşim etkilerinin TBF üzerinde istatistiksel olarak manidar etkisi olduğu görülmüştür. 200 maddelik banka için BOBUT uygulamalarında MPS’li madde bankasının uzun süre kullanımından kaynaklı, bazı maddelerin kolay yönde MPS gösterdiği durumda özellikle MPS büyüklüğü ve MPS büyüklüğü*MPS yüzdesi faktörlerinin TBF değerlerini önemli düzeyde etkilediği sonucu çıkarılabilir. Bir başka deyişle, MPS’li madde bankası testin verdiği bilgi miktarını yani yeterlik kestirimlerinin güvenilirliğini düşürmektedir. Bu etkiler düşük ve yüksek düzeydedir (Cohen, 1988, s.413). 1000 kişilik yanıtlayıcı örnekleminde 500 ve 1000 maddelik madde bankaları için de MPS büyüklüğü, MPS yüzdesi ve MPS büyüklüğü*MPS yüzdesi şeklinde iki faktörün etkileşim etkisi olduğu görüldü de bu etkiler düşük düzeydedir (Cohen, 1988, s.413). Yapılan Post-Hoc test sonuçları da anlamlı fark çıkan tüm faktörlerin tüm düzeyleri açısından fark olduğunu ortaya koymuştur. TBF üzerinde en fazla etkisi olan faktörün ise MPS yüzdesi olduğu görülmektedir.

Alanyazında MPS’nin TBF üzerindeki etkisinin incelendiği çalışmalarda bu bulguyu destekleyen çalışmalar olmakla birlikte (Chan vd., 1999), istatistiksel olarak manidar fark gözlenmeyen çalışmalar da mevcuttur (Deng ve Melican, 2009; Guo ve Wang, 2003; 2005). Chan ve diğerleri (1999) tarafından ASVAB’ın psikometrik tutarlılığının incelendiği çalışmada 200 maddeden 25 tanesi (%12.5) MPS göstermiş ve bu durum TBF değerlerini istatistiksel olarak manidar düzeyde etkilemiştir. Bir başka ifade ile madde bankasında %10 düzeyinde MPS görülmesi TBF değerlerini önemli düzeyde değiştirmiştir. Guo ve Wang (2003), gerçek ve simülatif verilerden iki zaman noktasında ölçüm aldıkları çalışmalarında ise test karakteristik eğrileri bakımından çok düşük miktarda farklar elde edilmiş olup bu farkların pratik anlamda etkisinin önemli olmadığını belirtilmişlerdir. Benzer şekilde Guo ve Wang (2005) tarafından yapılan çalışmada GMAT® sınavında iki zaman noktasında çeşitli simülasyon koşulları ile MPS incelenmiş, TBF açısından istatistiksel olarak manidar fark elde edilememiştir. Ancak MPS’nin etkisinin net bir şekilde gözlemlenebilmesi için ikiden fazla zaman noktasına ihtiyaç vardır (Babcock ve Albano, 2012; Chan vd., 1999). Dolayısıyla belirtilen çalışmalarda manidar fark çıkmamasının nedeni iki noktada ölçüm alınmış olmasından kaynaklanmış olabilir.

Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı son ölçüt (bağımlı değişken), test etkililiğine ilişkin değerleridir. Buna ilişkin bulgular, Tablo 10’da özetlenmiştir.

Tablo 10

Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Test Etkililiği Değerleri

Madde Bankası Büyüklüğü	MPS Büyüklüğü	MPS Yüzdesi	Ölçüm	Ortalama Test Uzunluğu	Maddelerin Ortalama Uygulanma Oranı	χ^2
200	0.00	%0	1.	23.47	0.2330	0.1460
			2.	23.47	0.2330	0.1460
			3.	23.47	0.2330	0.1460
		%5	1.	23.54	0.2336	0.2247
			2.	23.56	0.2345	0.2252
			3.	23.58	0.2345	0.2388
	0.50	%10	1.	23.56	0.2341	0.2451
			2.	23.61	0.2347	0.2472
			3.	23.64	0.2349	0.2688
		%20	1.	23.59	0.2343	0.2470
			2.	23.68	0.2359	0.2473
			3.	23.77	0.2370	0.2953
		%5	1.	23.54	0.2342	0.2340
			2.	23.58	0.2345	0.2388
			3.	23.62	0.2345	0.2537
	0.75	%10	1.	23.59	0.2346	0.2473
			2.	23.64	0.2346	0.2694
			3.	23.64	0.2349	0.2816
		%20	1.	23.63	0.2353	0.2563
			2.	23.77	0.2369	0.2948
			3.	23.80	0.2377	0.3374
		%5	1.	23.56	0.2345	0.2384
			2.	23.59	0.2348	0.2506
			3.	23.61	0.2349	0.2624
	1.00	%10	1.	23.61	0.2347	0.2476
			2.	23.62	0.2349	0.2772
			3.	23.65	0.2350	0.2897
		%20	1.	23.69	0.2357	0.3119
			2.	23.67	0.2367	0.3821
			3.	23.78	0.2374	0.4162
500	0.00	%0	1.	23.41	0.0932	0.1020
			2.	23.41	0.0932	0.1020
			3.	23.41	0.0932	0.1020
	0.50	%5	1.	23.43	0.0933	0.1465

(Devam ediyor)

Tablo 10 (Devam)

Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Test Etkililiği Değerleri

Madde Bankası Büyüklüğü	MPS Büyüklüğü	MPS Yüzdesi	Ölçüm	Ortalama Test Uzunluğu	Maddelerin Ortalama Uygulanma Oranı	X ²
500	0.50	%5	2.	23.44	0.0934	0.1468
			3.	23.47	0.0935	0.1484
			1.	23.44	0.0934	0.1472
		%10	2.	23.47	0.0937	0.1477
			3.	23.51	0.0938	0.1512
			1.	23.46	0.0936	0.1543
		%20	2.	23.54	0.0939	0.1548
			3.	23.64	0.0943	0.1567
			1.	23.43	0.0935	0.2280
	0.75	%5	2.	23.47	0.0935	0.2305
			3.	23.48	0.0936	0.2320
			1.	23.44	0.0934	0.4005
		%10	2.	23.50	0.0937	0.4041
			3.	23.51	0.0937	0.4113
			1.	23.48	0.0937	0.4111
		%20	2.	23.64	0.0943	0.4168
			3.	23.66	0.0945	0.4243
			1.	23.43	0.0935	0.4030
	1.00	%5	2.	23.47	0.0936	0.4074
			3.	23.49	0.0937	0.4090
			1.	23.45	0.0935	0.4032
		%10	2.	23.53	0.0938	0.4045
			3.	23.54	0.0939	0.4115
			1.	23.53	0.0939	0.4135
		%20	2.	23.71	0.0944	0.4213
			3.	23.73	0.0947	0.4378
1000	0.00	%0	1.	23.39	0.0467	0.3685
			2.	23.39	0.0467	0.3685
			3.	23.39	0.0467	0.3685
		%5	1.	23.41	0.0467	0.3685
			2.	23.44	0.0468	0.3706
			3.	23.48	0.0469	0.3706
		%10	1.	23.42	0.0468	0.3698
			2.	23.49	0.0469	0.3755
			3.	23.56	0.0470	0.3761
		%20	1.	23.43	0.0468	0.3713
			2.	23.55	0.0470	0.3727
			3.	23.68	0.0473	0.3770
	0.75	%5	1.	23.42	0.0468	0.3722
			2.	23.47	0.0469	0.3724
			3.	23.49	0.0469	0.3745

(Devam ediyor)

Tablo 10 (Devam)

Örneklem n=1000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Test Etkililiği Değerleri

Madde Bankası Büyüklüğü	MPS Büyüklüğü	MPS Yüzdesi	Ölçüm	Ortalama Test Uzunluğu	Maddelerin Ortalama Uygulanma Oranı	X^2
1000	0.75	%10	1.	23.46	0.0468	0.3737
			2.	23.56	0.0470	0.3763
			3.	23.63	0.0471	0.3763
		%20	1	23.49	0.0469	0.3735
			2	23.68	0.0473	0.3735
			3	23.79	0.0475	0.3798
	1.00	%5	1	23.44	0.0468	0.3725
			2	23.48	0.0469	0.3734
			3	23.49	0.0469	0.3746
		%10	1	23.49	0.0469	0.3770
			2	23.61	0.0472	0.3770
			3	23.67	0.0472	0.3781
		%20	1	23.55	0.0470	0.3799
			2	23.78	0.0475	0.3800
			3	23.81	0.0475	0.3806

Tablo 10 incelendiğinde; 1000 kişilik yanıtlayıcı örnekleminde, 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç farklı zaman noktasında elde edilen yetenek kestirimlerine ait *ortalama test uzunluğu*, *maddelerin ortalama uygulanma oranları* ve X^2 değerlerinin artma eğiliminde olduğu görülmektedir. X^2 değeri gözlenen ve beklenen madde kullanım sıklığı arasındaki farkı ifade etmektedir ve bu değer ne kadar düşükse, testler o kadar etkilidir. X^2 ve ortalama test uzunluğuna ilişkin eğilim madde bankası büyüklüğü arttıkça değişmemiştir. Ancak madde bankası büyüklüğü arttıkça maddelerin ortalama uygulanma oranı azalmıştır. Test etkililiği değerlerindeki bu değişim sadece MPS'nin olmadığı (%0 MPS ve 0.00 MPS büyüklüğü koşulu) temel düzey veri seti için ortaya çıkmamıştır. Bu doğrultuda test etkililiğine ilişkin olarak MPS miktarı, MPS büyüklüğü ve MPS'li madde bankası ile yapılan ölçüm sayısı arttıkça daha az etkili testler elde edildiği belirtilebilir.

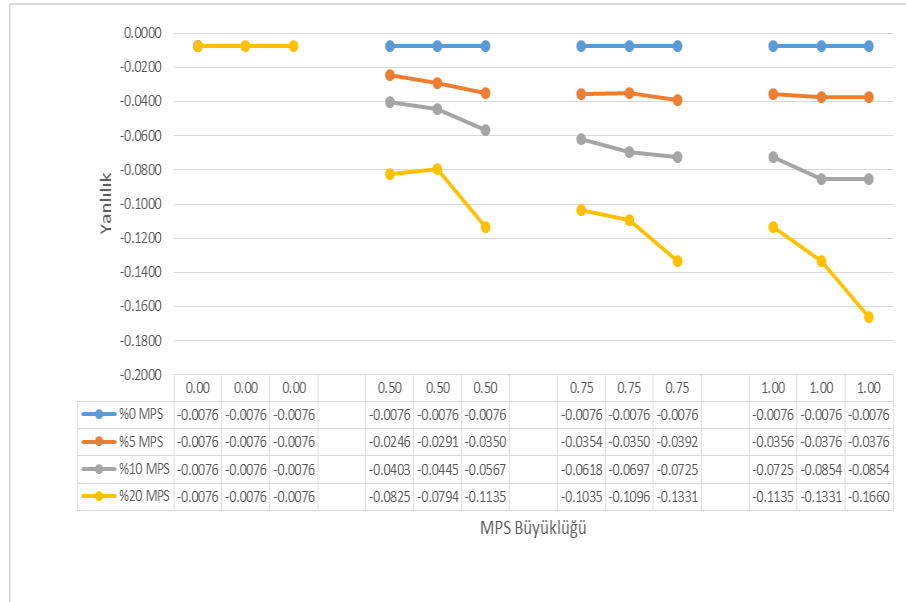
Benzer bulguya alanyazında yapılan bazı çalışmalarda da rastlanmaktadır (Bergstrom ve Stahl, 1992; Lunz ve Stahl, 1993). BOBUT'ta MPS'yi etkileyen önemli faktörlerden biri maddelerin görülme sıklığıdır. Bu doğrultuda etkili testler elde etmede

minimum madde bankası büyüklüğünün 200-300 olması gerektiği belirtilmektedir (Lunz ve Stahl; 1993). Bu çalışma kapsamında MPS'li madde bankasına ilişkin test etkililiği bakımından çok düşük düzeyde farklılıklar elde edilmesinin nedeni minimum 200 maddelik madde bankası ile çalışılması olabilir. Risk'in (2015) çalışmasında ise MPS'li maddelerin olduğu koşullarda daha etkili testler elde edildiği görülmüş ve bu durumun nedeni olarak simüle edilen koşullardan madde bankasındaki maddelerin, örnekleme uygun olmaması olarak belirtilmiştir.

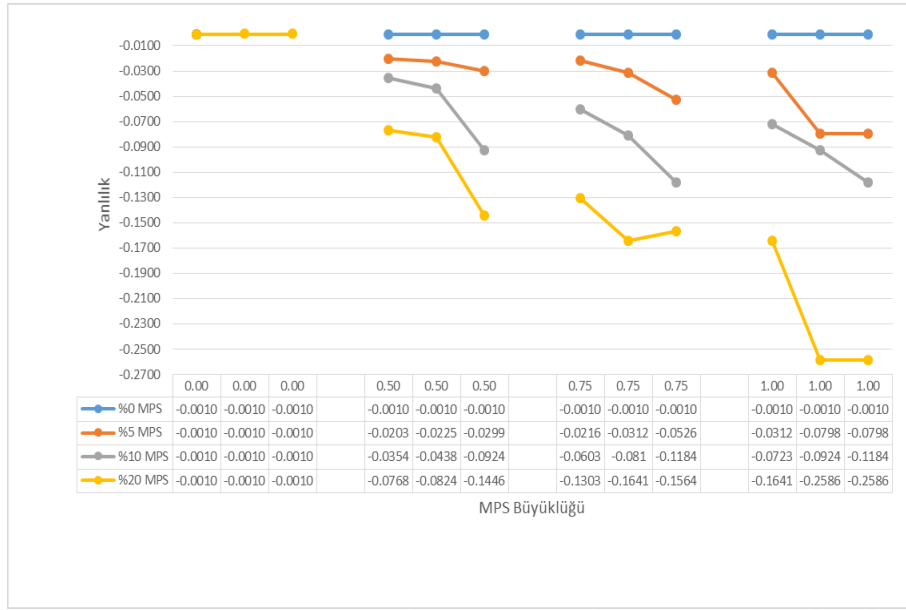
Örneklem Büyüklüğünün 5000 Olduğu Durumda Koşulların Karşılaştırılmasına İlişkin Bulgular

5000 kişilik örneklem büyüklüğü ve 200, 500 ve 1000 maddelik madde bankaları için bulgulara aşağıda yer verilmiştir.

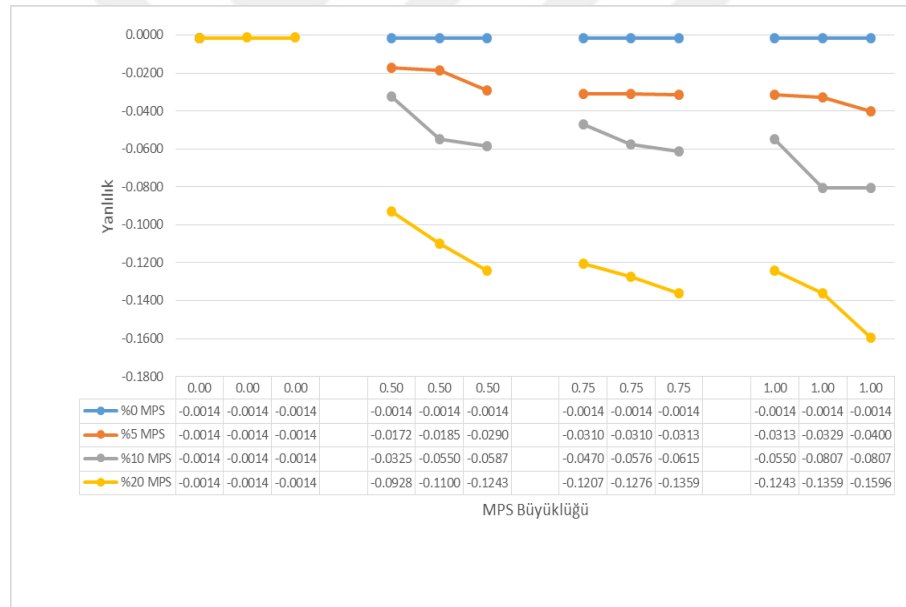
Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı ölçme kesinliğine ilişkin birinci ölçüt (bağımlı değişken), yeterli kestirimlerine ait yanlılık değerleridir. Yanlılık değerlerine ilişkin bulgular şekil 7.a, 7.b ve 7.c'de yer almaktadır. Bulguların üç madde bankası için de paralellik göstermesinden dolayı, üç şekile ilişkin yorumlar şekillerin ardından toplu olarak sunulmuştur.



Şekil 7.a. Örneklem n=5000 olduğunda 200 Maddelik Banka için Yanlılık Değerleri



Şekil 7.b. Örneklem n=5000 olduğunda 500 Maddelik Banka için Yanlılık Değerleri



Şekil 7.c. Örneklem n=5000 olduğunda 1000 Maddelik Banka için Yanlılık Değerleri

Şekil 7.a, 7.b ve 7.c. Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre Yanlılık Değerlerinin Karşılaştırılmasına İlişkin Grafikler

Yanlılığa ilişkin değerler incelendiğinde; 5000 kişilik yanıtlayıcı örnekleminde 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç farklı zaman noktasında

elde edilen yetenek kestirimlerine ait *yanlılık* değerlerinin negatif yönde artma eğiliminde olduğu görülmektedir. Bir başka deyişle MPS miktarı, MPS yüzdesi ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça daha yanlı yetenek kestirimleri elde edilmekte, dolayısıyla ölçme kesinliği de düşmektedir. Bu eğilim madde bankası büyüklüğü artsa da değişkenlik göstermemiştir. Bir diğer ifadeyle, BOBUT uygulamalı testlerde MPS’li madde miktarı ve bankadaki MPS’li madde yüzdesi arttığı sürece, madde bankasının büyümesi, yeterlik kestirimlerinin yanlı olmasını engellememektedir. Yanlılık değerlerindeki bu değişim sadece, MPS olmadığı için temel düzey veri seti için ortaya çıkmamıştır.

Negatif yanlılık değerleri, bireylerin kestirilen yetenek değerlerinin gerçek yeteneklerinden düşük olduğunu göstermektedir. MPS’nin kolay yönde gerçekleşmesinden dolayı yanlılık değerlerinin pozitif yönde artmış olması beklenirdi. Alanyazında MPS’nin yanlılık değerlerini negatif yönde etkilediğini gösteren çalışmalar mevcuttur (Aksu Dünya, 2017; Chen, 2013; Risk, 2015; Rupp ve Zumbo, 2003b). Rupp ve Zumbo (2003b) tarafından yapılan çalışmada özellikle madde güçlüğü’nün 0.40 logit kadar değiştiği durumda negatif yanlılık değerleri elde edildiğini ve bu durumun da test bilgi fonksiyonu değerini düşürdüğü belirtilmiştir. Risk (2015) de MPS’nin etkisini BOBUT ortamında incelediği çalışmasında, MPS’li madde bankasının yetenek kestirimlerini etkilediğini ve bu durumun da negatif yanlılık kestirimleri ortaya çıkardığını belirtmiştir. Bu çalışma kapsamında negatif yanlılık değerleri elde edilmesinin nedenleri; sadece kolay yönde gerçekleşen MPS durumunun ele alınmış olması (Aksu Dünya, 2017; Wei, 2013) veya bireylerin madde bankasında yeteneklerine uygun maddelerle karşılaşmalarına rağmen bunları yanlış yanıtlaması olabilir.

Madde parametre sapmasının yanlılık değerlerini çok fazla etkilemediğini gösteren çalışmalar da mevcuttur (Guo ve Wang, 2003; Kingsbury ve Wise, 2011). Örneğin Kingsbury ve Wise (2011) Measure of Academic Progress sınavında madde parametrelerinin sürekliliği ve bunların yetenek kestirimlerine etkisini inceledikleri çalışmalarında 15 yıllık süreç içerisinde altı zaman noktasında ölçüm almışlar ve madde güçlükleri açısından ölçümler arasında yüksek düzey korelasyonlar bulmuşlardır. Dolayısıyla bu durum da yetenek kestirimlerine ilişkin yanlılığı çok düşük ve önemli olmayan düzeyde etkilemiştir. Çünkü çalışmada madde güçlüklerindeki MPS miktarının standart sapması 0.01’den daha düşük çıkmıştır. Bir başka ifade ile MPS miktarı çok düşük düzeyde olduğu için yanlılık değerleri çok fazla etkilenmemiştir. Ancak özellikle 0.50 logit ve üzerinde MPS olan durumlar parametre kestirimlerini önemli düzeyde

etkilemektedir (Han ve Wells, 2007; Wollack vd., 2005). Bu çalışma kapsamında da 0.50 logit ve üzerinde MPS olan koşullar değerlendirildiği için yanlılık değerleri açısından düşük de olsa farklar elde edildiği belirtilebilir.

Yukarıda verilen yanlılık değerleri açısından elde edilen farkların istatistiksel olarak manidar olup olmadığını incelemek için bağımsız örneklemeler için yapılan üç faktörlü ANOVA sonuçları Tablo 11’de verilmiştir.

Tablo 11

Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Yanlılık Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
200	MPS Büyüklüğü	0.383	3	0.128	18571.59*	0.10
	MPS Yüzdesi	2.971	3	0.990	144063.18*	0.76
	Ölçüm	0.166	2	0.083	8049.31*	0.04
	MPS Büyüklüğü*MPS Yüzdesi	0.117	9	0.013	5673.31*	0.03
	MPS Büyüklüğü*Ölçüm	0.011	6	0.002	533.39*	0.00
	MPS Yüzdesi*Ölçüm	0.104	6	0.017	5042.94*	0.02
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.018	18	0.001	872.82*	0.00
	Hata	0.098	4752	0.000		
	Toplam	3.868	4799			
500	MPS Büyüklüğü	2.081	3	0.694	99888.00*	0.18
	MPS Yüzdesi	6.748	3	2.249	323904.00*	0.60
	Ölçüm	1.069	2	0.535	51312.00*	0.09
	MPS Büyüklüğü*MPS Yüzdesi	0.806	9	0.090	38688.00*	0.07
	MPS Büyüklüğü*Ölçüm	0.224	6	0.037	10752.00*	0.02
	MPS Yüzdesi*Ölçüm	0.183	6	0.031	8784.00*	0.01
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.212	18	0.012	10176.00*	0.02
	Hata	0.099	4752	0.000		
	Toplam	11.422	4799			
1000	MPS Büyüklüğü	0.228	3	0.076	11169.65*	0.04
	MPS Yüzdesi	4.403	3	1.468	215701.61*	0.88
	Ölçüm	0.160	2	0.080	7838.35*	0.03

(Devam ediyor)

Tablo 11 (Devam)

Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Yanlılık Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
1000	MPS Büyüklüğü*MPS Yüzdesi	0.033	9	0.004	1616.66*	0.00
	MPS Büyüklüğü*Ölçüm	0.017	6	0.003	832.82*	0.00
	MPS Yüzdesi*Ölçüm	0.049	6	0.008	2400.49*	0.01
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.007	18	0.000	342.93*	0.00
	Hata	0.097	4752	0.000		
	Toplam	4.994	4799			

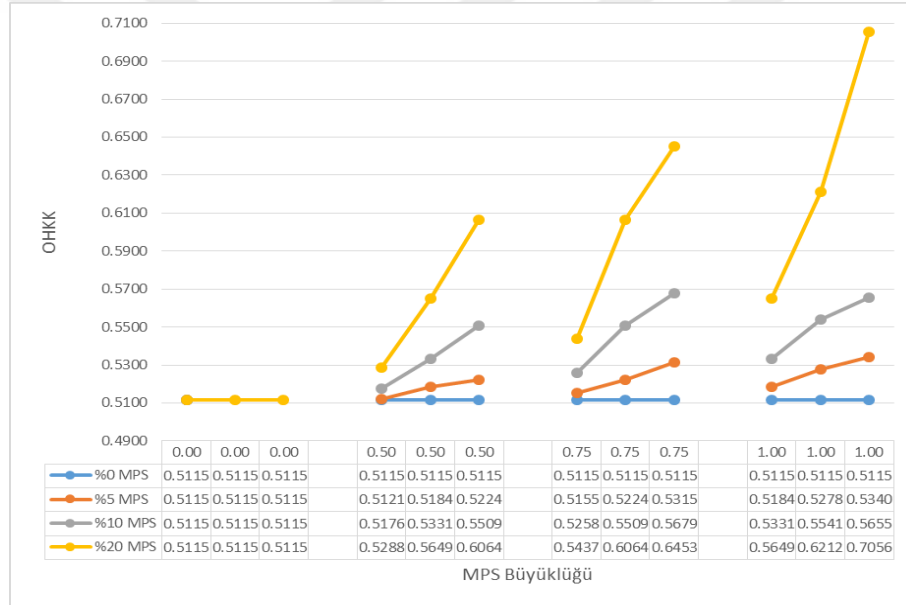
*p<.05

Yanlılığa ilişkin olarak bağımsız örneklem için üç faktörlü ANOVA sonuçları genel olarak incelendiğinde 5000 kişilik yanıtlayıcı örnekleme ve 200, 500 ve 1000 maddelik madde bankaları için MPS büyüklüğü, MPS yüzdesi, MPS’li madde bankası ile yapılan ölçüm sayısının hem temel etki, hem ikili etkileşim hem de üçlü etkileşim etkilerinin yanlılık üzerinde istatistiksel olarak manidar etkisi olduğu görülmüştür. Bunlar genel olarak düşük etki büyüklüklerine sahiptir (Cohen, 1988, s. 413). Yapılan Post-Hoc test sonuçları da tüm faktörlerin tüm düzeyleri açısından fark olduğunu ortaya koymuştur. Yeterlik kestirimlerinin yanlılığı üzerinde en fazla etkisi olan faktör; MPS yüzdesidir. Bir başka ifade ile MPS’ye ilişkin olarak karşılaştırılan faktörler yanlılığı ve dolayısıyla ölçme kesinliğini olumsuz yönde etkilemektedir. Buna göre, BOBUT uygulamalarında MPS’li madde bankasının uzun süre kullanımından kaynaklı, bazı maddelerin kolay yönde MPS gösterdiği durumda yetenek kestirimleri önemli düzeyde etkilenmekte ve yeterlikler yanlı olarak kestirilme eğiliminde olmaktadır. Benzer bulguya alanyazında yapılan bazı çalışmalarda da rastlanmaktadır (Abad vd., 2010; Babcock ve Albano, 2012).

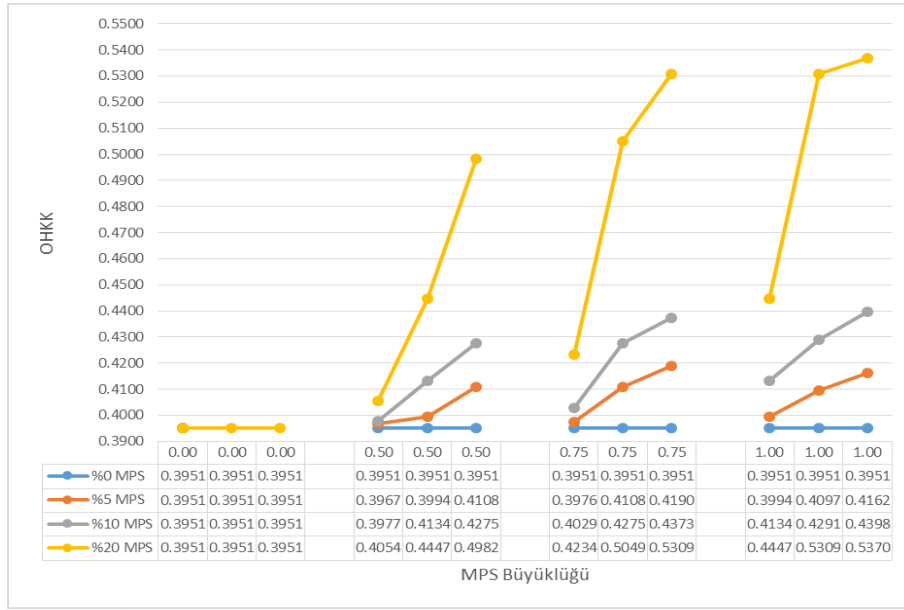
Abad ve diğerleri (2010) MPS’nin BOBUT üzerindeki etkilerini inceledikleri ve 3PLM’yi kullandıkları çalışmalarında, MPS’nin yanlılığı negatif yönde ve önemli düzeyde etkilediğini belirtmişlerdir. Madde güçlük ve ayırıcılık parametrelerindeki değişimin yetenek kestirimlerini, dolayısıyla da yanlılığı orta düzeyde etkilediği görülmüş ve bu nedenle madde bankalarının sürekli yenilenmesi gerektiğini

belirtmişlerdir. Chen (2013) ise MPS'nin ölçekleme maddelerinde görülmesi durumunda yanlılık değerlerinin negatif yönde arttığını belirtmişlerdir. Dolayısıyla MPS, ölçekleme sürecini de olumsuz yönde etkileyen bir faktör olarak ortaya çıkmaktadır. Ancak bazı araştırmalar MPS'nin yanlılık üzerinde önemli bir etkisinin olmadığını da göstermektedir (Guo ve Wang, 2010; Risk, 2015). Ancak bu çalışmalarda da iki zaman noktasında ölçüm alınmıştır.

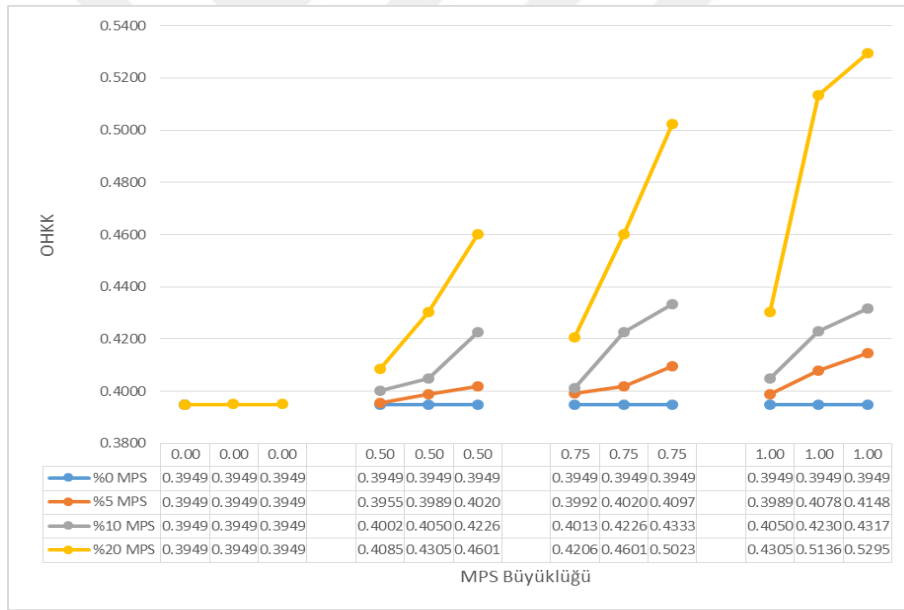
Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı ölçme kesinliğine ilişkin ikinci ölçüt (bağımlı değişken), yeterlik kestirimlerine ait OHKK değerleridir. OHKK değerlerine ilişkin bulgular şekil 8.a, 8.b ve 8.c'de yer almaktadır. Bulguların üç madde bankası için de paralellik göstermesinden dolayı, üç şekile ilişkin yorumlar şekillerin ardından toplu olarak sunulmuştur.



Şekil 8.a. Örneklem n=5000 olduğunda 200 Maddelik Banka için OHKK Değerleri



Şekil 8.b. Örneklem n=5000 olduğunda 500 Maddelik Banka için OHKK Değerleri



Şekil 8.c. Örneklem n=5000 olduğunda 1000 Maddelik Banka için OHKK Değerleri

Şekil 8.a, 8.b ve 8.c. Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre OHKK Değerlerinin Karşılaştırılmasına İlişkin Grafikler

Ortalama Hataların Kareköne ilişkin değerler incelendiğinde; 5000 kişilik yanıtlayıcı örnekleminde 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç

farklı zaman noktasında elde edilen yetenek kestirimlerine ait *OHKK* değerlerinin artma eğiliminde olduğu görülmektedir. *OHKK* açısından en yüksek değerler MPS büyüklüğünün 1.00, MPS yüzdesinin %20 olduğu ve üçüncü ölçümün alındığı noktada elde edilmiştir. Gerçek ve kestirilen yetenek değerleri arasındaki hatayı gösteren *OHKK* değeri sıfıra ne kadar yakınsa *BOBUT* uygulaması, gerçek yetenek değerine o kadar yakın kestirim yapmış demektir. Bu doğrultuda, MPS miktarı, MPS yüzdesi ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça daha hatalı yetenek kestirimleri elde edilmekte, dolayısıyla ölçme kesinliği de düşmektedir. Bu eğilim madde bankası büyüklüğü artsa da değişkenlik göstermemektedir. *OHKK* değerlerindeki bu değişim sadece MPS’nin olmadığı temel düzey veri seti için ortaya çıkmamıştır.

Şekil 8.a- 8.c incelendiğinde elde edilen bir diğer bulgu 200, 500 ve 1000 maddelik madde bankaları için madde bankası büyüklüğü arttıkça *OHKK* değerleri azalma eğilimi göstermiştir. Risk (2015) tarafından yapılan çalışmada da 300, 500 ve 1000 maddelik madde bankaları kullanılmış ve madde bankası büyüklüğü arttıkça *OHKK* değerlerinin azaldığı görülmüştür. Bu durum, madde bankası büyüklüğü arttıkça gerçek ve kestirilen yetenek değeri arasındaki hatanın bir miktar düştüğünü göstermektedir. Bir başka ifade ile madde bankası büyüklüğü arttıkça, bireyler gerçek yeteneklerine daha yakın maddelerle karşılaşmış ve bu nedenle hata miktarı düşmüş olabilir.

Madde parametre sapmasına ilişkin koşulların *OHKK* değerlerini arttırdığına, alanyazında yapılan bazı çalışmalarda da rastlanmaktadır (Aksu Dünya, 2017; Babcock ve Albano, 2012; Chen, 2013; Risk, 2015; Wells vd., 2002). Bu çalışmadaki duruma benzer şekilde Babcock ve Albano (2012) da en düşük *OHKK* değerini temel düzey veri seti için elde etmiş olup, özellikle madde bankasındaki MPS’li madde sayısı arttıkça hata değerinin daha fazla arttığını belirtmişlerdir. Aksu Dünya (2017) de madde bankasındaki MPS’li madde sayısı ve MPS’li maddelerden etkilenen birey sayısı arttıkça *OHKK* değerlerinin bu çalışmadakine benzer şekilde artmasının, ölçme kesinliğini düşürdüğünü belirtmiştir.

Yukarıda verilen *OHKK* değerleri açısından elde edilen farkların istatistiksel olarak manidar olup olmadığını incelemek için bağımsız örneklemeler için yapılan üç faktörlü ANOVA sonuçları Tablo 12’de verilmiştir.

Tablo 12

Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre OHKK Değerlerinin Karşılaştırılması

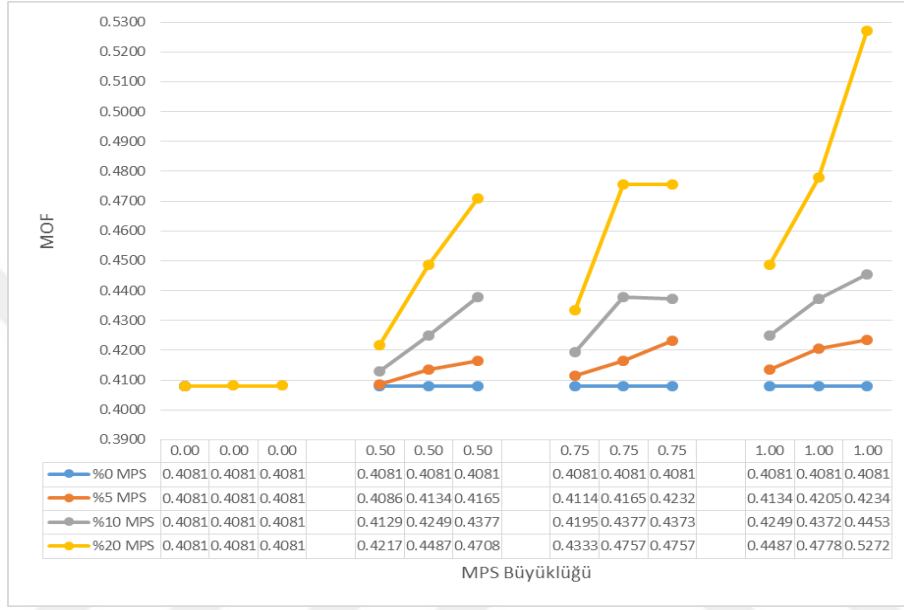
Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
200	MPS Büyüklüğü	0.408	3	0.136	7400.06*	0.07
	MPS Yüzdesi	2.760	3	0.920	50059.24*	0.47
	Ölçüm	1.226	2	0.613	22236.46*	0.21
	MPS Büyüklüğü*MPS Yüzdesi	0.267	9	0.030	4842.69*	0.04
	MPS Büyüklüğü*Ölçüm	0.044	6	0.007	798.05*	0.00
	MPS Yüzdesi*Ölçüm	0.706	6	0.118	12805.01*	0.12
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.085	18	0.005	1541.68*	0.01
	Hata	0.262	4752	0.000		
	Toplam	5.758	4799			
500	MPS Büyüklüğü	0.160	3	0.053	11880.00*	0.03
	MPS Yüzdesi	2.723	3	0.908	202182.75*	0.53
	Ölçüm	1.089	2	0.545	80858.25*	0.21
	MPS Büyüklüğü*MPS Yüzdesi	0.087	9	0.010	6459.75*	0.01
	MPS Büyüklüğü*Ölçüm	0.151	6	0.025	11211.75*	0.03
	MPS Yüzdesi*Ölçüm	0.602	6	0.100	44698.50*	0.11
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.206	18	0.011	15295.50*	0.04
	Hata	0.064	4752	0.000		
	Toplam	5.082	4799			
1000	MPS Büyüklüğü	0.299	3	0.100	22916.90*	0.08
	MPS Yüzdesi	1.702	3	0.567	130450.06*	0.48
	Ölçüm	0.673	2	0.337	51582.19*	0.19
	MPS Büyüklüğü*MPS Yüzdesi	0.240	9	0.027	18394.84*	0.06
	MPS Büyüklüğü*Ölçüm	0.063	6	0.011	4828.65*	0.02
	MPS Yüzdesi*Ölçüm	0.372	6	0.062	28512.00*	0.10
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.064	18	0.004	4905.29*	0.02
	Hata	0.062	4752	0.000		
	Toplam	3.475	4799			

*p<.05

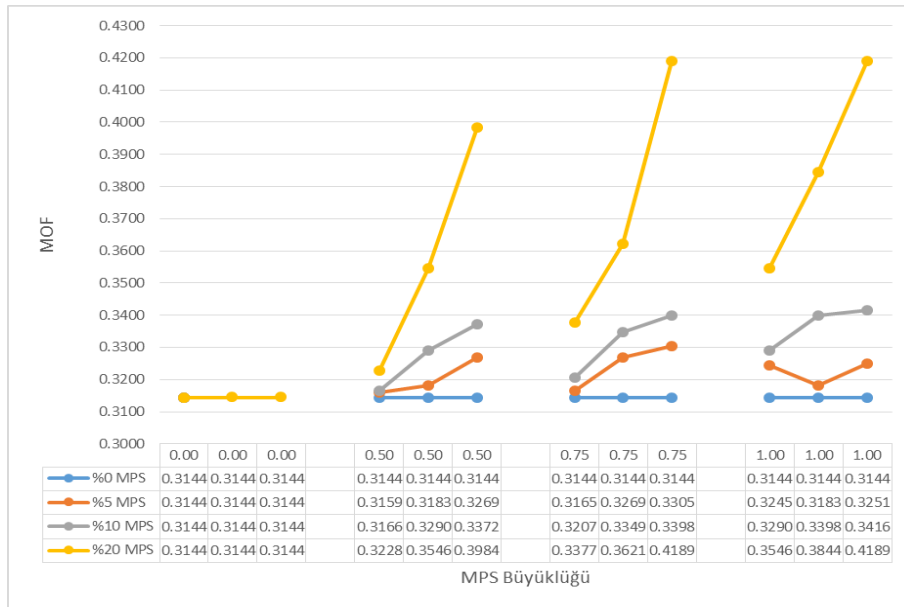
Ortalama Hataların Karekökü değerlerine ilişkin bağımsız örneklemeler için üç faktörlü ANOVA sonuçları genel olarak incelendiğinde; 5000 kişilik yanıtlayıcı örneklemi ve 200, 500 ve 1000 maddelik madde bankaları için MPS büyüklüğü, MPS yüzdesi, MPS'li madde bankası ile yapılan ölçüm sayısının hem temel etki, hem ikili etkileşim hem de üçlü etkileşim etkilerinin OHKK üzerinde manidar etkisi olduğu bulunmuştur. Bunlar çoğunlukla düşük ve yüksek etki büyüklüklerine sahiptir (Cohen, 1988, s.413). Yapılan Post-Hoc test sonuçları da tüm faktörlerin tüm düzeyleri açısından fark olduğunu ortaya koymuştur. Yeterlik kestirimlerine ilişkin OHKK üzerinde en fazla etkisi olan faktör; MPS yüzdesidir. Bir başka ifade ile MPS'ye ilişkin belirtilen faktörler OHKK'yı yani hata değerlerini ve dolayısıyla ölçme kesinliğini olumsuz yönde etkilemektedir. Bu doğrultu BOBUT uygulamalarında MPS'li madde bankasının uzun süre kullanımından kaynaklı, bazı maddelerin madde güçlük parametresinin kolay yönde MPS gösterdiği durumda yetenek kestirimleri önemli düzeyde etkilenmekte ve daha hatalı sonuçlar elde edilmektedir. Bu bulgu alanyazında bazı çalışmalarla da desteklenmektedir (Aksu Dünya; 2017; Babcock ve Albano, 2012; Risk, 2015;). Babcock ve Albano (2012) çalışmasında MPS'li koşulların OHKK değerlerini arttırdığını ve bu değer üzerinde en fazla etkisi olan faktörün MPS büyüklüğü olduğunu belirtmişlerdir. Çalışmada bu bulgunun yanında özellikle, test maddelerinin görece kolay olduğu test durumlarında, zor yönde MPS gözlemlenmesinin yetenek kestirimlerine ilişkin OHKK değerlerini daha olumsuz etkilediği de belirtilmiştir (Babcock ve Albano, 2012). Bu çalışma kapsamında ise görece düşük miktarlarda OHKK değişiminin gözlenmesinin nedeni sadece kolay yönde değişen MPS koşullarının ele alınmış olmasından kaynaklı olabilir.

Alanyazında MPS'nin OHKK üzerinde istatistiksel olarak manidar etkisinin olmadığını belirten çalışmalar da mevcuttur (Chen, 2013; Wells vd., 2002). Örneğin, Chen (2013) madde bankasındaki MPS'li madde yüzdesi arttıkça OHKK değerleri artsa da bu artışın düşük miktarlarda olduğunu ve istatistiksel olarak manidar olmadığını belirtmiştir. Ancak, belirtilen çalışmalarda simüle edilen örneklem büyüklüğü, madde bankası büyüklüğü ve MPS koşulları değişkenlik göstermektedir. Bu doğrultuda, BOBUT uygulamalarında madde bankasında MPS'li madde bulunması yetenek kestirimlerine ilişkin OHKK değerini olumsuz yönde etkilerken hangi faktörün bu değeri olumsuz yönde etkilediği ve/veya etkilemediği, kullanılan MTK modeli, örneklem büyüklüğü, madde bankası büyüklüğü veya MPS koşuluna göre değişkenlik gösterebilir.

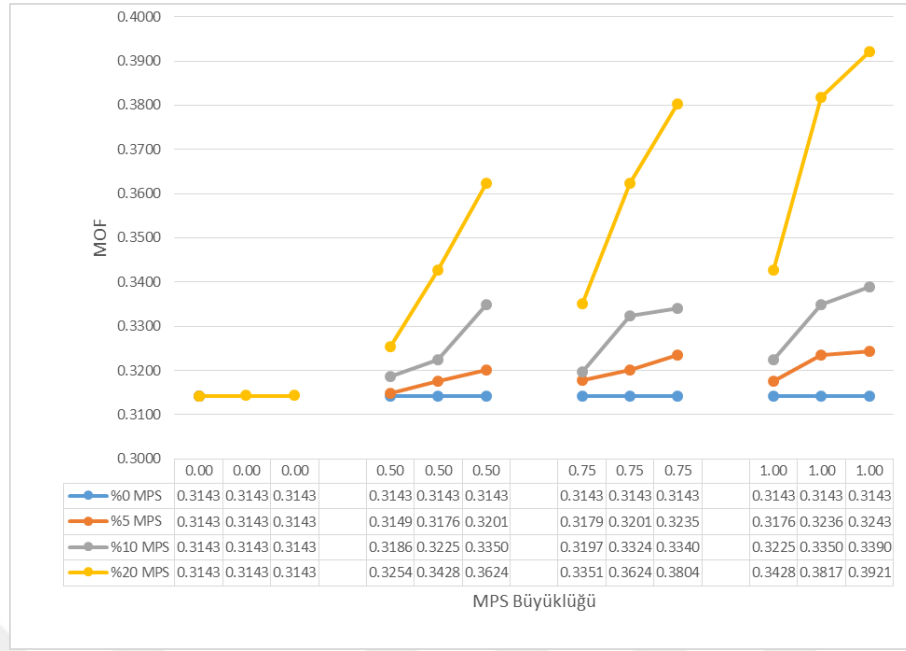
Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı ölçme kesinliğine ilişkin üçüncü ölçüt (bağımlı değişken), yeterlik kestirimlerine ait MOF değerleridir. MOF değerlerine ilişkin bulgular şekil 9.a, 9.b ve 9.c’de yer almaktadır. Bulguların üç madde bankası için de paralellik göstermesinden dolayı, üç şekile ilişkin yorumlar şekillerin ardından toplu olarak sunulmuştur.



Şekil 9.a. Örneklem n=5000 olduğunda 200 Maddelik Banka için MOF Değerleri



Şekil 9.b. Örneklem n=5000 olduğunda 500 Maddelik Banka için MOF Değerleri



Şekil 9.c. Örneklem n=5000 olduğunda 1000 Maddelik Banka için MOF Değerleri

Şekil 9.a, 9.b ve 9.c. Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre MOF Değerlerinin Karşılaştırılmasına İlişkin Grafikler

Maksimum Ortalama Fark değerlerine ilişkin grafikler incelendiğinde; 5000 kişilik yanıtlayıcı örnekleminde 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç farklı zaman noktasında elde edilen yetenek kestirimlerine ait MOF değerlerinin artma eğiliminde olduğu görülmektedir. MOF açısından en yüksek değerler MPS büyüklüğünün 1.00, MPS yüzdesinin %20 olduğu ve üçüncü ölçümün alındığı noktada elde edilmiştir. Gerçek ve kestirilen edilen yetenek değerleri arasındaki ortalama yanlılığı ifade eden MOF değerleri sıfıra ne kadar yakınsa o kadar doğru yetenek kestirimi yapıldığı anlamına gelmektedir. Bu doğrultuda, MPS miktarı, MPS yüzdesi ve MPS'li madde bankası ile yapılan ölçüm sayısı arttıkça daha yanlı yetenek kestirimleri elde edilmekte, dolayısıyla ölçme kesinliği de düşmektedir. Bu eğilim madde bankası büyüklüğü artsa da değişkenlik göstermemektedir. MOF değerlerindeki bu değişim sadece, MPS olmadığı için temel düzey veri seti için ortaya çıkmamıştır.

Alanyazında MOF ile ilgili bu bulguyu destekleyen çalışmalar bulunmaktadır (Chen, 2013; Park vd., 2016; Risk, 2015; Wells vd., 2002). Bu çalışma kapsamında sadece madde güçlük parametresinde kolay yönde değişim olması durumunda MOF

değerlerinde bir artış gözlemlenirken, Wells ve diğerleri (2002) tarafından yapılan ve 3PLM'nin kullanıldığı çalışmada, hem güçlük hem de ayırıcılık parametresinde aynı anda MPS görülmesi durumunda MOF değerlerinin daha çok etkilendiği sonucuna ulaşılmıştır.

Elde edilen bir diğer bulgu, madde bankası büyüklüğü arttıkça MOF değerleri hem temel düzey veri seti hem de MPS'li veri setleri için azalmaktadır. Bir başka ifade ile madde bankası büyüklüğü arttıkça ortalama yanlışlık değerleri açısından daha yansız yetenek kestirimleri elde edilmiş, dolayısıyla ölçme kesinliği artmıştır. Bu durumun nedeni; madde bankası büyüklüğü arttıkça, bireyler gerçek yeteneklerine daha yakın maddelerle karşılaşmış ve bu nedenle hata miktarı düşmüş olabilir.

Yukarıda verilen MOF değerleri açısından elde edilen farkların istatistiksel olarak manidar olup olmadığını incelemek için bağımsız örneklemeler için yapılan üç faktörlü ANOVA sonuçları Tablo 13'te verilmiştir.

Tablo 13

Örneklem n=5000 olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre MOF Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
200	MPS Büyüklüğü	0.148	3	0.049	2704.98*	0.06
	MPS Yüzdesi	1.094	3	0.365	19994.95*	0.48
	Ölçüm	0.389	2	0.195	7109.72*	0.17
	MPS Büyüklüğü*MPS Yüzdesi	0.089	9	0.010	1626.65*	0.04
	MPS Büyüklüğü*Ölçüm	0.027	6	0.005	493.48*	0.01
	MPS Yüzdesi*Ölçüm	0.181	6	0.030	3308.12*	0.08
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.048	18	0.003	877.29*	0.02
	Hata	0.260	4752	0.000		
	Toplam	2.236	4799			
	MPS Büyüklüğü	0.076	3	0.025	9504.00*	0.03
500	MPS Yüzdesi	1.266	3	0.422	158316.63*	0.53
	Ölçüm	0.499	2	0.250	62401.26*	0.21
	MPS Büyüklüğü*MPS Yüzdesi	0.052	9	0.006	6502.74*	0.02
	MPS Büyüklüğü*Ölçüm	0.013	6	0.002	1625.68*	0.00
	MPS Yüzdesi*Ölçüm	0.390	6	0.065	48770.53*	0.16

(Devam ediyor)

Tablo 13 (Devam)

Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre MOF Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
500	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.012	18	0.001	1500.63*	0.00
	Hata	0.038	4752	0.000		
	Toplam	2.346	4799			
1000	MPS Büyüklüğü	0.079	3	0.026	10146.16*	0.06
	MPS Yüzdesi	0.728	3	0.243	93498.81*	0.58
	Ölçüm	0.205	2	0.103	26328.65*	0.16
	MPS Büyüklüğü*MPS Yüzdesi	0.054	9	0.006	6935.35*	0.04
	MPS Büyüklüğü*Ölçüm	0.021	6	0.004	2697.08*	0.01
	MPS Yüzdesi*Ölçüm	0.108	6	0.018	13870.70*	0.08
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.022	18	0.001	2825.51*	0.01
	Hata	0.037	4752	0.000		
	Toplam	1.254	4799			

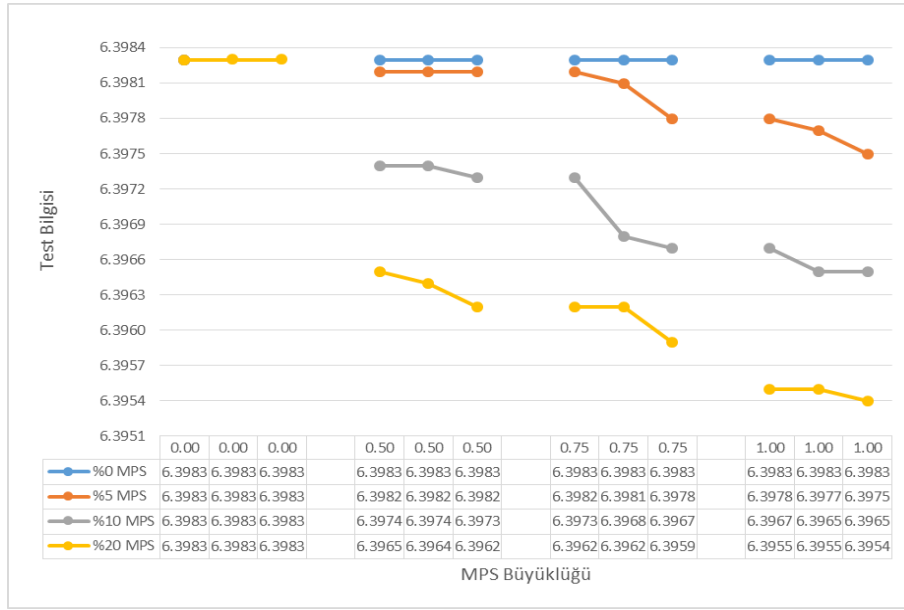
*p<.05

Maksimum Ortalama Fark değerlerine ilişkin bağımsız örneklem için üç faktörlü ANOVA sonuçları genel olarak incelendiğinde 5000 kişilik yanıtlayıcı örnekleme ve 200, 500 ve 1000 maddelik madde bankaları için MPS büyüklüğü, MPS yüzdesi, MPS’li madde bankası ile yapılan ölçüm sayısının hem temel etki, hem ikili etkileşim hem de üçlü etkileşim etkilerinin MOF üzerinde istatistiksel olarak manidar etkisi olduğu görülmüştür. Bunlar genelde düşük etki büyüklüklerine sahiptir (Cohen, 1988, s.413). Yapılan Post-Hoc test sonuçları da tüm faktörlerin tüm düzeyleri açısından fark olduğunu ortaya koymuştur. Yeterlik kestirimlerine ilişkin MOF üzerinde en fazla etkisi olan faktör; MPS yüzdesidir. Bu doğrultta BOBUT uygulamalarında MPS’li madde bankasının uzun süre kullanımından kaynaklı, bazı maddelerin kolay yönde MPS gösterdiği durumda yetenek kestirimleri önemli düzeyde etkilenmekte ve ortalama yanlılık açısından daha yanlı sonuçlar elde edilmektedir. Bu durum da ölçme kesinliğini düşürmektedir. Ulaşılabilen alanyazında MOF değerleri açısından bağımsız örneklem için faktöriyel ANOVA sonucu istatistiksel olarak manidar sonuç elde edilen çalışmaya

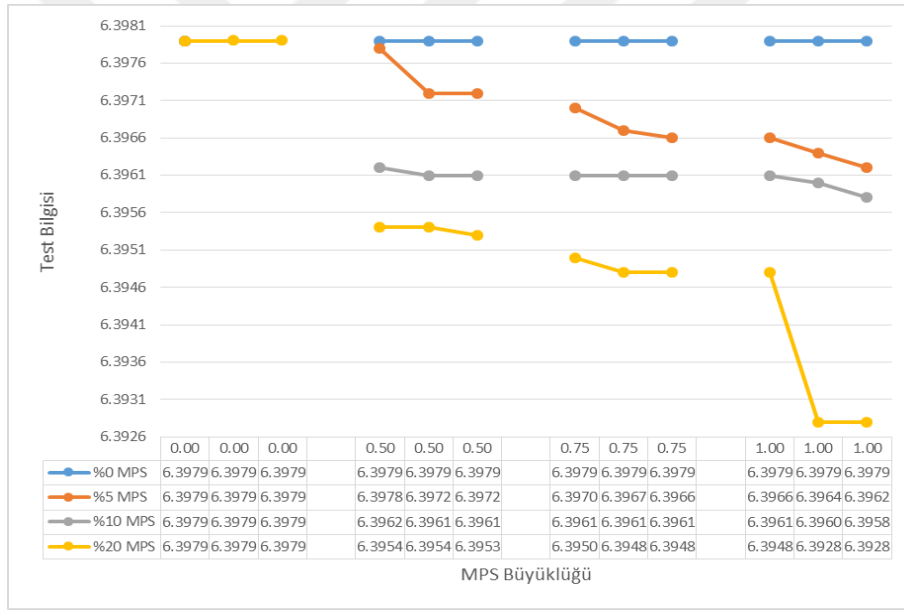
rastlanmamıştır. Sadece Han ve Guo (2011) tarafından MPS'nin BOBUT üzerindeki etkisinin incelendiği çalışmada, MPS eğer %30 ve daha fazla bireyi etkiliyorsa, o zaman MOF açısından istatistiksel olarak manidar sonuçlar elde edildiği bulgusuna ulaşılmıştır. Ancak belirtilen çalışmada MPS'nin tüm bireyleri değil, belli oranda bireyi etkilediği durumlar ele alınmıştır.

Madde parametre sapması açısından MOF değerlerinin incelendiği çalışmalarda örneklem büyüklüğü, MPS büyüklüğü, MPS'li madde sayısı gibi farklı faktörler açısından manidar sonuçlar elde edilmemesinin nedeninin, MPS koşullarından çok MTK ölçeğinin MPS koşullarına dayanıklılığı ve/veya simüle edilen örneklem ve/veya madde bankasının yapaylığından kaynaklı olduğu belirtilmektedir (Risk, 2015; Wells vd., 2002). Örneğin, Risk (2015) 500 kişilik örneklem büyüklüğü, 300, 500 ve 1000 maddelik madde bankası büyüklüğü, MPS büyüklüğünün 0.50, 0.75, 1.00 logit büyüklüğünde olduğu ve 50, 75 ve 100 maddede MPS görüldüğü koşul ile çalışmıştır. Çalışma sonunda belirtilen faktörler arasında MPS büyüklüğünün MOF değerlerini daha çok değiştirdiğini ancak bu değişimin istatistiksel olarak manidar olmadığı sonucuna ulaşmıştır. Bu nedenle çalışılan örneklem büyüklüğü, madde bankası büyüklüğü, BOBUT ve MPS koşullarına göre elde edilen bulgular farklılık gösterebilir. Bu çalışma kapsamında ilgili alanyazından MPS'yi en fazla etkileyebilecek koşullar seçildiği için istatistiksel olarak manidar farklar elde edildiği düşünülmektedir.

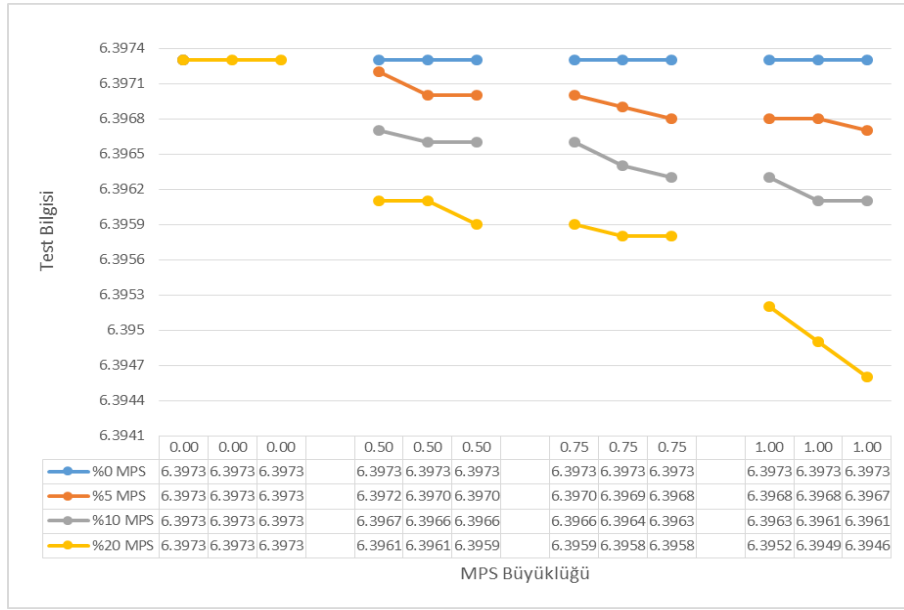
Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı dördüncü ölçüt (bağımlı değişken) TBF değerleridir. TBF değerlerine ilişkin bulgular şekil 10.a, 10.b ve 10.c'de yer almaktadır. Bulguların üç madde bankası için de paralellik göstermesinden dolayı, üç şekile ilişkin yorumlar şekillerin ardından toplu olarak sunulmuştur.



Şekil 10.a. Örneklem n=5000 olduğunda 200 Maddelik Banka için TBF Değerleri



Şekil 10.b. Örneklem n=5000 olduğunda 500 Maddelik Banka için TBF Değerleri



Şekil 10.c. Örneklem n=5000 olduğunda 1000 Maddelik Banka için TBF Değerleri

Şekil 10.a, 10.b ve 10.c. Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zaman Noktasına göre TBF Değerlerinin Karşılaştırılmasına İlişkin Grafikler

Test Bilgi Fonksiyonuna ilişkin değerler incelendiğinde; 5000 kişilik yanıtlayıcı örnekleminde 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç farklı zaman noktasında elde edilen yetenek kestirimlerine ait *test bilgi* değerlerinin azalma eğiliminde olduğu görülmektedir. TBF açısından en düşük değerler MPS büyüklüğünün 1.00, MPS yüzdesinin %20 olduğu ve üçüncü ölçümün alındığı noktada elde edilmiştir. TBF değerleri standart hata ile ters orantılıdır ve standart hata değeri ne kadar düşüğe, testin verdiği bilgi miktarı da o kadar artacaktır (Anıl, 2004). Bu doğrultuda, MPS miktarı, MPS yüzdesi ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça test bilgi fonksiyonu bir başka deyişle testin verdiği bilgi miktarı düşmektedir. Bu eğilim madde bankası büyüklüğü artsa da değişkenlik göstermemektedir. TBF değerlerindeki bu değişim sadece MPS olmadığı için temel düzey veri setinde ortaya çıkmamıştır.

Alanyazında, MPS olduğu durumda TBF değerlerinin düşük düzeyde de olsa değişme eğiliminde olduğunu gösteren çalışmalar mevcuttur (Chan vd., 1999; Deng ve Melican, 2009; Guo ve Wang, 2003; 2005). Deng ve Melican (2009) BOBUT uygulaması olan ACCUPLACER® testinde MPS’yi inceledikleri çalışmalarında yıllar içerisinde bazı yetenek düzeylerinde testin verdiği bilgi miktarının düştüğünü belirtmişlerdir. Bu

durumun nedeni incelendiğinde ise maddelerin sık görülmesi değil de müfredat değişimi veya başka bir sebepten kaynaklı olabileceğini belirtmişlerdir. Çünkü MPS gösteren maddeler incelendiğinde, maddelerin çok sık uygulanmadığını görmüşlerdir. Guo ve Wang (2005) da benzer şekilde MPS'nin test bilgi fonksiyonunu düşük miktarlarda da olsa değiştirdiğini belirtmişlerdir. Bu doğrultuda MPS'li madde bankaları olduğu durumda BOBUT uygulamalarında standart hatanın artmasına bağlı olarak test bilgisinin azaldığı belirtilebilir.

Yukarıda verilen TBF değerleri açısından elde edilen farkların istatistiksel olarak manidar olup olmadığını incelemek için bağımsız örneklemeler için yapılan üç faktörlü ANOVA sonuçları Tablo 14'te verilmiştir.

Tablo 14

Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre TBF Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
200	MPS Büyüklüğü	0.000	3	0.000	0.00*	0.00
	MPS Yüzdesi	0.032	3	0.010	38016.00*	0.45
	Ölçüm	0.002	2	0.001	2376.00*	0.03
	MPS Büyüklüğü*MPS Yüzdesi	0.016	9	0.002	19008.00*	0.22
	MPS Büyüklüğü*Ölçüm	0.007	6	0.001	8316.00	-
	MPS Yüzdesi*Ölçüm	0.003	6	0.000	3564.00	-
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.007	18	0.000	8316.00	-
	Hata	0.004	4752	0.000		
	Toplam	0.071	4799			
500	MPS Büyüklüğü	0.000	3	0.000	0.00*	0.00
	MPS Yüzdesi	0.002	3	0.000	2376.00*	0.01
	Ölçüm	0.000	2	0.000	0.00*	0.00
	MPS Büyüklüğü*MPS Yüzdesi	0.000	9	0.000	0.00*	0.00
	MPS Büyüklüğü*Ölçüm	0.054	6	0.009	64152.00*	0.50
	MPS Yüzdesi*Ölçüm	0.046	6	0.008	54648.00*	0.43
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.000	18	0.000	0.00*	0.00

(Devam ediyor)

Tablo 14 (Devam)

Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre TBF Değerlerinin Karşılaştırılması

Madde Bankası Büyüklüğü	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	Etki Büyüklüğü (η^2)
500	Hata	0.004	4752	0.000		
	Toplam	0.106	4799			
	MPS Büyüklüğü	0.000	3	0.000	0.00*	0.00
	MPS Yüzdesi	0.001	3	0.000	1584.00*	0.01
	Ölçüm	0.022	2	0.011	34848.00*	0.20
1000	MPS Büyüklüğü*MPS Yüzdesi	0.067	9	0.007	106128.00*	0.63
	MPS Büyüklüğü*Ölçüm	0.001	6	0.000	1584.00	-
	MPS Yüzdesi*Ölçüm	0.003	6	0.001	4752.00	-
	MPS Büyüklüğü*MPS Yüzdesi*Ölçüm	0.010	18	0.000	15840.00	-
	Hata	0.003	4752	0.000		
	Toplam	0.106	4799			

*p<.05

Test Bilgi Fonksiyonu değerlerine ilişkin bağımsız örneklemeler için üç faktörlü ANOVA sonuçları genel olarak incelendiğinde 5000 kişilik yanıtlayıcı örnekleme, 500 maddelik madde bankası için MPS büyüklüğü, MPS yüzdesi, MPS’li madde bankası ile yapılan ölçüm sayısının hem temel etki, hem ikili etkileşim hem de üçlü etkileşim etkilerinin TBF üzerinde istatistiksel olarak manidar etkisi olduğu görülse de bu etkiler genel olarak düşük düzeydedir (Cohen, 1988, s.413). 5000 kişi, 200 ve 1000 maddelik madde bankaları için de MPS büyüklüğü, MPS yüzdesi, ölçüm ve MPS büyüklüğü*MPS yüzdesi şeklinde iki faktörün etkileşim etkisinin TBF üzerinde istatistiksel olarak manidar etkisi olduğu görülmektedir (Cohen, 1988, s.413). Yapılan Post-Hoc test sonuçları da anlamlı fark çıkan tüm faktörlerin tüm düzeyleri açısından fark olduğunu ortaya koymuştur. Bu doğrultta BOBUT uygulamalarında MPS’li madde bankasının uzun süre kullanımından kaynaklı, bazı maddelerin kolay yönde MPS gösterdiği durumda özellikle MPS büyüklüğü, MPS yüzdesi, ölçüm ve MPS büyüklüğü*MPS yüzdesi faktörlerinin TBF değerlerini önemli düzeyde etkilediği belirtilebilir. Bir başka deyişle MPS’li madde bankası testin verdiği bilgi miktarını düşürmektedir. Chan ve diğerleri (1999) tarafından ASVAB’ın psikometrik tutarlılığının incelendiği çalışmada da 200 maddeden yaklaşık

%10'unun MPS gösterdiği durumda TBF değerlerinin istatistiksel olarak manidar düzeyde etkilendiği bulunmuştur.

Alanyazında MPS'nin TBF üzerinde istatistiksel olarak manidar etkisinin olmadığını gösteren çalışmalar da mevcuttur (Deng ve Melican, 2009; Guo ve Wang, 2003; 2005). Guo ve Wang (2003) gerçek ve simülatif verilerden iki zaman noktasında ölçüm aldıkları çalışmalarında test karakteristik eğrileri bakımından çok düşük miktarda farklar elde edilmiş olup bu farkların pratik anlamda etkisinin önemli olmadığını belirtilmişlerdir. Ancak çalışma kapsamında hem iki zaman noktası için ölçüm alınmış hem de sınava katılan 169.111 kişi arasından rastgele on tane 1000 kişilik örneklem çekerek, bu on replikasyon üzerinden çalışmışlardır. Çalışmanın sonunda da en az 100 replikasyon üzerinden sonuçların değerlendirilmesi gerektiğini belirtmişlerdir. Benzer şekilde Guo ve Wang (2005) GMAT® sınavında iki zaman noktasında çeşitli simülasyon koşulları ile MPS'yi inceledikleri çalışmalarında da TBF açısından istatistiksel olarak manidar fark elde etmemişlerdir. Ancak MPS'nin etkisinin net bir şekilde gözlemlenebilmesi için ikiden fazla zaman noktasına ihtiyaç vardır (Babcock ve Albano, 2012; Chan vd., 1999). Belirtilen çalışmalarda iki zaman noktası için ölçüm alınmışken, bu çalışma kapsamında da üç zaman noktası için ölçüm alınmıştır. Dolayısıyla belirtilen çalışmalarda manidar fark çıkmamasının nedeni iki noktada ölçüm alınmış olmasından kaynaklanmış olabilir.

Araştırmada ele alınan bağımsız değişken koşullarının karşılaştırıldığı son ölçüt (bağımlı değişken), test etkililiğine ilişkin değerleridir. Buna ilişkin bulgular, Tablo 15'te özetlenmiştir.

Tablo 15

Örneklem n=5000 olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Test Etkililiği Değerleri

Madde Bankası Büyüklüğü	MPS Büyüklüğü	MPS Yüzdesi	Ölçüm	Ortalama Test Uzunluğu	Maddelerin Ortalama Uygulanma Oranı	X ²
200	0.00	%0	1.	23.52	0.2018	0.3115
			2.	23.52	0.2018	0.3115
			3.	23.52	0.2018	0.3115
	0.50	%5	1.	23.52	0.2056	0.3124
			2.	23.54	0.2056	0.3146

(Devam ediyor)

Tablo 15 (Devam)

Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Test Etkililiği Değerleri

Madde Bankası Büyüklüğü	MPS Büyüklüğü	MPS Yüzdesi	Ölçüm	Ortalama Test Uzunluğu	Maddelerin Ortalama Uygulanma Oranı	X ²
200	0.50	%5	3.	23.56	0.2058	0.3162
			1.	23.54	0.2058	0.3172
			2.	23.58	0.2059	0.3187
		%10	3.	23.65	0.2059	0.3190
			1.	23.57	0.2060	0.3192
			2.	23.70	0.2060	0.3192
		%20	3.	23.80	0.2060	0.3193
			1.	23.53	0.2060	0.3236
			2.	23.56	0.2060	0.3246
	0.75	%5	3.	23.60	0.2060	0.3282
			1.	23.56	0.2060	0.5402
			2.	23.65	0.2060	0.5415
		%10	3.	23.69	0.2060	0.5424
			1.	23.63	0.2060	0.5439
			2.	23.80	0.2060	0.5467
		%20	3.	23.81	0.2060	0.5468
			1.	23.54	0.2060	0.5521
			2.	23.59	0.2060	0.5525
	1.00	%5	3.	23.63	0.2060	0.5533
			1.	23.58	0.2060	0.5595
			2.	23.66	0.2060	0.5598
		%10	3.	23.68	0.2061	0.5625
			1.	23.70	0.2061	0.5667
			2.	23.84	0.2062	0.5738
		%20	3.	23.98	0.2064	0.5856
			1.	23.41	0.0825	0.2353
			2.	23.41	0.0825	0.2353
500	0.00	%0	3.	23.41	0.0825	0.2353
			1.	23.43	0.0826	0.2873
			2.	23.43	0.0826	0.2883
	0.50	%5	3.	23.44	0.0826	0.2888
			1.	23.44	0.0827	0.2889
			2.	23.45	0.0827	0.2896
		%10	3.	23.45	0.0827	0.2900
			1.	23.45	0.0827	0.2903
			2.	23.46	0.0826	0.2922
		%20	3.	23.47	0.0826	0.2925
			1.	23.47	0.0827	0.2925
			2.	23.50	0.0827	0.2925

(Devam ediyor)

Tablo 15 (Devam)

Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Test Etkililiği Değerleri

Madde Bankası Büyüklüğü	MPS Büyüklüğü	MPS Yüzdesi	Ölçüm	Ortalama Test Uzunluğu	Maddelerin Ortalama Uygulanma Oranı	X ²
500	0.75	%5	3.	23.50	0.0827	0.2933
			1.	23.51	0.0827	0.2937
			2.	23.51	0.0827	0.2941
		%10	3.	23.52	0.0827	0.2942
			1.	23.53	0.0827	0.3003
			2.	23.53	0.0827	0.3016
		%20	3.	23.55	0.0827	0.3082
			1.	23.55	0.0827	0.3082
			2.	23.57	0.0827	0.3108
	1.00	%5	3.	23.57	0.0827	0.3108
			1.	23.58	0.0827	0.3260
			2.	23.70	0.0827	0.3260
		%10	3.	23.71	0.0827	0.3290
			1.	23.71	0.0827	0.3290
			2.	23.81	0.0827	0.3303
		%20	3.	23.83	0.0827	0.3303
			1.	23.40	0.0405	0.7318
			2.	23.40	0.0405	0.7318
1000	0.50	%0	3.	23.40	0.0405	0.7318
			1.	23.41	0.0412	0.7429
			2.	23.41	0.0413	0.7429
		%5	3.	23.42	0.0413	0.7442
			1.	23.43	0.0413	0.7460
			2.	23.44	0.0413	0.7463
		%10	3.	23.44	0.0413	0.7463
			1.	23.45	0.0413	0.7466
			2.	23.45	0.0413	0.7468
	0.75	%20	3.	23.45	0.0413	0.7468
			1.	23.46	0.0413	0.7469
			2.	23.46	0.0413	0.7473
		%5	3.	23.47	0.0413	0.7474
			1.	23.47	0.0413	0.7474
			2.	23.48	0.0414	0.7474
		%10	3.	23.48	0.0414	0.7474
			1.	23.51	0.0414	0.7475
			2.	23.51	0.0414	0.7476
	1.00	%20	3.	23.52	0.0414	0.7485
			1.	23.53	0.0414	0.7485
			2.	23.54	0.0414	0.7486
		%5	3.	23.54	0.0414	0.7486

(Devam ediyor)

Tablo 15 (Devam)

Örneklem n=5000 Olduğunda, Farklı Büyüklükteki Madde Bankaları için Farklı MPS Büyüklüğü ve Farklı MPS Yüzde Değerlerinde, Üç Farklı Zamana göre Test Etkililiği Değerleri

Madde Bankası Büyüklüğü	MPS Büyüklüğü	MPS Yüzdesi	Ölçüm	Ortalama Test Uzunluğu	Maddelerin Ortalama Uygulanma Oranı	X ²
1000	1.00	%10	1.	23.55	0.0414	0.7489
			2.	23.65	0.0414	0.7490
			3.	23.65	0.0414	0.7499
		%20	1.	23.75	0.0414	0.7507
			2.	23.76	0.0414	0.7538
			3.	23.76	0.0414	0.7549

Tablo 15 incelendiğinde; 5000 kişilik yanıtlayıcı örnekleminde 200, 500 ve 1000 maddelik madde bankaları için MPS miktarı büyüdükçe (0.00, 0.50, 0.75, 1.00) ve MPS yüzdesi arttıkça (%0, %5, %10, %20), üç farklı zaman noktasında elde edilen yetenek kestirimlerine ait *ortalama test uzunluğu*, *maddelerin ortalama uygulanma oranları* ve *X²* değerlerinin çok düşük miktarlarda da olsa artma eğiliminde olduğu görülmektedir. *X²* ve ortalama test uzunluğuna ilişkin eğilim madde bankası büyüklüğü arttıkça değişmemiştir. Ancak madde bankası büyüklüğü arttıkça maddelerin ortalama uygulanma oranı azalmıştır. Test etkililiği değerlerindeki bu değişim sadece MPS olmadığı için temel düzey veri seti için ortaya çıkmamıştır. Bu doğrultuda test etkililiğine ilişkin olarak MPS miktarı, MPS büyüklüğü ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça daha az etkili testler elde edildiği belirtilebilir. Benzer bulguya alanyazında yapılan bazı çalışmalarda da rastlanmaktadır (Bergstrom ve Stahl, 1992; Lunz ve Stahl; 1993).

Risk (2015) tarafından yapılan çalışmada, bu çalışmada elde edilen bulguya zıt olarak MPS’li koşullar altında daha etkili testler elde edildiği görülmüştür. Simülasyon çalışması olarak yürütülen bu çalışmada, bu durumun nedeni olarak da maddelerin, örnekleme uygun olmaması ihtimali üzerinde durulmuştur. BOBUT’ta önemli olan faktörlerden biri maddelerin görülme sıklığıdır. Maddelerin görülme sıklığının düşük olması için minimum madde bankası büyüklüğünün 200-300 civarında olması gerektiği belirtilmektedir (Lunz ve Stahl; 1993). Bu çalışma kapsamında da test etkililiği açısından düşük düzeyde farklılıklar elde edilmesinin nedeni minimum 200 maddelik madde bankası ile çalışılmış olması olabilir.

Elde edilen bulgular ölçme kesinliği açısından genel olarak değerlendirildiğinde beklenen yönde sonuçlar elde edilmiştir. Bir diğer ifadeyle MPS büyüklüğü, MPS yüzdesi ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça ölçme kesinliği azalmıştır çünkü madde bankasında MPS’li maddeler olduğu için yetenek kestirimlerinden sapmalar meydana gelmiştir. Madde bankasına ilişkin MPS koşulları arttıkça yanlılık değerleri negatif yönde, OHKK ve MOF değerleri de pozitif yönde artmıştır. Temel düzey veri seti ile karşılaştırıldığında yanlılık, OHKK ve MOF açısından en yüksek değerler ölçümün üçüncü noktasında, MPS büyüklüğünün 1.00 ve MPS’li madde yüzdesinin %20 olduğu durumda elde edilmiştir. Bir başka ifade ile MPS’li koşulların en yüksek olduğu durumda, ölçme kesinliği en düşük değerini almıştır. Yapılan bağımsız örneklem için üç faktörlü ANOVA analizi de ölçme kesinliğine ilişkin bu faktörler açısından çoğunlukla istatistiksel olarak manidar sonuçlar elde edildiğini ve ölçme kesinliğini en çok etkileyen faktörün madde bankasındaki MPS’li madde sayısı olduğunu göstermiştir.

Alanyazında MPS’nin BOBUT uygulamalarında ölçme kesinliği üzerindeki etkisini araştıran araştırma bulgularıyla da tutarlı olarak (Abad vd., 2010; Aksu Dünya, 2017; Babcock ve Albano, 2012; Chan vd., 1999; McCoy, 2009; Risk, 2015; Wells vd., 2002) MPS büyüklüğünün artması ölçme kesinliğini azaltmıştır. MPS koşullarının değişmesi ölçme kesinliğini etkilerken, örneklem büyüklüğünün artması durumunda yanlılık, OHKK ve MOF değerlerinde artan veya azalan yönde bir örüntü gözlenmemiş ancak MPS koşullarına ilişkin eğilim, bu faktörleri her iki örneklem büyüklüğünde de benzer şekilde etkilemiştir. Madde bankası büyüklüğü için de, madde bankası büyüklüğü arttıkça yanlılık değerleri açısından herhangi bir eğilim gözlenmezken 1000 kişilik örnekleme OHKK’nın temel düzey veri seti açısından, 5000 kişilik örnekleme de MOF ve OHKK’nın hem temel düzey hem de MPS’li veri setleri açısından madde bankası büyüklüğü arttıkça OHKK ve MOF değerlerinin azaldığı görülmüştür. Bu durumun nedeni olarak da madde bankası büyüklüğü arttıkça MPS koşullarının, yetenek kestirimlerine ilişkin hata ve ortalama yanlılık değerlerini daha az etkilediği şeklinde düşünülmüştür.

Çalışma kapsamında ölçme kesinliğini en fazla etkileyen faktörün MPS yüzdesi olduğu görülmüştür. Alanyazında ölçme kesinliği üzerinde en fazla etkisi olan faktörün MPS yüzdesi olduğunu belirten çalışmalar olmakla birlikte (Babcock ve Albano, 2012), MPS büyüklüğü (Risk, 2015); örneklem büyüklüğü ve MPS yüzdesinin birlikte (Wells vd., 2002) etkisinin olduğunu belirten çalışmalar mevcuttur. Örneğin, Risk (2015) çalışmasında Rasch model kullanarak farklı MPS koşullarının ölçme kesinliği üzerindeki

etkisini incelemiş ve ölçme kesinliğini en fazla etkileyen faktörün madde bankasındaki MPS'li madde sayısından çok MPS büyüklüğü olduğunu ancak bu etkinin manidar olmadığını belirtmiştir. Benzer şekilde Wells ve diğerleri (2002) de 2PLM ile MPS'nin yetenek kestirimlerine etkisini inceledikleri çalışmalarında, yetenek kestirimlerini en fazla etkileyen faktörün örneklem büyüklüğü ve MPS yüzdesi olduğunu belirtmişlerdir. Ancak, belirtilen çalışmalarda simüle edilen örneklem büyüklüğü, madde bankası büyüklüğü, MPS koşulları ve kullanılan MTK modeli değişkenlik göstermektedir. Bu doğrultuda, BOBUT uygulamalarında madde bankasında MPS'li madde bulunması ölçme kesinliğini olumsuz yönde etkilerken hangi faktörün bu değeri olumsuz yönde etkilediği, kullanılan MTK modeli, örneklem büyüklüğü, madde bankası büyüklüğü veya MPS koşuluna göre değişkenlik gösterebilir.

Çalışmada ölçme kesinliğinin yanlışlık faktörü açısından beklenenin aksi yönde bir bulgu elde edilmiştir. Yanlışlık değerleri MPS koşulları arttıkça negatif yönde artmıştır. Ancak, madde bankasındaki bir kısım maddenin kolay yönde MPS göstermesinden dolayı, bireylerin kestirilen yetenek değerlerinin, gerçek yeteneklerinden yüksek olması bir başka deyişle yanlışlık değerlerinin pozitif yönde artmış olması beklenirdi. Benzer bulguya Aksu Dünya (2017), Guo ve Wang (2003) ve Risk (2015) tarafından yapılan çalışmada da rastlanmaktadır. Bu bulgunun bir nedeni BOBUT uygulamasında düşük veya yüksek yetenek düzeyindeki hangi bireylerin hangi maddelerle karşılaştığı ile ilgili bir durum olabilir. Örneğin, yeteneklerine uygun maddelerle karşılaşmasına rağmen bunları yanlış yanıtlamışsa, kestirilen yetenek değerleri düşük çıkacaktır. Bu durumun bir diğer nedeni de BOBUT kapsamında ele alınan madde seçim yöntemi olabilir. Bu çalışma kapsamında BOBUT koşullarına ilişkin madde seçim yöntemi olarak Kullback-Leibler yöntemi seçilmiştir. Seçilen bu yöntem de bireylerin karşılaşılabilecekleri maddeleri etkileyen bir faktör olabilmektedir. Dolayısıyla madde seçim yöntemi de yanlışlık değerlerinin negatif yönde artmasının bir diğer nedeni olabilir.

Test bilgi fonksiyonu açısından bulgular değerlendirildiğinde, BOBUT uygulamalarında madde bankasında yer alan MPS'li madde sayısı, MPS büyüklüğü ve MPS'li madde bankası ile yapılan ölçüm sayısı arttıkça testin verdiği bilgi miktarının nispeten azaldığı görülmüştür. Tüm koşullar için en yüksek TBF değeri, MPS'nin olmadığı temel düzey veri seti için, en düşük TBF değerleri de ölçümün üçüncü noktasında, MPS büyüklüğünün 1.00 ve MPS'li madde yüzdesinin %20 olduğu durumda elde edilmiştir. Ölçüm sayısı ve MPS'ye ilişkin koşullar arttıkça, testin verdiği bilgi

miktarı azalmış ancak örneklem büyüklüğü ve madde bankası büyüklüğü değiştikçe TBF değerlerinde artan veya azalan yönde bir eğilim gözlenmemiştir.

Örneklem büyüklüğü ve madde bankası büyüklüğü TBF'yi etkilemezken, ölçüm sayısı ve MPS koşullarının TBF'yi etkilediği belirtilebilir. Elde edilen TBF değerlerinin istatistiksel manidarlığı incelendiğinde de genelde faktörlerin temel düzey ve MPS büyüklüğü*MPS yüzdesi faktörleri için istatistiksel olarak manidar sonuçlar elde edilmiştir. Alanyazında MPS'nin TBF üzerindeki etkisinin incelendiği çalışmalarda, bu bulguyu destekleyenler olmakla birlikte (Chan vd., 1999), istatistiksel olarak manidar fark elde edilmeyen çalışmalar da mevcuttur (Deng ve Melican, 2009; Guo ve Wang, 2002). Chan ve diğerleri (1999) tarafından yapılan ASVAB'ın psikometrik tutarlılığının incelendiği çalışmada beş zaman noktasında inceleme yapılmış ve test karakteristik eğrileri karşılaştırılmıştır. Karşılaştırma sonucu karakteristik eğrileri arasındaki farkın çok düşük miktarda ve istatistiksel olarak manidar olduğu, ancak pratik anlamda bu farkların düşük düzeyde etkisinin olduğu belirtilmiştir. Guo ve Wang (2003) tarafından yapılan, gerçek ve simülatif verilerin kullanılarak ölçek sapmasının BOBUT'a etkisinin araştırıldığı çalışmada da test karakteristik eğrileri karşılaştırılmıştır. Ancak iki zaman noktasında ölçüm alınmasından dolayı TBF açısından çok düşük farklar elde edildiği ve bu farkların manidar olmadığı belirtilmiştir.

Test etkililiğine ilişkin değerler genel olarak değerlendirildiğinde BOBUT uygulamalarında MPS büyüklüğü, MPS yüzdesi ve MPS'li madde bankası ile yapılan ölçüm sayısı arttıkça ortalama test uzunluğu, maddelerin ortalama uygulanma oranı ve X^2 değerlerinin çok küçük miktarlarda arttığı görülmektedir. Tüm koşullar için ortalama test uzunluğu optimum 23.00 olarak elde edilmiştir. Tüm koşullar için maddelerin ortalama uygulanma oranı incelendiğinde ise örneklem büyüklüğünün arttığı durumda herhangi bir örüntü elde edilmezken, madde bankası büyüklüğü arttıkça maddelerin ortalama uygulanma oranlarının düştüğü görülmektedir. X^2 değerlerinde ise örneklem büyüklüğü ve madde bankası büyüklüğü arttıkça herhangi bir artış ya da azalış gözlemlenmese de MPS koşulları arttıkça gözlenen ve beklenen madde kullanım sıklığı arasındaki farkın arttığı görülmektedir. Bu doğrultuda, MPS yüzdesi, MPS büyüklüğü ve yapılan ölçüm sayısı arttıkça, madde bankasındaki maddelerin daha az etkili bir şekilde kullanıldığı bir diğer ifadeyle maddelerin dengeli bir şekilde kullanılmadığı belirtilebilir.

Bu çalışma kapsamındaki koşullar altında MPS miktarı, MPS büyüklüğü ve MPS'li madde bankası ile yapılan ölçüm sayısı arttıkça daha az etkili testler elde edilse de aradaki farklar çok düşük düzeydedir. MPS koşulları arttıkça ölçme kesinliği ve test

bilgisinin azalmasına paralel olarak daha az etkili testler elde edilmiştir. Benzer bulguya alanyazında yapılan bazı çalışmalarda da rastlanmaktadır (Bergstrom ve Stahl, 1992; Lunz ve Stahl; 1993). BOBUT'ta MPS'yi etkileyen önemli faktörlerden biri maddelerin görülme sıklığıdır. Bu doğrultuda etkili testler elde etmede minimum madde bankası büyüklüğünün 200-300 olması gerektiği belirtilmektedir (Lunz ve Stahl; 1993). Bu çalışma kapsamında da test etkililiği değerleri açısından çok düşük düzeyde farklılıklar elde edilmesinin nedeni minimum 200 maddelik madde bankası ile çalışması olabilir. Risk'in (2015) çalışmasında ise bu bulgunun aksi yönde, MPS'li maddelerin olduğu koşullarda daha etkili testler elde edildiği görülmüştür. Beklenenin aksi yönde sonuç elde edilmesinin nedeni olarak da simüle edilen koşullardan madde bankasındaki maddelerin, örnekleme uygun olmaması olarak belirtilmiştir.

Özetle, bu araştırma kapsamında çalışılan koşullar altında ölçme kesinliği, TBF ve test etkililiği değerlerinin MPS'den olumsuz yönde etkilendiği görülmüştür. Babcock ve Albano (2012), Rasch model kullanıldığı durumda BOBUT'un, MPS'nin etkisine bir miktar dayanıklı olduğunu belirtse de MPS'li madde bankası ile yapılan ölçüm sayısı arttıkça madde ve birey parametrelerine ilişkin kestirimlerden sapmalar gözlemlenmiştir. BOBUT uygulamalarında Rasch modelin MPS'ye dayanıklı olduğu belirtilse de bu çalışmanın bulgusuna da paralel olarak, en çok MPS yüzdesinin yetenek kestirimlerini ve Rasch ölçeğinin sürekliliğini etkilediğini belirtilmektedir (Babcock ve Albano, 2012; Wells vd., 2002).

Bu noktada, kullanılan MTK modeli ve BOBUT uygulamalarının avantajlarından önce, madde bankasına ilişkin test geliştirme ve gözden geçirme süreçlerinin önemini bir kez daha vurgulamak gerekir. Çünkü madde bankasında kaliteli maddeler olduğu doğrultuda kullanılan modeller veya tercih edilen uygulamalar (BOBUT, kâğıt-kalem gibi) doğru sonuçlar verecektir. Test geliştirme ve gözden geçirme süreci dinamik bir süreçtir ve özellikle önemli kararların alındığı uygulamalarda -MPS'nin yetenek kestirimlerini etkilediğini göz önünde bulundurarak- madde bankalarına ilişkin sürekli kontroller ve güncellemeler yapılması gerekmektedir.

BÖLÜM 4

SONUÇLAR VE ÖNERİLER

Bu bölümde araştırmadan elde edilen bulgulara yönelik sonuçlara ve önerilere yer verilmiştir.

Sonuçlar

Örneklem büyüklüğünün 1000 ve 5000 olduğu BOBUT uygulamalarında;

1. Madde bankası büyüklüğünün 200, 500 ve 1000 olduğu durumda, madde bankasına ilişkin MPS büyüklüğü, MPS yüzdesi arttıkça ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça yanlışlık değerleri negatif yönde; OHKK ve MOF değerleri de pozitif yönde artmaktadır. Bu doğrultuda madde bankasında MPS’li maddeler olduğu durumda daha yanlış ve daha hatalı sonuçlar elde edilmektedir. Bir başka ifade ile madde bankasında MPS’li madde olması durumunda ölçme kesinliği düşmektedir.
2. Madde bankası büyüklüğünün 200, 500 ve 1000 olduğu durumda ölçme kesinliğine ilişkin olarak en iyi sonuçlar MPS’nin olmadığı temel düzey veri seti için elde edilmişken, geçerlik ve güvenilirlik açısından istenmeyen değerler MPS büyüklüğünün 1.00 logit, MPS yüzdesinin %20 olduğu ve üçüncü ölçümün alındığı noktada elde edilmiştir. Bir başka deyişle MPS’li maddelerin olmadığı durumda ölçme kesinliği en yüksek iken, madde havuzundaki MPS’li madde sayısı, MPS büyüklüğü arttıkça ve MPS’li madde bankası uzun süre kullanıldıkça ölçme kesinliği azalmaktadır.
3. Madde bankası büyüklüğünün 200, 500 ve 1000 olduğu durumda ölçme kesinliğini en çok etkileyen faktör, madde bankasındaki MPS’li madde sayısıdır. Bir başka ifade ile MPS büyüklüğü ve yapılan ölçüm sayısından çok madde bankasındaki MPS’li madde sayısının fazla olması ölçme kesinliğini düşürmektedir.

4. Madde bankası büyüklüğünün 200, 500 ve 1000 olduğu durumda, madde bankasına ilişkin MPS büyüklüğü, MPS yüzdesi ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça, ölçmenin standart hatası arttığından testin verdiği bilgi miktarı azalmaktadır. Bir başka deyişle madde bankasında MPS’li maddelerin olması, test bilgi fonksiyonunu düşürmektedir.
5. Madde bankası büyüklüğünün 200, 500 ve 1000 olduğu durumda en yüksek test bilgi değeri MPS’nin olmadığı temel düzey veri seti için elde edilmişken, en düşük test bilgi değeri MPS büyüklüğünün 1.00 logit, MPS yüzdesinin %20 olduğu ve üçüncü ölçümün alındığı noktada elde edilmiştir. Bir başka deyişle MPS’li maddelerin olmadığı durumda test bilgisi en yüksek iken, madde havuzundaki MPS’li madde sayısı, MPS büyüklüğü arttıkça ve MPS’li madde bankası uzun süre kullanıldıkça testin verdiği bilgi miktarı azalmaktadır.
6. Madde bankası büyüklüğünün 200, 500 ve 1000 olduğu durumda TBF değerini en çok etkileyen faktör, madde bankasındaki MPS’li madde sayısıdır. Bir başka ifade ile MPS büyüklüğü ve yapılan ölçüm sayısından çok madde bankasındaki MPS’li madde sayısının fazla olması test bilgisini düşürmektedir.
7. Madde bankası büyüklüğünün 200, 500 ve 1000 olduğu durumda, madde bankasına ilişkin MPS büyüklüğü, MPS yüzdesi ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça ortalama test uzunluğu ve gözlenen ve beklenen madde kullanım sıklığı değerleri artmaktadır. Bu doğrultuda test etkililiği azalmaktadır. Bir başka deyişle madde bankasında MPS’li madde olması durumunda daha az etkili testler elde edilmektedir.
8. 1000 ve 5000 kişilik yanıtlayıcı örnekleminde madde bankası büyüklüğünün 200, 500 ve 1000 olduğu durumda, hem temel düzey veri seti hem de MPS’li veri setleri için test etkililiğinin göstergesi olarak, test uzunluğu açısından uygulanan optimum madde sayısı 23’tür.
9. Madde bankasına ilişkin MPS büyüklüğü, MPS yüzdesi ve MPS’li madde bankası ile yapılan ölçüm sayısı arttıkça maddelerin ortalama uygulanma oranı artmakta, madde bankası büyüklüğü arttıkça da, maddelerin ortalama uygulanma oranı azalmaktadır. Bu doğrultuda aynı madde bankası içerisinde MPS koşulları arttıkça test etkililiği

azalsa da, madde bankası büyüdükçe maddelerin MPS gösterme olasılığı azalmakta dolayısıyla da MPS'nin maddelerin ortalama uygulanma oranı üzerindeki etkisi azalmaktadır.

10. Madde bankası büyüklüğünün 200, 500 ve 1000 olduğu durumda en etkili testler MPS'nin olmadığı temel düzey veri seti için elde edilmişken, MPS büyüklüğünün 1.00 logit, MPS yüzdesinin %20 olduğu ve üçüncü ölçümün alındığı noktada test etkililiği azalmıştır. Bir başka deyişle MPS'li maddelerin olmadığı durumda test etkililiği en yüksek iken, madde havuzundaki MPS'li madde sayısı, MPS büyüklüğü arttıkça ve MPS'li madde bankası uzun süre kullanıldıkça, maddelerin kullanımı açısından daha az etkili testler elde edilmektedir.

Öneriler

Araştırmadan elde edilen bulgular ve sonuçlar doğrultusunda uygulayıcılar ve araştırmacılara yönelik öneriler aşağıda sunulmuştur.

Uygulayıcılara Yönelik Öneriler

1. Çalışma kapsamında BOBUT uygulamalarında çalışılan tüm MPS koşullarının yetenek kestirimlerinin yanlılığını artırdığı, kestirimlerin güvenilirliğini düşürdüğü bulunmuştur. Ancak bu koşullar içerisinde yetenek kestirimlerini en çok etkileyen faktörün madde bankasında yer alan MPS'li madde sayısı olduğu belirlenmiştir. MPS'li madde sayısı arttıkça, hem ölçme kesinliği hem testin verdiği bilgi miktarı hem de test etkililiği azalmıştır. Bu nedenle özellikle önemli kararların alındığı test uygulamaları yapan kuruluşlar, madde bankalarına ilişkin madde bankası büyüklüğü, test uygulamalarında bir maddenin kaç kere kullanıldığı, birden çok kullanılan maddelerin madde parametrelerinde değişim olup olmadığı gibi durumlara ilişkin kontrolleri belirli aralıklarla yaparak, MPS'li madde sayısının artmasının önüne geçmeleri önerilebilir.
2. Bir madde havuzunda belirli maddelerde MPS olması, ölçülmek istenen davranışa farklı durumların karışması anlamına gelir. Bu durum ölçmelere hata karışmasına neden olarak güvenilirliği ve geçerliği düşürür. Dolayısıyla MPS'nin varlığı ölçme

kesinliđi, test bilgi fonksiyonu ve test etkililiđini olumsuz ynde etkileyen bir unsurdur. Bu nedenle test uygulayan kuruluřlara, belirli aralıklarla madde bankasındanki maddeleri MPS'nin bu olumsuz etkilerine karřı yenilemeleri nerilebilir.

Arařtırmacılara Ynelik neriler

1. Bu arařtırmada MPS'nin lme kesinliđi, test bilgisi ve test etkililiđi zerindeki etkisi simlatif veriler zerinden incelenmiřtir. MPS'nin bu faktrler zerindeki etkisi nemli kararlara dayanarak alınan ve yıllar iinde aynı/benzer soruların sorulduđu geniř lekli test uygulamaları iin arařtırılabilir.
2. Bu arařtırmadan elde edilen bulgular 1000 ve 5000 kiřilik rneklemler ile sınırlıdır. alıřma kapsamında 1000 kiřilik rneklem byklđu dřk rneklem byklđu olarak ele alınsa da bazı lisans ve/veya sertifika verilen programlar iin 500 kiřilik rneklem byklđu de ideal olarak ele alınmaktadır (Kim, Barton ve Choi, 2010). Dolayısıyla, MPS'nin etkileri 1000'den kk rneklem byklđu iin arařtırılabilir.
3. Bu alıřma kapsamında 1000 ve 5000 kiřilik rneklem byklkleri normal dađılıma gre simle edilmiřtir. Ancak sertifika verilen sınavlar gibi bazı sınavlarda ve eđitim alanında, normal dađılıma uymayan veya u deđerlerin olduđu kořullarla da sıklıkla karřılařılmaktadır. Dolayısıyla, arpık dađılım olduđu kořulda MPS'nin lme kesinliđi, test bilgisi ve test etkililiđi zerindeki etkisi arařtırılabilir.
4. Bu arařtırma Rasch model ile sınırlıdır. 1PLM, 2PLM, 3PLM vb. gibi farklı MTK modelleri kullanılarak MPS'nin lme kesinliđi, test bilgisi ve test etkililiđi zerindeki etkisi arařtırılabilir.
5. Arařtırmada MPS byklđu olarak 0.50, 0.75, 1.00; MPS gsteren madde yzdesi %5, %10, %20 ve MPS'nin kolay ynde deđiřim gsterdiđi durum ele alınmıřtır. Farklı byklk ve yzdelerde, farklı ynlerde (zor veya her iki ynde) MPS'nin olduđu durumların lme kesinliđi, test bilgisi ve test etkililiđi zerindeki etkisi arařtırılabilir.

6. BOBUT bağlamında yetenek kestirim yöntemi olarak beklenen sonsal dağılım, başlama kuralı olarak ön testten aldıkları puanlara göre önsel θ dağılımları, madde seçim yöntemi olarak Kullbak-Leibler ve durdurma kuralı olarak da en düşük sayıda madde ve standart hatanın 0.40'ın altına düştüğü durum ele alınmıştır. Bu bağlamda BOBUT uygulamalarında farklı yetenek kestirim yöntemi, başlama kuralı, madde seçim yöntemi ve durdurma kuralı seçilerek MPS'nin etkileri incelenebilir.
7. Bu çalışma kapsamında maddelerin görülme sıklığına (item exposure) ilişkin herhangi bir sınırlama getirilmemiştir. Maddelerin görülme sıklığına ilişkin olarak farklı yöntemler kullanılarak ve çeşitli sınırlamalar getirilerek MPS'li madde havuzu olması durumunda, yetenek kestirimlerine etkisi incelenebilir.
8. Bu çalışma kapsamında tüm bireylerin MPS'li maddelerle karşılaşabileceği durum ele alınmıştır. Ancak gerçek uygulamalarda müfredat değişikliği veya pratik yapmadan kaynaklı olarak belirli bir grup bireyin aldığı maddelerde MPS görülebilmektedir. Bu doğrultuda, bazı bireylerin aldığı maddelerde MPS varken, diğer bireylerin aldıkları maddelerde MPS'nin olmadığı durumun yetenek kestirimlerine etkisi incelenebilir.

KAYNAKLAR

- Abad, F. J., Olea, J., Aguado, D., Ponsoda, V., & Barrada, J. R. (2010). Deterioro de parámetros de los ítems en tests adaptativos informatizados: estudio con eCAT [Item parameter drift in computerized adaptive testing: Study with eCAT]. *Psicothema*, 22, 340-7.
- Aksu Dünya (2017). *Item parameter drift in computer adaptive testing due to lack of content knowledge within sub-populations* (Unpublished doctoral dissertation). University of Illinois, Chicago.
- American Educational Research Association, American Psychological Association, ve National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anıl, D. (2005). Test Güvenirliği ve Test Bilgi Fonksiyonunun Klasik ve Örtük Özellikler Test Teorilerine Göre Kestirilmesi. *Eğitim ve Bilim*, 30 (135), 23-33.
- Babcock, B., & Albano, A. D. (2012). Rasch scale stability in the presence of item parameter and trait drift. *Applied Psychological Measurement*, 36(7), 565-580.
- Babcock, B., & Weiss, D.J. (2012). Termination criteria in computerized adaptive tests: do variable-length CAT's provide efficient and effective measurement? *International Association for Computerized Adaptive Testing*, 1, 1-18.
- Barrada, J. R., Olea, J., Ponsoda, V., & Abad, F. J. (2010). A method for the comparison of item selection rules in computerized adaptive testing. *Applied Psychological Measurement*, 34, 438-452.
- Bergstrom, B.A., & Lunz, M.E. (1994). Equivalence of rasch item calibrations across modes of administration. M.Wilson (Ed.). *Objective theory into practice*, Norwood, NJ: Ablex.
- Bergstrom, B. A., Stahl, J., & Netzky, B. A. (2001, April). *Factors that influence parameter drift*. Paper presented at American Educational Research Association. Seattle, WA.
- Blais, J. & Raiche, G. (2002, April). Features of the sampling distribution of the ability estimate in computerized adaptive testing according to two stopping rules, *International Objective Measurement Workshop*, New Orleans.
- Bock, D. B., Muraki, E., & Pfeifferberger, W. (1988). Item pool maintenance in the presence of item parameter drift. *Journal of Educational Measurement*, 25(4), 275-285.

- Burton, A., Altman, D.G., Royston, P., & Holder, R.L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25, 4279-4292.
- Chan, K. Y., Drasgow, F., & Sawin, L. L. (1999). What is the shelf life of a test? The effect of time on the psychometrics of a cognitive ability test battery. *Journal of Applied Psychology*, 84(4), 610-619.
- Chang, H., & Ying, Z. (1999). A-stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, 23, 211-222.
- Chang, H.-H., Qian, J., & Ying, Z. (2001). a-stratified multistage computerized adaptive testing with b blocking. *Applied Psychological Measurement*, 25(4), 333-341. doi: 10.1177/01466210122032181.
- Chang, S.-W., & Ansley, T. N. (2003). A comparative study of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 40, 71-103.
- Chen, S.Y., Ankenmann, R.D., & Chang, H.H. (2000). A comparison of item selection rules at the early stages of computerized adaptive testing. *Applied Psychological Measurement*, 24, 241-255.
- Chen, Q. (2013). *Remove or keep: linking items showing item parameter drift*. (Unpublished Doctoral Dissertation). Michigan State University.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: Lawrence Erlbaum.
- Çıkrıkçı-Demirtaşlı, N. (1999). Psikometride Yeni Ufuklar: Bilgisayar Ortamında Bireye Uyarlanmış Test. *Türk Psikoloji Bülteni*, 5(13), 31-36.
- DeMars, C. E. (2004). Detection of item parameter drift over multiple test administrations. *Applied Measurement in Education*, 17(3), 265-300.
- Deng, H., Ansley, T., & Chang, H. (2010). Stratified and maximum information item selection procedures in computer adaptive testing. *Journal of Educational Measurement*, 47(2), 202-226.
- Deng, H., & Melican, G. (2010, April). *An investigation of scale drift in computer adaptive test*. Paper presented at Annual Meeting of National Council on Measurement in Education. San Diego, CA.
- Demirus, K. ve Uysal, İ. (2016, Eylül). *Madde Parametre Kaymasının Test Eşitleme Üzerindeki Etkisi*. 5. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sunulmuş bildiri, Akdeniz Üniversitesi, Antalya.
- Donoghue, J. R., & Isham, S. P. (1998). A comparison of procedures to detect item parameter drift. *Applied Psychological Measurement*, 22(1), 33-51.

- Eggen, T. J. H. M. (1999). Item selection in adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement*, 23(3), 249–61.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Eroğlu, M.G. (2013). *Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Farklı Sonlandırma Kurallarının Ölçme Kesinliği ve Test Uzunluğu Açısından Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Filho, N. H., Machado, W.L., & Damasio, B. F. (2014). Effects of statistical models and items difficulties on making trait-level inferences: A simulation study. *Psicologia: Reflexão e Crítica*, 27(4).
- Flaugher, R. (2000). Item Pools. Wainer, H. (Ed.) *Computerized adaptive testing: A Primer*. Mahwah, NJ: Erlbaum.
- Gershon, R. C. (2005). Computer adaptive testing. *Journal of Applied Measurement*, 6(1), 109–127.
- Giordano, C., Subhiyah, R., & Hess, B. (2005, April). An analysis of item exposure and item parameter drift on a take-home recertification exam. *Paper presented at the Annual Meeting of the American Educational Research Association*, Montreal, Canada.
- Goldstein, H. (1983). Measuring changes in educational attainment over time: Problems and possibilities. *Journal of Educational Measurement*, 20(4), 369–377.
- Guo, J., Tay, L., & Drasgow, F. (2009). Conspiracies and test compromise: An evaluation of the resistance of test systems to small-scale cheating. *International Journal of Testing*, 9(4), 283–309.
- Guo, F., & Wang, L. (2003, April). *Online calibration and scale stability of a CAT program*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Guyer, R., & Thompson, N.A. (2011). *Item response theory parameter recovery using Xcalibre 4.1*. St. Paul MN: Assessment Systems Corporation.
- Hagge, S., Woo, A., & Dickison, P. (2011, October). *Impact of item drift on candidate ability estimation*. Paper presented at the Annual Conference of the International Association for Computerized Adaptive Testing, Pacific Grove, CA.
- Han, K. T., & Guo, F. (2011). *Potential impact of item parameter drift due to practice and curriculum change on item calibration in computerized adaptive testing* (R-11-02). Graduate Management Admission Council Research Report.
- Hatfield, J.P., & Nhouyvanisvong, A. (2005, April). *Parameter drift in a high-stakes computer adaptive licensure examination: An analysis of anchor items*. Paper

presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Huang, C., & Shyu, C. (2003, April). *The impact of item parameter drift on equating*. Paper presented at National Council on Measurement in Education. Chicago, IL.

Jiang, G., Tay, L., & Drasgow, F. (2009). Conspiracies and test compromise: An evaluation of the resistance of test systems to small-scale cheating, *International Journal of Testing*, 9 (4), 283-309.

Jodoin, M.G. (2003). *Psychometric properties of several computer-based test designs with ideal and constrained item pools* (Unpublished doctoral dissertation). University of Alberta, Canada.

Jones, P. E., & Smith, R. W. (2006, April) *Item parameter drift in certification exams and its impact on pass-fail decision making*. Paper presented at National Council of Measurement in Education. San Francisco, CA.

Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability* (Unpublished Doctoral Dissertation). Middle East Technical University, Ankara.

Kaptan, F. (1993). *Yetenek Kestiriminde Adaptive (Bireyselleştirilmiş) Test Uygulaması ile Geleneksel Kağıt-Kalem Testi Uygulamasının Karşılaştırılması* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.

Karasar, N. (2009). *Bilimsel Araştırma Yöntemi*. Pegem Akademi, Ankara

Keller, A.L. (2000). *Ability estimation procedures in computerized adaptive testing*. Technical Report, American Institute of Certified Public Accountants-AICPA Research Consortium-Examination Teams, May.

Kezer, F. (2013). *Bilgisayar Ortamında Bireye Uyarlanmış Test Stratejilerinin Karşılaştırılması* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara

Kim, D., Barton, K., & Choi, S. (2010, May). *Sample size impact on screening methods in the Rasch model*. Paper presented at the Annual Meeting of the American Educational Research Association, Denver, CO.

Kim, S. H., & Cohen, A. S. (1992). Effects of linking methods on detection of DIF. *Journal of Educational Measurement*, 29, 51–66.

Kingsbury, G. G., & Wise, S. L. (2011). Creating a K-12 adaptive test: Examining the stability of item parameter estimates and measurement scales. *Journal of Applied Testing Technology*, 12.

Kingsbury, G. G., & Zara, A. R. (2009). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, 2(4), 359-375.

- Kolen, M., & Brennan, R. (1995). *Test equating methods and practices*. New York: Springer-Verlag.
- HyeSun Lee & Kurt F. Geisinger (2019). Item parameter drift in context questionnaires from international large-scale assessments, *International Journal of Testing*, 19(1), 23-51, Doi: 10.1080/15305058.2018.1481852
- Lee, H., & Dodd, B. G. (2012). Comparison of exposure controls, item pool characteristics, and population distributions for CAT using the partial credit model. *Educational and Psychological Measurement*, 72(1), 159-175.
- Li, X. (2008). An investigation of the item parameter drift in the examination for the certificate of proficiency in English (ECPE). *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, 6, 1–28.
- Linda, T. (1996, April). *A comparison of the traditional maximum information method and the global information method in CAT item selection*. Paper presented at National Council on Measurement in Education, New York.
- Linden, W. J., & Glas, G. A. W. (2002). *Computerized adaptive testing: Theory and practice*. USA: Kluwer Academic Publishers.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: ETS.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- Lunz, M. E., & Stahl, J. A. (1993). The effect of rater severity on person ability measure: A rasch model analysis. *The American Journal Of Occupational Therapy: Official Publication Of The American Occupational Therapy Association*, 47(4), 311-7
- Lunz, M. E., Stahl, J. A., & Bergstrom, B. A. (1993, April). *Targeting, test length, test precision, and decision accuracy for Computerized Adaptive Tests*. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.
- Martineau, J. A. (2004). *The effects of construct shift on growth and accountability models* (Unpublished doctoral dissertation). Michigan State University, East Lansing, MI.
- McCoy, K. M. (2009). *The impact of item parameter drift on examinee ability measures in a computer adaptive environment* (Unpublished doctoral dissertation). University of Illinois at Chicago, Chicago, IL.
- McDonald, P.L. (2002). *Computer adaptive test for measuring personality factors using item response theory* (Unpublished doctoral dissertation). The University Western of Ontario, London.

- Meng, H., Steinkamp, S., & Matthews-Lopez, J. (2010). *An investigation of item parameter drift in computer adaptive testing*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Denver, CO.
- Meyers, J., Miller, G. E., & Way, W. D. (2009, April). *Item position and item difficulty change in an IRT based common item equating design*. Paper presented at the American Educational Research Association, San Francisco, CA.
- Moyer, E. L., Galindo, J. L., & Dodd, B. G. (2012). Balancing flexible constraints and measurement precision in computerized adaptive testing. *Educational and Psychological Measurement*.
- Nydic, S.W. (2015). An R package for simulating IRT-based computerized adaptive tests.
- Patton, J.M., Cheng, Y., Yuan, K.H., & Diao (2013). The influence of item calibration error on variable-length computerized adaptive testing. *Applied Psychological Measurement*, 37(1), 24–40.
- Ranganathan, K., & Foster, I. (2003). Simulation studies of computation and data scheduling algorithms for data grids. *Journal of Grid Computing*, 1, 53-62.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- Reckase, M. D. (2011). Computerized adaptive assessment (CAA): The way forward. In *The road ahead for state assessments, policy analysis for California education and Rennie Center for Education Research & Policy* (pp.1-11). Cambridge, MA: Rennie Center for Education Research & Policy.
- Risk, N.M. (2015). *The impact of item parameter drift in computer adaptive testing (CAT)* (Unpublished doctoral dissertation), University of Illinois, Chicago.
- Roznowski, M. (1989). Examination of measurement properties of the job descriptive index with experimental items. *Journal of Applied Psychology*, 74, 805-814
- Rudner, L.M., & Guo, F. (2011). Computer adaptive testing for small scale programs and instructional systems. *Graduate Management Council (GMAC)*, 11(01), 6-10
- Rupp, A. A., & Zumbo, B. D. (2003a). Which model is best? Robustness properties to justify model choice among unidimensional IRT models under item parameter drift. *The Alberta Journal of Educational Research*, 49(3), 264-276.
- Rupp, A. A., & Zumbo, B. D. (2003b). *Bias coefficients for lack of invariance in unidimensional IRT models*. Vancouver: University of British Columbia.
- Rupp, A. A., & Zumbo, B. D. (2004). A note on how to quantify and report whether item parameter invariance holds: When Pearson correlations are not enough. *Educational and Psychological Measurement*, 64, 588-599.

- Rupp, A. A., & Zumbo, B. D. (2006). Understanding parameter invariance in unidimensional IRT models. *Educational and psychological measurement*, 66(1), 63-84.
- Schulz, W., & Fraillon, J. (2009, September). The analysis of measurement equivalence in international studies using the rasch model. *Paper presented at the European Conference on Educational Research (ECER)*, Vienna.
- Scullard, M.G. (2007). *Application of item response theory based computerized adaptive testing to the strong interest inventory* (Unpublished doctoral dissertation). University of Minnesota, USA.
- Segall, D. O. (2004). Computerized adaptive testing. K. Kempf-Lenard (Ed.), *The Encyclopedia of social measurement*. San Diego, CA: Academic Press.
- Skorupski, W. P. (2006, April). *The effects of item parameter drift on equating test scores*. Paper presented at the Annual Meeting of the National Council on Measurement, San Francisco, CA.
- Song, T., & Arce-Ferrer, A. (2009, April). *Comparing IPD detection approaches in common-item nonequivalent group equating design*. Paper presented at the Annual Meeting of The National Council on Measurement, San Diego, CA.
- Stahl, J. A., & Muckle, T. (2007, April). *Investigating displacement in the Winsteps Rasch calibration application*. Paper presented at the Annual Meeting of The American Educational Research Association, Chicago, IL.
- Sulak, S. (2013). *Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Kullanılan Madde Seçme Yöntemlerinin Karşılaştırılması* (Yayımlanmamış Doktora Tezi) Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Svetina, D., Crawford, A.V., Levy, R., Green, S.B., Scott, L., Thompson, M., Gorin, J.S., Fay, D., & Kunze, K.L. (2013). Designing small-scale tests: A simulation study of parameter recovery with the 1-PL. *Psychological Test and Assessment Modeling*, 55(4), 335-360.
- Şahin, A. (2012). *Madde Tepki Kuramı'nda Test Uzunluğu ve Örneklem Büyüklüğünün Model Veri Uyumu, Madde Parametreleri ve Standart Hata Değerlerine Etkisinin İncelenmesi* (Yayımlanmamış Doktora Tezi), Hacettepe Üniversitesi, Ankara.
- Urry, V. (1977). Tailored testing: A successful application of latent trait theory. *Journal of Educational Measurement*, 14, 181-196.
- van der Linden, W. (1998). Bayesian item selection criteria for adaptive testing. *Psychometrika*, 63(2), 201-216.
- van der Linden, W. J., & Glas, C. A. (2010). *Elements of adaptive testing*. New York, NY: Springer.

- Veldkamp, B.P., & Linden van der, W. (2006) Designing item pool for computerized adaptive testing. In *Designing Item Pools* (s.149-166). The Netherlands: University of Twente.
- Wainer, H. (1993). Some practical considerations when converting a linearly administered test to an adaptive format. *Educational Measurement: Issues and Practice*, 12(1), 15–20.
- Wainer, H. (2000). *Computerized adaptive testing*. London: Lawrence Erlbaum Assc.
- Wainer, H., Dorans, N. J., Eignor, D., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg, L., & Thissen, D. (2010). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- Wang, T. (1997, March). *Essential unbiased EAP estimates in computerized adaptive testing*. Paper presented at the American Educational Association, Chicago.
- Wang, W. C., & Chen, C. T. (2005). Item parameter recovery, standard error estimates, and fit statistics of the WINSTEPS program for the family of Rasch models. *Educational and Psychological Measurement*, 65(3), 376-404.
- Wang, H-P., Kuo, B-C., Tsai, Y-H., & Liao, C-H. (2012). A Cerf-Based computerized testing system for chinese proficiency. *TOJET: The Turkish Journal of Educational Technology*, 11(4), 1–12.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361–375.
- Wells, C. S., Subkoviak, M. J., & Serlin, R. C. (2002). The effect of item parameter drift on examinee ability estimates. *Applied Psychological Measurement*, 26(1), 77-87.
- Wise, S.L., & Kingsbury, G.G. (2000). Practical issues in developing and maintaining a computerized adaptive testing program, *Psicológica*, 21(2000), 135-155.
- Witt, E. A., Stahl, J. A., Bergstrom, B. A., & Muckle, T. (2003, April). *Impact of item drift with nonnormal distributions*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Wollack, J. A., Sung, H. J., & Kang, T. (2005) *Longitudinal effects of item parameter drift*. Paper presented at Annual Meeting of the National Council on Measurement in Education, Montreal, CA.
- Yao, L. (2013). Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. *Applied Psychological Measurement*, 37(1), 3-23.
- Yi, Q., Wang, T., & Ban, J. C. (2001). Effects of scale transformation and test termination rule on the precision of ability estimation in computerized adaptive testing. *Journal of Educational Measurement*, 38, 267-292.

Zhang, J. (2014). A sequential procedure for detecting compromised items in the item pool of a CAT system. *Applied Psychological Measurement*, 38(2), 87-104.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). BILOG-MG 3.0 [computer software]. Lincolnwood, IL: Scientific Software International.





EK 1. Etik Kurul Muafiyet Onayı

ANKARA ÜNİVERSİTESİ SOSYAL BİLİMLER ALT ETİK KURULU KARAR ÖRNEĞİ

Karar Tarihi : 22/04/2019

Toplantı Sayısı : 05

Karar Sayısı : 181

181- Üniversitemiz Eğitim Bilimleri Enstitüsü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı doktora öğrencisi **Merve Şahin Kürşad**'ın "Farklı Koşullar Altında Madde Parametre Sapmasının Bireye Uyarlanmış Testlerden Elde Edilen Yetenek Kestirimlerine Etkisinin İncelenmesi" başlıklı tezi ile ilgili 11/04/2019 tarihli "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

Üniversitemiz Eğitim Bilimleri Enstitüsü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı doktora öğrencisi **Merve Şahin Kürşad**'ın "Farklı Koşullar Altında Madde Parametre Sapmasının Bireye Uyarlanmış Testlerden Elde Edilen Yetenek Kestirimlerine Etkisinin İncelenmesi" başlıklı tezinin, insan üzerinde yapılan bir çalışma olmaması nedeniyle etik kurul onayına ihtiyaç duymadığına oy birliği ile karar verildi.

ASLININ AYNI DİR
22/04/2019


Prof. Dr. Muharrem ÖZEN
Ankara Üniversitesi
Etik Kurulu Başkanı

BENZERLİK BİLDİRİMİ

“Madde Parametre Sapmasının Bireye Uyarlanmış Testlerden Elde Edilen Yetenek Kestirimlerine Etkisinin Farklı Koşullar Altında 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	: 25.12.2020
Gönderim Numarası	: 1481259346
Sayfa Sayısı	: 86
Sözcük Sayısı	: 21773
Karakter Sayısı	: 144191
Benzerlik Oranı	: %2
Savunma Tarihi	: 17.12.2020

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: Merve ŞAHİN KÜRŞAD

Tarih: 25.12.2020

İmza: 

ÖZGEÇMİŞ

Adı ve Soyadı: Merve ŞAHİN KÜRŞAD

E-posta adresi: sahinmerv@gmail.com

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversitesi	Yıl
Lisans	Fen Bilgisi Öğretmenliği	ODTÜ	2006-2011
Yüksek Lisans	Ölçme ve Değerlendirme	Abant İzzet Baysal Üniversitesi	2013-2014

İş Deneyimi:

Unvan	Görev Yeri	Yıl
Araştırma Görevlisi	Abant İzzet Baysal Üniversitesi	2013-2014
Araştırma Görevlisi	Ankara Üniversitesi	2015- 2020
Öğretim Görevlisi	Milli Savunma Üniversitesi	2020- Devam ediyor