



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

BİREYSELLEŞTİRİLMİŞ ÇOK AŞAMALI TESTLERİN TEST BİRLEŞTİRME
YÖNTEMLERİNE GÖRE İNCELENMESİ

Ebru DOĞRUÖZ

Doktora Tezi

Ankara, 2018



Liderlik, arařtırma, inovasyon, kaliteli eęitim ve deęiřim ile

Daha ileriye... En İyiyeye...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

BİREYSELLEŞTİRİLMİŞ ÇOK AŞAMALI TESTLERİN TEST BİRLEŞTİRME
YÖNTEMLERİNE GÖRE İNCELENMESİ

INVESTIGATION OF ADAPTIVE MULTISTAGE TEST BASED ON TEST
ASSEMBLY METHODS

Ebru DOĞRUÖZ

Doktora Tezi

Ankara, 2018

Kabul ve Onay

Eđitim Bilimleri Enstitüsü MÜdürlüğüne,

Ebru DOĐRUÖZ'ün hazırladığı "Bireyselleştirilmiş Çok Aşamalı Testlerin Test Birleřtirme Yöntemlerine Göre İncelenmesi" başlıklı bu çalışma jürimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

Jüri Başkanı

Prof. Dr., Selahattin GELBAL



Jüri Üyesi (Danışman)

Prof. Dr., Hülya KELECIOĐLU



Jüri Üyesi

Doç. Dr., Nuri DOĐAN



Jüri Üyesi

Yrd. Doç. Dr., Ömer KUTLU



Jüri Üyesi

Yrd. Doç. Dr., H. Deniz GÜLLEROĐLU



Bu tez Hacettepe Üniversitesi Lisansüstü Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 19 / 02 / 2018 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca / / tarihinde kabul edilmiştir.

Prof. Dr. Ali Ekber ŞAHİN
Eđitim Bilimleri Enstitüsü Müdürü

Öz

Bu arařtırmada, bireyselleřtirilmiř ok ařamalı testlerde (BAT) rneklem byklėė, panel deseni ve modl uzunluėuna gre yukarıdan-ařaėıya ve ařaėıdan-yukarıya test birleřtirme yntemleri karřılařtırılmıřtır. Her iki test birleřtirme yntemine gre oluřturulan BAT'lar sınıflandırma testi aısından da deėerlendirilmiřtir. Arařtırma kapsamında PISA 2015 yılı uygulamasına ait veriler kullanılmıř, bu verilerden kestirilen parametrelere gre simlasyon alıřması yapılmıřtır. Madde havuzunu, 2015 yılında uygulanan PISA verilerinden elde edilen iki kategorili veri seti oluřturmuřtur. Arařtırma deėiřkenlerinden modl uzunluėu deėiřkeni 6 ve 12, panel deseni deėiřkeni "1-2", "1-2-2" ve "1-2-3", rneklem byklėė deėiřkeni 250 ve 2000 olarak belirlenmiřtir. Analiz sonuları ortalama hata, yanlılık ve sınıflandırma doėruluėu deėerleri aısından karřılařtırılmıřtır. Arařtırmanın yukarıdan-ařaėıya test birleřtirme yntemine gre oluřturulan BAT simlasyonundan elde edilen sonularına gre kk ve byk rneklem uygulanan BAT'ın modl uzunluėu arttıėında ve panel deseni "1-2"den "1-2-2" ve "1-2-3"e doėru getiėinde ortalama hata deėerleri dřmřtr. rneklem byklėėnn artması, "1-2" ve "1-2-2" panel desenlerinin kısa ve orta modl uzunluėunda ortalama hata deėerlerinde bir miktar dřř oluřturmuřtur. Modl uzunluėunun artması her iki rneklem trnde panel desenlerinin yanlılıėını azaltmıřtır. Panel deseninin "1-2"den "1-2-2" ve "1-2-3"e doėru geiři her iki rneklem byklėė ve modl uzunluėunda ortalama yanlılık deėerlerini dřrmřtr. rneklem byklėėnn artması aynı Őekilde tm kořullar iin yanlılıėı kk bir miktar azaltmıřtır. BAT'ı yanıtlayan bireylerin %70, %50 ve %30 geme oranlarında yksek sınıflandırma oranları ve Kappa uyum katsayıları hesaplanmıřtır. Arařtırmanın ařaėıdan-yukarıya test birleřtirme yntemine gre oluřturulan BAT simlasyonundan elde edilen sonularına gre kk ve byk rneklem uygulanan BAT'larda modl uzunluėu ve panel deseni arttıėa ortalama hata deėeri azalmıřtır. rneklem byklėėnn ve modl uzunluėunun artması kk ve byk rneklem uygulanan "1-2", "1-2-2" ve "1-2-3" panel desenlerinin yanlılıėını azaltmıřtır. Her iki rneklem uygulanan kısa ve orta uzunlukta modln olduėu "1-2", "1-2-2" ve "1-2-3" panel desenlerinden oluřan BAT'ların; %70, %50 ve %30 geme oranlarında olduka yksek dzeyde bireyleri doėru sınıflandırdıėı sonucuna ulařılmıřtır.

Anahtar sözcükler: bireyselleştirilmiş çok aşamalı test, test birleştirme, sınıflandırma doğruluğu, PISA



Abstract

In this research, the performance of top-down and bottom-up test assembly methods were compared according to the sample size, panel design and module length of the Adaptive Multistage Tests (MST). MSTs formed according to the both test assembly methods were also evaluated in terms of the classification test. Within the context of the research, data for PISA 2015 were utilized and a simulation study was conducted according to the parameters predicted based on these data. Item parameters calibrated according to the 2 PL model from the two-category data set obtained the PISA data applied in the year 2015 comprised the item pool of the research. Of the research variables, module length variables 6 and 12, panel design variable “1-2”, “1-2-2” and “1-2-3”, and the sample size variables 250 and 2000 were determined in the study. Analysis results for each condition were compared in terms of mean error, bias and classification accuracy. According to the results obtained from the MST simulation formed based on top-down test assembly method, it was determined that the mean error value decreased when the module length of MST applied to the small and large samples increased. The mean error values decreased in case of the transition from “1-2” panel design to panel design “1-2-2” and “1-2-3” took place. The increase in the sample size lead to a certain amount of decrease in the mean error values of “1-2” and “1-2-2” panel design in short and moderate module lengths. The increase in the module length lowered the bias values of the panel design in both sample types. The transition of panel design from “1-2” to “1-2-2” and “1-2-3” decreased the mean bias values in both sample sizes and module lengths. The increase in sample size also reduced the bias slightly for all conditions. Individuals who filled the MST were compared with the passing ratios 70%, 50% and 30% respectively and the high classification ratios and Cohen’s Kappa coefficient of agreement. According to the results obtained from the MST simulation constructed according to bottom-up test assembly method, the mean error value decreased as the module length and panel design in MSTs applied to small and large samples increased. The increase in the sample size lead to a decrease in the mean error values of “1-2” panel design in short and moderate module lengths, “1-2-2” panel design in short module length and “1-2-3” panel design in moderate module length. The increase in sample size and module length decreased the bias values

of panel design “1-2”, “1-2-2” and “1-2-3” applied to the small and large samples. It was found that MSTs consisting of panel design “1-2”, “1-2-2” and “1-2-3” that contain short and moderate length modules applied to both samples classified the individuals at a high rate at 70%, 50% and 30% passing ratios.

Keywords: adaptive multistage test, test assembly, classification accuracy, PISA



Teşekkür

Doktora eğitimime başladığım günden bu yana bilgi ve tecrübeleriyle beni destekleyen, önerileri ve fikirleriyle beni yönlendiren, değerli katkıları bulunan, örnek aldığım, çok sevdiğim danışmanım Prof. Dr. Hülya KELEÇİOĞLU'na,

Eğitim-öğretim sürecimin her aşamasında yakın ilgilerini gördüğüm, desteklerini hiçbir zaman esirgemeyen değerli hocalarım Prof. Dr. Selahattin GELBAL ve Doç. Dr. Nuri DOĞAN'a,

Değerli görüş ve önerileri ile bu araştırmaya büyük katkı sunan sevgili hocalarım Yrd. Doç. Dr. Ömer KUTLU ve Yrd. Doç. Dr. Deniz GÜLLEROĞLU'na,

Desteğiyle yanımda olan ve beni yareklendiren, akademisyen olma yolunda kendisinde çok şey öğrendiğim, örnek aldığım, çok değerli, kıymetli hocam Doç. Dr. Murat ÖZDEMİR'e,

Doktora sürecinde üzerimde emeği olan ve ders aldığım tüm hocalarıma,

Bu araştırmayı tamamlayabilmem için beni destekleyen değerli hocam Yrd. Doç. Dr. Fatih GÜZEL'e,

Yardımlarına ne zaman ihtiyaç duysam yanımda olan sevgili arkadaşlarım Dr. Çiğdem AKIN ARIKAN ve Arş. Gör. Esmâ ACAR'a,

Bu uzun süreçte desteğini her zaman hissettiğim, en büyük destekçim ve güç kaynağım olan canım annem Güner DEMİRCİOĞLU ve canım babam Osman DEMİRCİOĞLU'na,

Araştırmanın her aşamasında destek olan sevgili abim Ersin DEMİRCİOĞLU'na,

Araştırmanın birçok kısmında fikrine başvurduğum, motivasyonumun kırıldığı anlarda cesaretlendiren, heyecanımı daima paylaşan ve önerilerini hiç esirgemeyen, rol modelim, sevgili ablam Yasemin AKDEVELİOĞLU'na ve

Bu çalışmanın oluşum sürecinde özverisi, anlayışı ve yardımı ile her zaman destek olan, beni cesaretlendiren, uzaklıkları yakın kılan, kalben hep yanımda hissettiğim, tezin en önemli paydaşı olan yol arkadaşım, sevgili eşim Cihan DOĞRUÖZ'e

sonsuz teşekkürlerimi sunuyorum.

İçindekiler

Öz.....	ii
Abstract.....	iv
Teşekkür.....	vi
Tablolar Dizini.....	ix
Şekiller Dizini.....	x
Simgeler ve Kısaltmalar Dizini.....	xi
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	6
Araştırma Problemi.....	9
Sayıltılar.....	10
Sınırlılıklar.....	10
Tanımlar.....	10
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	11
Madde Tepki Kuramı.....	11
Bireyselleştirilmiş Çok Aşamalı Test (BÇAT).....	14
PISA (Program for International Student Assessment-Uluslararası Öğrenci Başarısını Belirleme Programı).....	30
BÇAT ile ilgili çalışmalar.....	33
Otomatik Test Birleştirme İle İlgili Çalışmalar.....	38
İlgili Araştırmalar Özet.....	41
Bölüm 3 Yöntem.....	42
Araştırma Modeli.....	42
Çalışma Grubu.....	42
Verilerin Analizi.....	43
Verilerin Analizinin Birinci Aşaması.....	44

Verilerin Analizinin İkinci Aşaması.....	48
Bölüm 4 Bulgular	60
Bölüm 5 Tartışma, Sonuç ve Öneriler	75
Tartışma.....	75
Sonuç.....	80
Öneriler	82
Kaynaklar	84
EK-A: PISA 2015 Yılı Uygulamasına Ait Veri Setinin Faktör Yükleri.....	96
EK-B. PISA 2015 Yılı Uygulamasına Ait Veri Setinin 2 PL Modelde Kestirilen Madde Parametreleri.....	97
EK-C. Yukarıdan-aşağıya Test Birleştirme Yöntemine Göre Kestirilen Madde Ayırt Edicilik Parametrelerinin Ortalama ve Standart Sapma Değerleri	99
EK-Ç. Yukarıdan-aşağıya Test Birleştirme Yöntemine Göre Kestirilen Madde Güçlük Parametrelerinin Ortalama ve Standart Sapma Değerleri	101
EK-D. Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Kestirilen Madde Ayırt Edicilik Parametrelerinin Ortalama ve Standart Sapma Değerleri	103
EK-E. Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Kestirilen Madde Güçlük Parametrelerinin Ortalama ve Standart Sapma Değerleri	105
EK-F: Etik Komisyonu Onay Bildirimi	107
EK-G: Etik Beyanı	108
EK-Ğ: Doktora Tez Çalışması Orijinallik Raporu	109
EK-H: Dissertation Originality Report	110
EK-I: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	111

Tablolar Dizini

Tablo 1 <i>Kağıt-kalem Testi, Bireyselleştirilmiş Bilgisayarlı Test ve BÇAT'ın Özellikleri</i>	27
Tablo 2 <i>Bileşenlere Ait Açıklanan Varyans Oranı</i>	44
Tablo 3 <i>Modellere İlişkin -2 Log (Olabilirlik) Değerleri</i>	45
Tablo 4 <i>2015 Yılında Gerçekleştirilen PISA Uygulamasına Ait Madde ve Yetenek Parametrelerine İlişkin Betimleyici İstatistikler</i>	46
Tablo 5 <i>Tesadüfi Olarak Atanan 11 Gruba Ait Madde Parametreleri Arasındaki Korelasyonlar</i>	47
Tablo 6 <i>Yetenek Parametreleri Arasındaki Korelasyonlar</i>	47
Tablo 7 <i>BÇAT Değişkenleri</i>	48
Tablo 8 <i>Panel Tasarımlarına Ait Modül ve Test Uzunlukları</i>	52
Tablo 9 <i>Olası Sınıflandırma Kararı Durumlarının Değerlendirilmesi</i>	58
Tablo 10 <i>Yukarıdan-aşağıya Test Birleştirme Yöntemine Göre Çeşitli Simülasyon Koşullarına Göre Oluşturulan BÇAT'lara İlişkin Ortalama RMSE ve Yanlılık Değerleri</i>	61
Tablo 11 <i>Yukarıdan-aşağıya Test Birleştirme Yöntemi Seçildiğinde Elde Edilen Ortalama RMSE ve Yanlılık Değerlerine Ait ANOVA Sonuçları</i>	64
Tablo 12 <i>Yukarıdan-aşağıya Test Birleştirme Yöntemine Göre Modül Uzunluğu, Panel Deseni ve Örneklem Büyüklüklerine Ait Sınıflandırma Doğruluğunun Karşılaştırılması</i>	66
Tablo 13 <i>Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Çeşitli Simülasyon Koşullarına Göre Oluşturulan BÇAT'lara İlişkin Ortalama RMSE ve Yanlılık Değerleri</i>	68
Tablo 14 <i>Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Çeşitli Simülasyon Koşulları İçin Oluşturulan BÇAT'lara İlişkin Ortalama RMSE ve Yanlılık Değerlerine Ait ANOVA Sonuçları</i>	71
Tablo 15 <i>Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Modül Uzunluğu, Panel Deseni ve Örneklem Büyüklüklerine Ait Sınıflandırma Doğruluğunun Karşılaştırılması</i>	73

Şekiller Dizini

Şekil 1. BÇAT Bileşenleri.....	4
Şekil 2. “1-3-3” Panel Deseni.....	16
Şekil 3. Örnek BÇAT Yapıları.....	27
Şekil 4. İki Panelden Oluşan İki Aşamalı “1-2” Panel Deseni.	49
Şekil 5. İki Panelden Oluşan Üç Aşamalı “1-2-2” Panel Deseni.	50
Şekil 6. İki Panelden Oluşan Üç Aşamalı “1-2-3” Panel Deseni.	51
Şekil 7. “1-2-3” BÇAT Panel Desenine İlişkin OMB Yönlendirme Yöntemi.....	54
Şekil 8. Yukarıdan-aşağıya Test Birleştirme Yöntemine Göre Elde Edilen Ortalama RMSE ve Yanlılık Değerlerine İlişkin Grafikler.....	63
Şekil 9. Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Elde Edilen Ortalama RMSE ve Yanlılık Değerlerine İlişkin Grafikler.....	70

Simgeler ve Kısaltmalar Dizini

1 PL: 1 Parametrelili Lojistik

2 PL: 2 Parametrelili Lojistik

3 PL: 3 Parametrelili Lojistik

ASM: Ağırlıklı Sapma Modeli (Weighted Deviation Model -WDM)

BBT: Bireyselleştirilmiş Bilgisayarlı Test (Computer Adaptive Test -CAT)

BBST: Bilgisayarda Bireyselleştirilmiş Sıralı Test (Computer Adaptive Sequential Test -CAST)

BÇAT: Bireyselleştirilmiş Çok Aşamalı Test (Multistage Adaptive Test -MST)

BSD: Beklenen Sonsal Dağılım (Expected a Posteriori -EAP)

CPA: Uniform Certified Public Accountant Examination (Mali Müşavirlik Sınavı)

ESL: English as a Second Language (İngilizce Yerleştirme ve Tanımlayıcı Test)

ETS: Educational Testing Service (Eğitimsel Test Hizmetleri)

GMAT: Graduate Management Admission Test (İşletme Alanında Lisansüstü Eğitime Giriş Sınavı)

GRE: Graduate Record Examination (Mühendislik ve Sosyal Bilimler Alanında Lisansüstü Giriş Sınavı)

KTK: Klasik Test Kuramı (Classical Test Theory)

KTP: Karma Tam Sayı Programlama (Mixed Integer Programming -MIP)

LSAT: Law School Admission Test (Hukuk Fakültelerine Kabul Sınavı)

MOK: Maksimum Olabilirlik Kestirimi (Maximum Likelihood Estimation -MLE)

MÖE: Maksimum Öncelikli İndeks (Maximum Posterior Index -MPI)

MSD: Maksimum Sonsal Dağılım (Maximum Posterior Index -MPI)

MTK: Madde Tepki Kuramı (Item Response Theory)

NAEP: The National Assessment of Educational Progress (Ulusal Eğitimdeki Gelişmelerin Değerlendirilmesi)

NAMSS: Normalleştirilmiş Ağırlıklı Mutlak Sapma Sezgisel (Normalized Weighted Absolute Deviation Heuristics -NWADH)

NCLEX: Nurses Certification and Licensure Examination (Hemşirelik Lisansına Kayıt Sınavı)

OECD: Organization of Economic Cooperation and Development (Ekonomik İşbirliği ve Kalkınma Örgütü)

OMB: Ortalama Maksimum Bilgi (Approximate Maximum Information -AMI)

OTB: Otomatik Test Birleştirme (Automated Test Assembly -ATA)

PISA: Programme for International Student Assessment (Uluslararası Öğrenci Değerlendirme Programı)

RMSE: Root Mean Squared Error (Hataların Ortalama Karekökü)

SOOT: Sıralı Olasılık Oran Testi (Sequential Probability Ratio Test -SPRT)

TBF: Test Bilgi Fonksiyonu

TOEFL: Test of English as a Foreign Language (Yabancı bir Dil olarak İngilizce Sınavı)

TSBÇAT: Test Sırasında Test Birleştirmenin Yapıldığı Bireyselleştirilmiş Çok Aşamalı Test (On-The-Fly Assembled Multistage Adaptive Test -OMST)

USMLE: The U.S. Medical Licensure Examination (ABD'de Tıp Alanında Uzmanlık Sınavı)

Bölüm 1

Giriş

Bu bölümde problem durumu, araştırmanın amacı ve önemi, problem cümlesi, alt problemler, sayılılar, sınırlılıklar ve tanımlar yer almaktadır.

Problem Durumu

Bilgisayar teknolojisi ve Madde Tepki Kuramı (MTK) ile test uygulamalarının bütünleşmiş olması, Bireyselleştirilmiş Bilgisayarlı Test (BBT -Computer Adaptive Test) desenlerini ortaya çıkarmıştır. Bu testlerde bilgisayarın kullanılması, bireye özgü testlerin oluşturulmasını sağlarken; MTK, bireyin yeteneğini göz önünde bulundurarak testin hazırlanması, uygulanması ve değerlendirilmesi olanağını sağlamaktadır. Geniş ölçekli testlerin uygulandığı bireylerin yetenekleri, oldukça geniş bir ölçek aralığında sıralanır. Bu bireylere kağıt-kalem testi uygulandığında testin içerisindeki maddeler yüksek yetenek düzeyindeki bireyler için oldukça kolay, düşük yetenek düzeyindeki bireyler için oldukça zor olabilir. Ancak aynı bireylere BBT uygulandığında, bireyler yalnızca yetenek düzeylerine uygun güçlükteki maddeleri yanıtlar. Bu nedenle Amerika Birleşik Devletleri'nde (ABD) uygulanan *Lisansüstü Eğitime Giriş Sınavı (Graduate Management Admission Test -GMAT)*, *Hemşirelik Lisansına Kayıt Sınavı (Nurses Certification and Licensure Examination -NCLEX)* ve ikinci dili İngilizce olan öğrenciler için *İngilizce Yerleştirme ve Tanımlayıcı Testi (English as a Second Language -ESL)* gibi birçok test, kağıt-kalem test türünden BBT'ye dönüşmüştür.

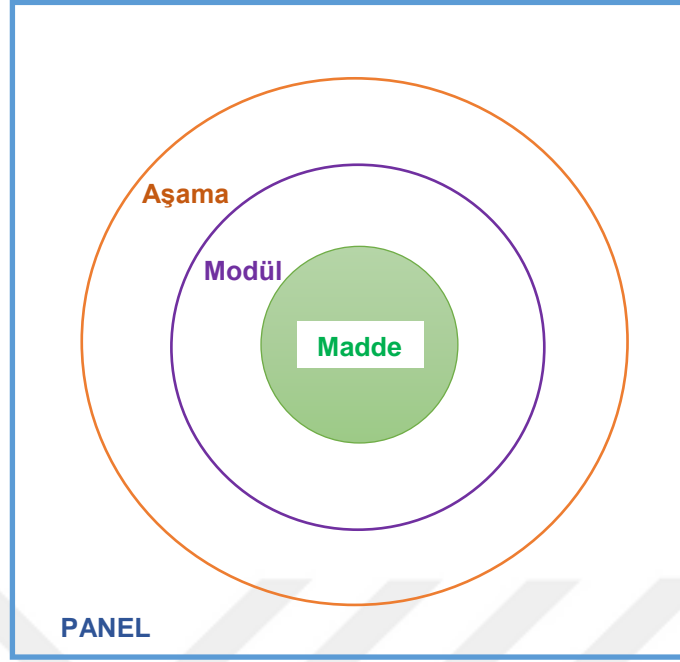
BBT'nin uygulanması bireyin yeteneğinin kestirilmesi ile başlar. Bu doğrultuda bireye yöneltilen genellikle orta güçlükteki bir maddenin yanıtlanmasının ardından bireyin yeteneğinin ölçek üzerindeki yeri kestirilir. Ardından madde havuzundan bireyin kestirilen yeteneğine uygun madde seçilir. Seçilen maddenin bireye yöneltilmesinden sonra bireyin yeteneği yeniden kestirilir. Bu kestirilen değer yeni bir maddenin seçilmesi için kullanılır. Test, önceden belirlenen bir ölçme doğruluğu elde edildiğinde veya maksimum test uzunluğuna erişildiğinde sonlandırılır. BBT'nin uygulanma süresi, geleneksel kağıt-kalem testlerine göre çok daha kısadır. BBT'de kısa uzunluktaki testler ile en uygun ölçme duyarlılığı elde edilir. Bununla birlikte test uzunluğunun kısalması bireyin yorgunluğunu azaltarak test performansını yükseltebilmektedir. BBT'nin bilgisayar aracılığıyla gerçekleştirilmesi bireye testin tamamlanmasından sonra test başarısı

ile ilgili geribildirim verme, test yöneticisine ise yeni madde türlerini madde havuzuna ekleme olanağını sunar. Test formu bireyselleştirildiğinden, kağıt-kalem testinde olduğu gibi testi yanıtlayan bireyin testi yanıtlayan diğer bireylerle birlikte hareket etmesi gerekli değildir (Rudner, 1998). Bu açıklamalar doğrultusunda BBT'nin özellikleri; testin kısa sürede puanlanabilmesi, test yönetiminin standardizasyonu, çeşitli bilgi türlerinin toplanması ve depolanması, yenilikçi madde türlerinin benimsenmesi ve test güvenliğinin daha iyi kontrol edilmesi şeklinde özetlenebilir (Chalhoub-Deville & Deville, 1999).

Bilgisayar ortamında bireyselleştirilmiş testlerin tarihsel gelişimi incelendiğinde, ilk uygulamanın Reckase tarafından 1974 yılında gerçekleştirildiği görülmektedir (Wise & Kingsbury, 2000). Eğitimsel Test Hizmetleri (Educational Testing Service -ETS) 1993 yılında, ilk geniş ölçekli BBT uygulamasını ABD'deki birçok üniversitenin yüksek lisans programlarına başvurulması için gerekli olan *Mühendislik ve Sosyal Bilimler Alanında Lisanüstü Giriş Sınavı (Graduate Record Examination -GRE)* ile gerçekleştirmiştir. 1997 yılında ETS, işletme programlarında yükseköğrenim için GMAT'ın BBT versiyonunu uygulamaya koymuştur. 1998 yılında *Yabancı bir Dil olarak İngilizce Sınavı (Test of English as a Foreign Language -TOEFL)*, BBT'ye göre düzenlenmeye başlamış ve bu uygulama 2001 yılı itibariyle kağıt-kalem testlerinin yerini almıştır (Slater, 2001; Weissmann, 2003). 1990-2000 yılları arasında ölçme alanında BBT'ye olan ilgi artmış ve bu testin kullanımı yaygınlaşmıştır. Ancak gerek test formunun test sonlanana kadar birleştirilememesi, gerekse de testi alan bireylere yanıtladıkları önceki maddeleri inceleme olanağı tanımamasından; günümüzde BBT uygulamaları giderek yerini Bireyselleştirilmiş Çok Aşamalı Testlere (BÇAT) bırakma eğilimine girmiştir. Bu eğilimin ardındaki nedenlerden biri BÇAT'ın kağıt-kalem ve bilgisayarlı testler gibi doğrusal test formları ile madde düzeyinde bireyselleştirilmiş bilgisayarlı test formları arasında köprü görevini üstlenmesidir. Bir diğer ifadeyle BÇAT, her iki test formunun üstünlüklerini kendi bünyesinde birleştirmektedir. BÇAT, hem bireyselleştirilmiş bir testtir; hem de test geliştiricisine önceden test formlarını inceleme, bireye de yanıtlarını kontrol etme imkanı sunan bir testtir. Bu önemli avantajlarından dolayı, BÇAT'ın popülerliği son yıllarda giderek artış göstermektedir (Yan, von Davier & Lewis, 2014).

BÇAT yıllar içinde farklı isimler ile adlandırılmıştır. Bu adlandırmalar arasında iki aşamalı test (two-stage testing, Adema, 1990; Kim & Plake, 1993), bilgisayarlı yeterlik testi (computerized mastery testing, Lewis & Sheehan, 1990), çok aşamalı test (multistage testing, Patsula, 1999), yığınlı çok aşamalı bilgisayarlaştırılmış test (bundled multistage adaptive test, Luecht, 2003), çoklu form yapıları (multiple forms structure, Armstrong, Jones, Koppel & Pashley, 2004), bilgisayarlaştırılmış çok aşamalı test (computerized multistage testing, Ariel, Veldkamp & Breithaupt, 2006) ve çok aşamalı bireyselleştirilmiş test (multistage adaptive testing, Zheng, Nozawa, Gao & Chang, 2012) yer almaktadır. Bu araştırmada ise BÇAT, Bireyselleştirilmiş Çok Aşamalı Test (Multistage Adaptive Test) olarak adlandırılmıştır.

BÇAT, bir testin güçlüğüne testi alan bireyin yetenek düzeyine göre uyarlanabilmesine izin veren spesifik bir değerlendirme türü olarak tanımlanabilir. Bu değerlendirme türü *modül*, *aşama*, *panel* ve *yol* olmak üzere birbiri içerisinde kümelenmiş bileşenlerden oluşur. Bu kümenin en küçük elemanı *modüldür*. Modül, maddelerin bir araya gelmesiyle oluşan madde takımıdır. Bir modülün içerisindeki maddeler içerik açısından farklı olsa da benzer istatistiksel özellikler gösterir. Modül veya modüllerin bulunduğu düzey, *aşama* olarak adlandırılır. Bir aşamanın içerisinde yer alan modül veya modüllerin istatistiksel özellikleri testin amacına göre belirlenir. Aşamaların da bir araya gelmesiyle oluşan desen *panel* olarak adlandırılır. Panel, BÇAT'ın en büyük bileşenidir. Örneğin; birinci aşamada 1 modül, ikinci aşamada 2 modül ve üçüncü aşamada 3 modülün olduğu bir panel, "1-2-3" BÇAT panel deseni olarak adlandırılır. Bir bireyin panel içerisindeki aşamalar ve modüller arasında izleyeceği sıra *yol* (*pathway*) olarak adlandırılır. Her birey test süresince yalnızca bir yolu izler (Zenisky & Hambleton, 2014). BÇAT bileşenlerine ilişkin şematik gösterim Şekil 1'de sunulmuştur.



Şekil 1. BÇAT bileşenleri.

BÇAT'ın oluşturulma sürecindeki belki de en önemli ve en zor adım, testin birleştirilmesidir (test assembly). Testin birleştirilmesi modül ve panellerin bir araya getirilmesi amacıyla gerçekleştirilir. BÇAT'ın uygun bir şekilde birleştirilmesi doğru sonuçların elde edilmesi için oldukça önemlidir. El ile yapılan hesaplamalarla testin birleştirilmesi, zaman ve doğruluk yönünden uygun bir seçenek olmayabilir. Ancak, dijital bilgisayar teknolojisinin üstünlüklerinden yararlanılarak otomatik test birleştirme (OTB) düzeneği yardımı ile testin birleştirilmesi gerçekleştirilebilir. OTB bilgisayar yazılımı, çoklu panelleri BÇAT'da aynı anda üretmek için doğrusal (linear) programlama gibi optimizasyon algoritmalarını kullanır. BÇAT uygulamasında önce madde havuzundaki maddeler birleştirilerek modüller, sonrasında modüller birleştirilerek paneller birleştirilir. Aynı anda üç farklı amacın gerçekleştirilmesi gerekliliği testin birleştirilmesini karmaşıktırır. Bu amaçlar:

- bir aşamada yer alan farklı güçlük düzeyindeki modüllerin bilgi eğrilerinin aşamalar arasında farklı olmasının sağlanması,
- farklı panellerdeki aynı yolların bilgi eğrilerinin paralel paneller oluşturabilmek için yeterince benzer olmasının sağlanması,
- panelde takip edilecek olan olası tüm yollar için birleştirme kısıtlamalarının karşılanmasıdır (Zheng, Wang, Culbertson & Chang, 2014).

Örneğin; madde havuzu boyutunun küçük olması ve izlenecek yolların sayısının fazla olması bu üç amacın da gerçekleştirilebilme olasılığını düşürür.

BÇAT'ı oluşturan ve onu diğer test türlerinden ayıran birtakım özellikler vardır. Bu özelliklerden ilki BÇAT'da maddelerin önceden belirlenen modüllerin içerisinde olmasıdır. Böylelikle sonraki modüllerin seçimi tek bir maddenin performansına göre değil, daha önce yöneltilen modüllerin performansına göre gerçekleştirilir (Magis, Yan & von Davier, 2016). Bu durum bireyselleştirmenin modül düzeyinde yapılmasını sağlar.

BÇAT'ın bir diğer özelliği test formlarının önceden yapılandırılmasıdır. Bu özellik sayesinde madde kullanım sıklığı (item exposure) ve içerik ağırlıklandırılması (content balancing) gibi test nitelikleri kontrol edilebilir. Ayrıca bu özellik test yöneticisine, BÇAT üzerinde çeşitli test gereksinimlerinin sağlanması için ayrıntılı ayarlama ve değişim (modifikasyon) yapabilme olanağı sunar. Diğer bir anlatımla, test formlarındaki maddeler için güçlü bir kalite-kontrol süreci gerçekleştirilebilir. Bu süreçte ölçme ve değerlendirme uzmanları ve test kapsamını belirleyen alan uzmanları tarafından test formunda yer alan maddeler değerlendirilerek, maddelerin istatistiksel özellikleri incelenir (Hendrickson, 2007; Luecht & Nungester, 1998; Patsula, 1999). Test formlarının önceden yapılandırılması bireylere, modüllerin içerisindeki maddeleri tekrar gözden geçirme ve yanıtları değiştirilebilme izni verir. Bireyin modülü tamamlamasının ardından da yetenek kestirimi gerçekleştirilir.

BÇAT'ın bir başka özelliği, madde havuzunun yüksek nitelikteki maddelerin kullanımı açısından pratik ve ekonomik olmasıdır. Testlerin hazırlanmasında bilgisayardan yararlanıldığından her birey için madde yönetimi özel olarak gerçekleştirilir. Kolay olan maddeler zor olan maddelere göre daha sık kullanılır. Ayrıca madde kullanım sıklığının artması test güvenliğini tehdit eder. Bu nedenle sıklıkla kullanılan maddelerin yönetimi kısıtlanarak madde havuzu daha etkili kullanılır.

BÇAT, sınıflandırma testleri için etkili bir bireyselleştirilmiş test yaklaşımı olarak düşünülebilir. Sınıflandırma testleri geçme/kalma, sertifika alabilir/alamaz gibi iki veya temel/orta/ileri düzey gibi çok yönlü kararlar vermek için yaygın şekilde kullanılmaktadır. Bu testler bireyin farklı test standartlarına göre

hedeflenen yetenek düzeyinde olup olmadığı hakkında bilgi verir (Bergstrom & Lunz, 1999; Parshall, Spray, Kalohn, & Davey, 2002; Thompson, 2009). Sınıflandırma kararının verileceği bir testte BÇAT, bireyselleştirilmiş bilgisayar aracılığıyla uygulanan bir test olması ve diğer test türlerinden farklı bileşenlere sahip olmasından dolayı tercih edilebilir.

Görüldüğü üzere, kağıt-kalem testi ve BBT ile karşılaştırıldığında BÇAT'ın birçok avantajı bulunmaktadır. Bu avantajlar da BÇAT'ın yaygınlaşmasını ve geniş ölçekli testler için tercih edilmesini sağlamaktadır. Ancak bir testin BÇAT'a göre oluşturulması sırasında dikkat edilecek birtakım özellikler vardır. Bu özelliklerden birisi olan 'test birleştirme' yöntemi ile ilgili alanyazın incelendiğinde çalışmalarda sıklıkla Luecht ve Nungester (1998) tarafından önerilen 'yukarıdan-aşağıya' ve 'aşağıdan-yukarıya' test birleştirme yöntemlerinden birisinin kullanıldığı görülmektedir. Ancak bu test birleştirme yöntemlerinin farklı koşullarda nasıl sonuç verdiğine ilişkin deneysel çalışmalar bulunmamaktadır. Dolayısıyla seçilecek test birleştirme yönteminin belirlenmesinde doğru kararın verilip verilmediği ile ilgili bir soru işareti oluşmaktadır. Bu nedenle bu çalışmanın BÇAT'ın oluşturulması aşamasında 'yukarıdan-aşağıya' veya 'aşağıdan-yukarıya' test birleştirme yöntemlerinden hangisinin tercih edileceği konusunda araştırmacıları doğru yönlendirebileceği düşünülmektedir.

Araştırmanın Amacı ve Önemi

Bu çalışmada, Uluslararası Öğrenci Değerlendirme Programı (Programme for International Student Assessment -PISA) 2015 uygulamasındaki verilerden elde edilen Madde Tepki Kuramı'na dayalı birey ve yetenek parametreleri kestirilen, çeşitli koşullara göre oluşturulan BÇAT yapılarında, iki farklı test birleştirme yönteminin karşılaştırılması amaçlanmıştır. Bu doğrultuda *yukarıdan-aşağıya* ve *aşağıdan-yukarıya* test birleştirme yöntemleri modül uzunluğu, panel deseni ve örneklem büyüklüğüne göre irdelenmiştir. Bu çerçevede BÇAT yapılarında minimum hataya ve yanlılığa sahip kestirimler yapabilmek için gerekli olan modül uzunluğu, panel deseni ve örneklem büyüklüğüne ilişkin öneriler geliştirilmiştir. Bununla birlikte, araştırma kapsamında her iki test birleştirme yöntemine göre oluşturulan BÇAT'lar sınıflandırma testi olarak da

değerlendirilmiştir. Sonuç olarak bu çalışmada BÇAT'ın test birleştirme yöntemlerinin uygulanabilirliği test edilmiştir.

BBT ile BÇAT'ın performansının karşılaştırıldığı çeşitli çalışmalar bulunmaktadır (Davis & Dodd, 2003; Jodoin, 2003; Kim & Plake, 1993; Luecht, Nungester & Hadidi, 1996; Patsula, 1999). Bu çalışmaların sonuçlarında BBT'nin yetenek düzeylerinin ölçüm aralığında daha iyi ölçme kesinliği elde ettiği, BÇAT'ın ise özellikle test sürecinin yönetilmesinde bir modüldeki maddelerin gözden geçirilebilmesi, bir sonraki maddeye ilerlenebilmesi veya geriye dönülebilmesi, test içeriğindeki bir modülün önceden seçilebilmesi gibi üstünlüklere daha fazla sahip olduğu görülmüştür. Bu nedenle bu çalışmada, alternatif bir bireyselleştirilmiş test türü olan BÇAT çalışılmıştır.

Bireyselleştirilmiş test ile ilgili alanyazın temelinde, test birleştirme yöntemlerinin uygulanması ve incelenmesi konusunda kısa sürede çok fazla ilerleme kaydedildiğinden söz edilebilir. BÇAT, doğrusal test birleştirme yöntemleri kullanılarak birleştirilebilmesine rağmen, özellikle çoklu paralel panellerin birleştirilmesi gerektiğinde, BÇAT'ın birleştirilmesi doğrusal test formlarının birleştirilmesinden çok daha fazla karmaşıktır. Luecht ve Nungester (1998), bireyselleştirilmiş çok aşamalı testin oluşturulması için öncelikle modül ve panellerin en uygun şekilde bir araya getirilmesini sağlayan bir yonteme ihtiyacın olduğunu belirtmişlerdir. Bunun üzerine de test birleştiricisi olarak tanımlanan iki yöntem önermişlerdir. Bu yöntemler 'yukarıdan-aşağıya' ve 'aşağıdan-yukarıya' test birleştirme yöntemleridir. Her iki yöntem de otomatik test birleştirme bilgisayar yazılımı aracılığıyla gerçekleştirilir. Bu yöntemlerin işlevi madde havuzundan seçilen maddelerle modüller, sonrasında modüllerden paneller oluşturmaktır. Bu yöntemler arasında istatistiksel olarak birtakım farklılıklar vardır. BÇAT'ın bu aşaması, test geliştirme sürecinde oldukça önemli bir adımdır (Wendler & Walker, 2006). Bu nedenle, iyi tanımlanmış test koşullarının yerine getirilebilmesi için uygun bir test birleştirme yönteminin seçilmesinin öneminden bahsedilmelidir. Bu doğrultuda araştırmamızın odak noktası BÇAT çalışmalarında sıklıkla tercih edilen yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemlerinin karşılaştırılmasıdır. İlgili alanyazında BÇAT'ın mevcut uygulamalarında çoğunlukla aşağıdan-yukarıya test birleştirme yöntemi kullanılmıştır (Hembry, 2014; Jodoin, Zenisky & Hambleton, 2006; Lu, 2010; Luecht, Brumfield & Breithaupt, 2006;

Wang, 2013; Wang, 2017; Yang, 2016; Zheng, 2014). Yukarıdan-aşağıya test birleştirme yönteminin ise kullanıldığı birkaç çalışmaya rastlanmıştır (Davis & Dodd, 2003; Lynn Chen, 2010; Zheng, Nozawa, Zhu & Gao, 2016). Bu nedenle bu araştırma kapsamına yukarıdan-aşağıya test birleştirme yönteminin de dahil olmasının, uygulayıcılar ve araştırmacılar için ayrı bir katkı sunacağı düşünülmektedir.

Bu araştırma ülkemizde yapılan çalışmalar açısından da bir ilki bünyesinde barındırmaktadır. BÇAT'ın gelişimi 1970'li yıllara dayanırken, 2000'li yıllarda birkaç çalışma bulunmakta ve 2010 yılından itibaren çalışma sayısı giderek artmaktadır. Yurtdışındaki birçok geniş ölçekli testlerde BBT yerine, BÇAT kullanılmaya başlanmıştır. Bu testlere Lisansüstü Eğitim Sınavı (Graduate Record Examination -GRE), Hukuk Fakültelerine Kabul Sınavı (Law School Admission Test -LSAT), Ulusal Eğitimdeki Gelişmelerin Değerlendirilmesi (The National Assessment of Educational Progress -NAEP), ABD'de Tıp Alanında Uzmanlık Sınavı (The U.S. Medical Licensure Examination -USMLE), Mali Müşavirlik Sınavı (Uniform Certified Public Accountant Examination -CPA) örnek olarak verilebilir. BÇAT ile ilgili bir çalışmaya Türkiye'deki alanyazında rastlanmamıştır. Bu kapsamda ülkemizde de BÇAT ile ilgili uygulamaya dönük çalışmaların yapılmasının gerektiği ve bu konuda bir boşluk olduğu düşünülmektedir. Özellikle lise ve üniversiteye girmek, devlet kurumlarına kabul edilmek için gerçekleştirilen geniş ölçekli testlerin yüksek düzeyde geçerli ve güvenilir olması, daha fazla psikometrik nitelikleri bünyesinde barındırması için yurtdışında olduğu gibi ülkemizde de ileriki zamanlarda kullanılacak test türünün BÇAT olacağı öngörülmektedir.

Araştırma kapsamında PISA 2015 yılı uygulamasına ait veriler kullanılmış ve bu verilerden kestirilen parametrelere göre simülasyon çalışması yapılmıştır. PISA 2015 yılı uygulamasının gerek geçerlik ve güvenilirliği test edilmiş uluslararası bir sınav olması, gerekse bilgisayara dayalı gerçekleştirilen bu uygulamanın gelecek yıllarda kullanılması öngörülen BÇAT'ın temellerini oluşturması araştırmada PISA verilerinin tercih edilme nedenlerindedir. Böylece gerçek veriye dayalı olarak bir simülasyon (post-hoc simulation) çalışması gerçekleştirilmiştir. Bu durum, araştırmayı güçlü kılan önemli özelliklerden biridir. Bu doğrultuda araştırmada BÇAT'ın uygulanabilirliği araştırılmıştır. Ayrıca araştırmada, BÇAT'ın birkaç değişkeni ele alınmıştır. Bu değişkenler modül uzunluğu, panel deseni ve

örneklem büyüklüğüdür. Çünkü BÇAT uygulamasında ölçme kesinliği modül uzunluğu ve panel deseninden etkilenebilir (Zenisky & Hambleton, 2014). Araştırmanın bir başka önemli özelliği de sınıflandırma doğruluklarının incelenmesidir. Bu kapsamda araştırma değişkenlerine göre oluşturulan BÇAT'lar, geçme-kalma kararlarının sınıflandırılması açısından değerlendirilmiştir. Bu doğrultuda araştırmanın, kuramsal çalışmalara da katkıda bulunması beklenmektedir. Diğer yandan Luecht ve Nungester'in (1998) çalışmasından günümüze kadar ki zaman diliminde yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemlerini karşılaştıran bir çalışmaya rastlanmamıştır. Dolayısıyla araştırma sonuçlarının BÇAT çalışmalarına temel oluşturacağı, ülkemizde yapılacak olan bir BÇAT uygulamasına katkıda bulunabileceği öngörülmektedir.

Araştırma Problemi

PISA 2015 yılı fen okuryazarlıklarının Bireyselleştirilmiş Çok Aşamalı Test uygulamasında test birleştirme yöntemleri (yukarıdan-aşağıya ve aşağıdan-yukarıya) ile elde edilen yetenek kestiriminin kesinliği ve sınıflandırma kararının doğruluğu modül uzunlukları (6 ve 12), panel desenleri ("1-2", "1-2-2" ve "1-2-3") ve örneklem büyüklüklerine (250 ve 2000) göre nasıl bir değişim göstermektedir?

Alt problemler

1. BÇAT uygulamasında yukarıdan-aşağıya test birleştirme yönteminde modül uzunlukları (6 ve 12), panel desenleri ("1-2", "1-2-2" ve "1-2-3") ve örneklem büyüklüklerine (250 ve 2000) göre

- a) yetenek kestiriminin kesinliği ve
- b) sınıflandırma kararının doğruluğu

nasıl bir değişim göstermektedir?

2. BÇAT uygulamasında aşağıdan-yukarıya test birleştirme yönteminde modül uzunlukları (6 ve 12), panel desenleri ("1-2", "1-2-2" ve "1-2-3") ve örneklem büyüklüklerine (250 ve 2000) göre

- a) yetenek kestiriminin kesinliği ve
- b) sınıflandırma kararının doğruluğu

nasıl bir değişim göstermektedir?

Sayıtlılar

1. Araştırmada içerik dengeleme ve madde kullanım sıklığı göz ardı edilmiştir.

Sınırlılıklar

1. Araştırma, test birleştirme yöntemlerinden yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemleri ile sınırlıdır.
2. Araştırma, 2015 yılında uygulanan PISA fen okuryazarlığı testinde yer alan 1-0 şeklinde puanlanan 159 maddeye ilişkin madde ve yetenek parametreleri ile sınırlıdır.

Tanımlar

Modül: Birden fazla maddenin günlük düzeylerine göre biraraya gelerek oluşturduğu madde grubudur.

Aşama: İçerisinde en az bir modülü bulunduran basamaktır.

Panel: Birkaç aşamanın bir araya gelerek oluşturduğu desendir.

Yol: Bireyin panel içerisindeki aşamalar ve modüller arasında izleyeceği sıradır.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu araştırmanın kuramsal temeli, "Madde Tepki Kuramı", "Bireyselleştirilmiş Çok Aşamalı Test" ve "PISA" olmak üzere üç farklı konu başlığı çerçevesinde yapılandırılmıştır. Her bir konu içeriği, genel hatlarıyla açıklanmıştır. Ardından bu kapsamda alanyazında yer alan ilgili araştırmalara değinilmiştir.

Madde Tepki Kuramı

Madde Tepki Kuramı (MTK), bir bireyin test maddesine verdiği yanıt ve maddelerin özelliklerine göre tahmini yeteneğini kestirmek için geliştirilen bir modele dayalı ölçme yaklaşımıdır (de Ayala, 2009; Lord, 1980). MTK'ya göre oluşturulan bir ölçme aracında madde ve bireyler ortak bir ölçeğe yerleştirilirler. Böylece bir maddenin güçlüğünü gösteren ölçek, sınava giren bir bireyin yeteneğini kestirmek için kullanılan ölçek ile aynıdır (Kelecioğlu, 2001). Bu durum, bireylerin yetenek kestirimlerine göre sınav sonuçlarının karşılaştırılmasına izin verir (Embretson & Reise, 2000).

MTK'da bireyin gizil yetenek özelliği (θ), test maddelerine verilen yanıtlara göre kestirilir. Bu gizil özelliklerin kestirimi, aynı ve farklı yetenek düzeyindeki bireylerin sayısından ve uygulanan test maddelerinden bağımsızdır. Dolayısıyla, bir test önceki test yönetimlerinden kalibre edilmiş madde parametrelerini bulunduran bir madde setini içerdiğinde, yetenek düzeyi sıfır olan bir birey bir sonraki test uygulamasında yetenek düzeyi sıfır olan bir başka bireyin performansı ile benzer performans sergileyecektir. Bu durumun nedeni madde parametrelerinin birey gruplarından bağımsız hesaplanabilmesi olarak açıklanır (Lord, 1980).

MTK'nın *tek boyutluluk* ve *yerel bağımsızlık* olmak üzere iki temel varsayımı bulunmaktadır. *Tek boyutluluk varsayımı*, testin tek boyutlu olması, yani testi oluşturan maddelerin, yalnızca tek bir yeteneği ölçmesidir (Hambleton, Swaminathan & Rogers, 1991). Böylece bir maddenin doğru yanıtlanabilme olasılığı yalnızca madde parametresi ve birey yeteneğine bağlıdır. *Yerel bağımsızlık varsayımı* ise aynı yetenek düzeylerinde testi oluşturan maddelerin birbirinden bağımsız olmasıdır (Lord & Novick, 1968). Yerel madde varsayımı karşılanmadığında bazı maddelerin performansı θ yeteneğinden başka bir özelliğe bağlı olacaktır (Lord, 1980).

MTK, iki kategorili puanlanan bir maddenin doğru yanıt olasılığını modellemek için çeşitli olasılık modelleri kullanır. Madde çok kategorili puanlandırıldıysa da farklı madde yanıt düzeylerini modellemek için kullanılır. Bu olasılıklar, madde ve birey parametrelerine bağlı olarak değişir. Maddeler, çoktan seçmeli gibi iki kategorili olarak puanlandığında; en yaygın MTK modelleri tek parametrelili lojistik (1 PL, Rasch, 1960), iki parametrelili lojistik (2 PL, Birnbaum, 1968; de Ayala, 2009; Lord, 1980) ve üç parametrelili lojistik (3 PL, Birnbaum, 1968; de Ayala, 2009; Lord, 1980) modellerdir. θ yetenek düzeyindeki bir bireyin j maddesine verdiği doğru yanıt olasılığı madde tepki fonksiyonuna göre eşitlik 2.1, 2.2 ve 2.3'te yer alan fonksiyonlar ile ifade edilir:

$$1 \text{ PL} \quad P_j(\theta) = \frac{1}{1+\exp[-(\theta-b_j)]} \quad (2.1)$$

$$2 \text{ PL} \quad P_j(\theta) = \frac{1}{1+\exp[-a_j(\theta-b_j)]} \quad (2.2)$$

$$3 \text{ PL} \quad P_j(\theta) = \frac{1-c_j}{1+\exp[-a_j(\theta-b_j)]} \quad (2.3)$$

Yukarıdaki eşitlik 2.1, 2.2 ve 2.3'te $P_j(\theta)$, θ yetenek düzeyindeki bir bireyin j maddesini doğru yanıtama oranı; a_j , j maddesinin ayırt edicilik parametresi; b_j , j maddesinin güçlük parametresi; c_j , j maddesinin şans parametresidir. 1 PL model en basit, ancak en güçlü varsayımlara sahip modeldir. Tüm maddeler eşit ayırt edicilik gücüne sahiptir ve madde güçlüğü bakımından farklılaşmaktadır. Bireyin doğru yanıt verme olasılığı ile yetenek düzeyi arasındaki ilişki madde güçlüğü parametresi ile açıklanır. 1 PL modelde yetenek düzeyi arttıkça doğru yanıt olasılığı artarken, maddeler sadece madde güçlüğü açısından farklılaşır. 2 PL modelde ise tüm maddelerin madde güçlük ve ayırt edicilik parametreleri farklılaşır. 2 PL modelde bireyin doğru yanıt verme olasılığı ile yetenek düzeyi arasındaki ilişki madde güçlüğü ve ayırt edicilik parametresi ile açıklanır. 3 PL modelde bireyin doğru yanıt verme olasılığı ile yetenek düzeyi arasındaki ilişki madde güçlüğü, ayırt edicilik ve şans parametresi ile açıklanır.

Modül ve test bilgisi. Madde parametrelerinin değeri bir bireyin θ yetenek düzeyinin ölçme kesinliğini etkiler. MTK'da θ için ölçme kesinliği her yetenek düzeyinde aynı değildir. Ölçme kesinliği eşitlik 2.4'te yer alan bir maddenin bilgi fonksiyonu ile incelenir:

$$I(\theta) = \frac{P'(\theta)^2}{P(\theta)(1-P(\theta))} \quad (2.4)$$

Eşitlik 2.4'te $I(\theta)$, madde bilgi fonksiyonu; $P(\theta)$, θ yetenek düzeyindeki bir bireyin maddeye doğru yanıt verme olasılığı; $1 - P(\theta)$, θ yetenek düzeyindeki bir bireyin maddeye yanlış yanıt verme olasılığı; $P'(\theta)$, $P(\theta)$ 'nin birinci türevidir (Embretson & Reise, 2000). Madde bilgi fonksiyonu ölçülen özellik hakkında maddenin hangi düzeyde bilgi verdiğini gösteren bir fonksiyondur. Bu eşitlik, madde bilgisinin madde karakteristik eğrisinin eğimine karşılık gelen $P(\theta)$ 'nin birinci türevidirle ilişkili olduğunu gösterir. Dolayısıyla bir maddenin bilgi miktarı madde ayırt edicilik parametresiyle yakından ilişkilidir. Bir maddenin ayırt ediciliği arttıkça, madde bilgi fonksiyonunun değeri artar; madde güçlüğü θ 'ya yaklaştıkça da madde bilgisi artar (Embretson & Reise, 2000). Bu doğrultuda bireyin θ yetenek düzeyini doğru kestirmek ve bireyi bu yetenek düzeyine uygun maddelere yönlendirmek önem kazanır.

Yerel bağımsızlık varsayımından dolayı MTK'ya göre oluşturulan bir testteki maddelerin bilgi fonksiyonu bağımsızdır ve test bilgi fonksiyonunun elde edilmesi için toplanabilir. Bu fonksiyon eşitlik 2.5'deki gibidir:

$$TI(\theta) = \sum I(\theta) \quad (2.5)$$

Eşitlik 2.5'de $TI(\theta)$, test bilgi fonksiyonudur. Test bilgisi bir testin ölçme kesinliğini değerlendirmek için kullanılabilir. Bu, çeşitli yetenek düzeylerindeki yeteneğin standart hatası ($SE(\theta)$) incelenerek gerçekleştirilir. Yeteneğin standart hatası test bilgileriyle ilişkilidir ve fonksiyon eşitlik 2.6'daki gibidir:

$$SE(\theta) = \frac{1}{\sqrt{TI(\theta)}} \quad (2.6)$$

Eşitlik 2.6'dan da anlaşılacağı gibi standart hata ile test bilgi fonksiyonunun karekökü ters orantılıdır. Bu doğrultuda madde bilgi fonksiyonu değeri arttıkça, bireyin yeterlik ölçüsüne ilişkin hata değeri düşer. Test bilgileri ve standart hata, dolayısıyla yeteneğin ölçme kesinliği yetenek ölçeğinde sabit değildir. Birçok testte standart hatalar yetenek sürekliliğinin ortasında daha düşük olma eğilimindedir (Embretson & Reise, 2000).

Bireyselleştirilmiş Çok Aşamalı Test (BÇAT)

BÇAT, madde düzeyindeki bireyselleştirilmiş testin olumlu özelliklerinin birçoğunu içeren bir test uygulamasıdır. Bu uygulamada modül olarak adlandırılan madde setleri aracılığıyla test, bireye uyarlanır (Leucht & Nungester, 1998).

BÇAT uygulamalarına temel oluşturan fikirler oldukça eskiye dayanır. Cronbach ve Glaser (1965), Lord (1971, 1974), Weiss (1973) ve Loyd (1984), panel desenlerinin kağıt-kalem versiyonlarının kullanılmasını önermişlerdir. Bu versiyonun ilk örneği Cronbach ve Glaser (1956) tarafından personel seçimi için hazırlanan iki aşamalı örnekleme planında görülmektedir. Örnekleme planında bireylerin ilk aşamadaki performansları yeterliyse, bireylere sadece testin ikinci aşamasında verilen bir test senaryosu sunulmuştur. Bu test senaryosunun sonucunda bireyin ilk aşamadaki performansının sonraki seçimlerini engellediği tespit edilmiştir. Bu durumda da bireyi ikinci aşamaya yönlendirmeye gerek olmadığı sonucuna varılmıştır. Bu işlemler başvuranların önceden incelendiği seçme yöntemi ile benzerlik göstermektedir. Ancak bu desenlerin uygulanması BBT'nin geliştirilmesi ve uygulanması ile geri planda kalmıştır (Mead, 2006). Son yıllarda ise BÇAT'ın BBT uygulamalarının pratikteki eksikliklerine çözüm olduğu ile ilgili düşünceler bireyselleştirilmiş test çerçevesinde BÇAT'a olan ilgiyi yeniden canlandırmıştır.

BÇAT'da her bir modül doğrusal test formu olarak düşünülebilir. Her modül içerik özelliklerini karşılamak için bir araya getirilebilir ve test uygulanmadan önce gözden geçirilebilir. Bu durum, madde düzeyinde bireyselleştirilmiş test üzerinde önemli bir üstünlüktür (Zenisky, Hambleton & Luecht, 2010). Her aşamadan sonra bireylerin tüm performansı o noktaya ait yetenek kestirimini elde etmek için kullanılır. Kestirim aracılığıyla bir sonraki aşamadaki en uygun modül seçilir. Paneller birbiriyle değiştirilebilir şekilde tasarlanır.

BÇAT'da test güvenliğinin sağlanması amacıyla birden fazla panel geliştirilir. Panellerde bulunan aşama ve modüllerin sayısı aynıdır. Her bir panelin içerisinde bulunan aşamalarda yer alan modüller önceden bir araya getirilir (Luecht & Nungster, 1998). BÇAT'ın uygulanması sırasında öncelikle her birey tesadüfi olarak seçilen bir panele yönlendirilir. Her panelin ilk aşamasında orta güçlükte bir modül bulunur ve birey kendi panelinin içerisindeki ilk aşamadaki modülü yanıtlar. Bu doğrultuda ilk aşamadaki modülü yanıtlayan bireyler, kestirilen

yeteneklerine göre yeniden belirlenen yönlendirme kuralı ile bir sonraki aşamadaki modüle yönlendirilir. Bu nokta, BÇAT'da bireyselleştirmenin gerçekleştirildiği noktadır. Dolayısıyla her madde sonrasında değil, her modül sonrasında bireyselleştirilme gerçekleştirilir. Bu şekilde her modül sonrasında bireyselleştirmenin yapılması BÇAT'ın BBT'ye göre daha az bireyselleştirilmiş bir test olmasına neden olur. Ancak bir testteki maddeleri tüm bireylerin yanıtladığı kağıt-kalem testleri ile karşılaştırıldığında BÇAT, kağıt-kalem testine göre çok daha fazla bireyselleştirilmiş bir testtir (Berger, 1994). Bu doğrultuda BÇAT, kağıt-kalem testleri ve BBT arasındaki *orta yol* olarak düşünülebilir (Jodoin, Zenisky & Hambleton, 2006).

BÇAT'da test maddelerinin bulunduğu modüller önceden yapılandırıldığı için test geliştiriciler uygulamaya başlamadan önce tüm testi görebilir. Ayrıca bir modül içerisinde birey maddeler arasında ileri veya geriye atlayabilir ve önceden yanıtladığı maddelerde değişiklik yapabilir (Mead, 2006; Patsula, 1999). Modüllerin önceden yapılandırılması test yöneticisinin madde içeriği ve madde formatı ile ilgili düzenlemeler yapabilmesini sağlar (Hendrickson, 2007).

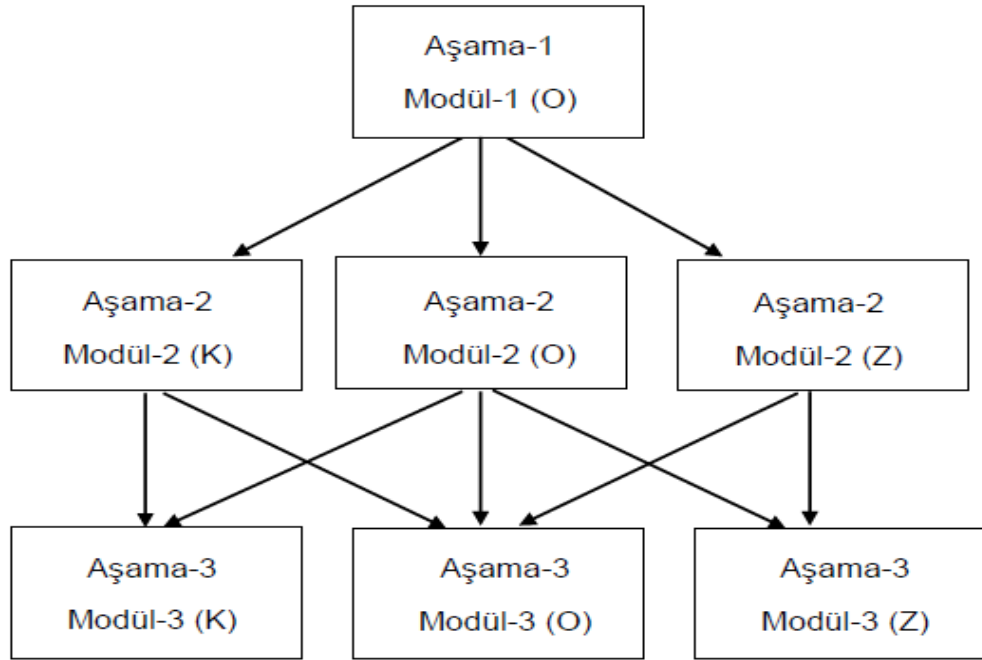
BÇAT'da bireyin test süresince izleyeceği yol ve bireye sunulacak modüller, uygulamadan önce oluşturulur. Belirli özelliklere göre modül ve panellerin test uygulamasından önce bir araya getirilmesiyle, test birleştirme (assembly) işlemi gerçekleşir.

BÇAT oluşturulurken modül, panel yapılandırılması, madde havuzu, aşama, test birleştirme, puanlama ve yönlendirme yöntemi gibi BÇAT bileşenleri ile ilgili temel ölçütler belirlenmelidir. Bu ölçütler aşağıda ayrıntılı olarak açıklanmaktadır.

Modüller. Önceden yapılandırılan madde setleri modül (module) (Luecht & Nungester, 1998) veya madde seti/takımı (testlet) (Wainer & Kiely, 1987) olarak adlandırılır. Modüller, homojen güçlük düzeyine sahip maddelerden oluşur ve farklı yetenek grupları için tasarlanır. Modüllerin ortalama madde güçlüğü gibi belirli istatistiksel özelliklere ve içerik dengesine sahip olması hedeflenir. Güçlük düzeylerine göre kolay, orta ve zor gibi kategorilere ayrılır.

Paneller. Modül ve yönlendirme kurallarının özel bir bileşimi panel olarak adlandırılır. Her panel belirlenen istatistiksel hedefleri ve içerik alanlarını kapsar. Birden fazla panel, modüllerin ve maddelerin kullanım sıklıklarını azaltmaya ve

maddelerin aşırı kullanımını engellemeye yardımcı olur (Luecht, 2003). Bu da test güvenilirliğini önemli derecede artırıcı bir etkidir. Her birey BÇAT yönetimindeki bir panele atanır. Modüllerin ve maddelerin madde kullanım sıklığını kontrol etmek için çoklu paneller, çoklu test formları gibi bireylere tesadüfi veya birleştirme yöntemi ile atanır. BÇAT’da oluşturulacak olan panel sayısına, testi alan birey sayısı, testin amacı gibi faktörlere bağlı olarak karar verilir. Genellikle panel sayısı 1 ile 40 arasında değişiklik göstermektedir (Yan, Lewis & von Davier, 2014). Şekil 2’de birinci aşamada 1 (1O), ikinci aşamada 3 (2K, 2O ve 2Z) ve üçüncü aşamada 3 (3K, 3O ve 3Z) modülü olan üç aşamalı BÇAT’ın bir paneli sunulmuştur.



K: Kolay, O: Orta, Z: Zor

Şekil 2. “1-3-3” panel deseni.

Şekil 2’deki desen “1-3-3” panel deseni olarak adlandırılır ve çeşitli çalışmalarda kullanılmıştır (Hambleton & Xing, 2006; Jodoin, Zenisky & Hambleton, 2006; Keng, 2008; Luecht & Nungester, 1998). K, O ve Z harfleri modüllerin ortalama güçlüğüne belirtir. Her satır ise belirli bir yolu temsil eder. Bireyin birinci aşamadan ikinci aşamaya yönlendirilmesine birinci aşamadaki performansına göre karar verilir. Düşük performanslı bireyler ikinci aşamadaki *kolay* modüle; orta yetenek düzeyindeki bireyler ikinci aşamadaki *orta* güçlükteki modüle ve yüksek performans gösteren bireyler ikinci aşamadaki güçlük düzeyi

zor olan modüle yönlendirilir. Üçüncü aşama için yönlendirme kuralları bireyin önceki iki aşamadaki performansına göre belirlenir. Bir sonraki aşamada bireye yöneltilecek olan modüle karar verilmesi için gerekli olan kural, yönlendirme kurallarının içerisinde inşa edilmiştir.

Şekil 1'deki panele göre bir bireyin takip edebileceği yedi olası yol; 1O-2K-3K, 1O-2K-3O, 1O-2O-3K, 1O-2O-3O, 1O-2O-3Z, 1O-2Z-3O ve 1O-2Z-3Z şeklindedir. Bu yolların tümünde madde sayısı, içerik dengesi gibi test düzeyi özellikleri bulunur. BÇAT uygulamasından önce şekildeki gibi birden fazla panel oluşturulabilir.

Madde Havuzu. BÇAT'ın psikometrik ve içerik özelliklerini karşılamak için bir araya getireceği modüllere ihtiyacı vardır. Panel desenlerinde farklı güçlükte modüller oluşturulmalıdır. Ayrıca panel deseninde modüllerin paralel olmasına da ihtiyaç duyulabilir. Bu nedenle madde havuzunun BÇAT'ın birleştirilmesine olanak sağlaması gerekir (Hendrickson, 2007).

Xing & Hambleton (2004), BÇAT'da madde havuzunun boyutu ve madde kalitesinin etkisini çalışmıştır. Çalışmalarında madde havuzunun boyutu 240 ve 480 olarak belirlenmiştir. Madde kalitesi ise havuzdaki maddelerin ayırt edicilik değerlerinin ortalaması olarak tanımlanmıştır. Çalışmada madde havuzundaki ortalama ayırt edicilik parametreleri 0.60, 1.00 ve 1.40 ölçütlerine göre değerlendirilmiştir. Bu düzeyler madde havuzunun sırasıyla zayıf, orta ve iyi düzeyde madde kalitesi olduğunu göstermiştir. Çalışmanın sonuçları 480 maddenin yer aldığı ve iyi düzeyde madde kalitesi olan madde havuzunun ölçme kesinliğini geliştirebileceğini göstermiştir. Jodoin, Zenisky & Hambleton'ın (2006) çalışmasında da 238 maddelik gerçek madde havuzunun kullanıldığı kağıt-kalem testi ile BÇAT karşılaştırılmıştır. Bu çalışmada 40 ve 60 maddeden oluşan testler bireylere uygulanmıştır. Çalışmanın sonucunda 60 maddeden oluşan her iki test türünde de doğru yetenek kestirimi ve benzer düzeyde karar doğruluğu oranları elde edilmiştir. Simülasyon ile üretilen daha büyük madde havuzunda (3222 madde) ise %30 geçme oranındaki BÇAT'ın sınıflandırma doğruluğu incelenmiştir (Zenisky, 2004). Çalışmanın sonucunda test bilgi miktarı azaldığında yanlış ve tutarsız sınıflandırma düzeyinin arttığı gözlenmiştir.

Yukarıda bahsedilen çalışmaların sonuçlarından yola çıkarak madde havuzu boyutu ve madde kalitesi bakımından zayıf madde havuzlarının BÇAT desenindeki ölçme kesinliği üzerinde doğrudan etkisi olduğundan bahsedilebilir. Ayrıca madde havuzunun, yetenek kestirimleri ve sınıflandırma kararlarının doğruluğu üzerinde etkisinin olduğu da söylenebilir.

Aşama. BÇAT deseninde bir birey, BÇAT'ın her aşamasındaki bir modülü ziyaret eder. Her aşama bir veya birkaç modül içerebilir. Teorik olarak aşamaların olası sayısı en az 1, en fazla toplam madde sayısı kadar olabilir. Ancak son yapılan çalışmalar ve uygulamalarda genellikle üç veya dört aşamanın kullanıldığı görülmektedir (Hendrickson, 2007). Aşama sayısının birden fazla olması BÇAT'ın bireye daha fazla uyarlanabilmesini sağlar. Ayrıca aşama sayısının artması panel deseninin etkililiği ve yetenek kestirimlerinin doğruluğunu da artırır (Patsula, 1999).

Her aşamadaki modüllerin sayısı. Birçok BÇAT çalışmasında ilk aşamada bir modül, sonraki aşamalarda iki veya üç modül kullanılmıştır. Patsula'ya (1999) göre ilk aşamadan sonraki aşamalarda üç modül yerine beş modülün bulunması, yetenek kestirimini artırır. Ancak yapılan birçok araştırma sonucunda bir panelin son aşamasında en fazla dört modülün kullanabileceği, üç güçlük düzeyinin kullanılmasının yeterli olabileceği görülmüştür (Armstrong, 2002; Armstrong, Jones, Koppel & Pashley, 2004).

Modüldeki madde sayısı. Alanyazında her bir modülün içerisinde 20 maddenin bulunduğu çalışmalar mevcuttur. Modüllerin uzunluğu her aşamada değişiklik gösterebilir (Hambleton & Xing, 2006; Jodoin, Zenisky & Hambleton, 2006). Bazı çalışmalarda testler birinci aşamada uzun bir modül, sonraki aşamalarda da daha kısa modüller içerir (Schnipke & Reese, 1997; Xing & Hambleton, 2004). Patsula (1999), çalışmasında toplam test uzunluğu aynı olacak şekilde her modüldeki madde sayısı dağılımının etkisini incelemiştir. Bu çalışmanın sonucunda birinci aşamadaki modülde 6, ikinci aşamadaki iki modülün her birinde 12 ve üçüncü aşamadaki beş modülün her birinde 18 maddenin bulunduğu "1-2-5" panel desenin diğer panel desenlerine göre daha doğru yetenek kestirimleri ürettiği bulgusuna ulaşılmıştır. Bir başka çalışmada da her aşamadaki modül uzunluğunun sabit bir sayı olması gerektiği önerilmiştir (Zenisky, 2004).

Test Birleştirme (Assembly). Madde havuzundan seçilen maddelerden oluşturulan test formlarının çeşitli istatistiksel özelliklere göre bir araya getirilmesi test birleştirme olarak adlandırılır. Test birleştirmenin ilk uygulamasına 1980'li yıllarda rastlanmıştır (Papadimitriou & Steiglitz, 1982; Theunissen, 1985; van der Linden & Boekkooi-Timminga, 1989). İlk test birleştirme yöntemlerinden birisi olan Eşleştirilen Rastgele Alt Test Yöntemleri (Gulliksen, 1950), Klasik Test Kuramı (KTK)'na göre oluşturulmuştur. Bu yöntemde öncelikle tüm maddeler KTK'ya göre hesaplanan madde güçlüğü ve ayırt edicilik parametrelerinden oluşan iki boyutlu bir ölçeğe yerleştirilir. Sonrasında maddeler ölçek üzerinde buldukları yere göre birbirlerine en yakın çiftleri oluşturur. Her çiftteki maddeler tesadüfi olarak iki alt teste atanır. Son yıllarda ise MTK'nın gelişimiyle birlikte test birleştirme, Test Bilgi Fonksiyonu (TBF) hesaplanarak kontrol edilmektedir. TBF, ölçmenin standart hatasının karesinin alt sınırına karşılık gelir. Lord (1977), hedef test bilgi eğrisinin altında kalan alanın doldurulması için test maddelerini sırasıyla seçen bir test birleştirme yönteminin uygulanmasını önerir.

Formların birleştirilmesi test birleştirme problemi olarak belirtilen en uygun duruma getirme (Combinatorial optimization -CO) problemi olarak formüle edilir (Papadimitriou & Steiglitz, 1982; Theunissen, 1985; van der Linden & Boekkooi-Timminga, 1989). CO, belirli fonksiyonu optimize eden sonlu bir kümenin elemanının araştırılmasıdır. CO problemi eşitlik 2.7'deki gibi formüle edilebilir:

$$\text{Maksimuma çıkarmak } \mathbf{F}(\mathbf{x}) \quad (2.7)$$

$$\mathbf{x} \in X$$

$\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, bir testi tanımlayan ikili karar vektörüdür. $x_i = 1$ olduğunda, i maddesi testte bulunur; $x_i = 0$ olduğunda ise i maddesi testte bulunmaz.

n , madde havuzundaki madde sayısıdır.

X , her biri uygun bir testi tanımlayan tüm ikili vektörleri içerir. Bu nedenle bu set *uygulanabilir set* olarak adlandırılır. Uygulamada uygulanabilir set açıkça verilmez, ancak \mathbf{x} karar vektörünü sınırlayan bir eşitsizlik ve kapsama listesiyle örtük olarak belirtilir. Bu liste doğrudan test özelliklerinden oluşur. Örneğin 5 ile 10 arasında madde içeren bir uygulanabilir set aşağıdaki eşitlik 2.8'de sunulmuştur:

$$5 \leq \sum_{i=1}^n x_i \leq 10 \quad (2.8)$$

$$x_i \in \{0,1\}$$

Bu uygulanabilir set, ikinci kısıtlamanın herhangi CO problemine dahil edildiği yerdir. Örneğin, her bir uygun çözüm $x = (x_1, x_2, \dots, x_n)^T$ bir ikili vektör olmalıdır.

$F(x)$, olası bir vektör fonksiyonu, yani hedef fonksiyondur (Veldkamp, 1999). Örneğin, θ' yetenek kestiriminde bir testin Maksimum Fisher Bilgisi eşitlik 2.9'daki fonksiyon ile hesaplanır:

$$\text{Maksimuma çıkarmak } \sum_{i=1}^n I_i(\theta') x_i \quad (2.9)$$

$I_i(\theta')$, θ' yetenek düzeyindeki i maddesinin Fisher bilgisidir (Lord, 1980).

Bu hesaplamalar yardımıyla BÇAT'da farklı güçlük düzeyindeki modüller ve panellerin 'birleştirilmesi' gerçekleştirilmektedir. BÇAT'ın geliştirilmesi sürecinde paralel BÇAT panellerinin birleştirilmesi en önemli ve en zor adımlardan birisidir. BÇAT'da test birleştirme iki adımda gerçekleştirilir: (1) madde havuzundan modüllerin birleştirilmesi ve (2) modüllerden oluşan panellerin birleştirilmesi. Bu süreçte karmaşık bir yapının olması üç aşamanın aynı anda gerçekleştirilme zorunluluğundan kaynaklanır. Bu aşamaların birincisi, bir aşamadaki farklı modüllerin test bilgi eğrilerinin geçerli bireyselleştirmeyi sağlamak için yeterince farklı olması; ikincisi, paneller boyunca karşılık gelen yolların test bilgi eğrilerinin paralel panelleri elde etmek için benzer olması; üçüncüsü, her paneldeki her yol için tüm test birleştirme sınırlamalarının (içerik dengeleme, madde kullanım sıklığı gibi) gerçekleştirilmesidir (Zheng, Wang, Culbertson & Chang, 2014). Bu aşamalar kağıt-kalem kullanılarak da hesaplanabilirken, geniş ölçekli testlerde modül uzunluğu sayısının yüksek olmasından dolayı bu işlem test geliştiricilerin otomatik test birleştirme (OTB) yöntemini tercih etmelerine neden olur (Breithaupt & Hare, 2007; Breithaupt, Ariel & Veldkamp, 2005; Luecht, 2000; Luecht, 2006; Luecht, Brumfield & Breithaupt, 2006; Luecht & Nungester, 1998; van der Linden, 2005).

Otomatik Test Birleştirme Yöntemi. Test geliştiricilerin hesaplama işlemlerini kolaylaştıran OTB bilgisayar yazılımları (örneğin; CASTISEL, ConTEST) ile otomatik test birleştirme yöntemi uygulanabilmektedir. Bu yazılım, belirli içerik veya istatistiksel hedefleri ve test içeriği ile ilgili kuralları karşılamak için optimizasyon algoritmalarını veya sezgisel yöntemleri ya da her iki yöntemi

birden uygulamak üzere tasarlanmıştır (Breithaupt & Hare, 2007; Luecht, 2000, 2006). Bu işlemler test uygulamasından önce gerçekleştirilir. Böylece test geliştiricisi işlemleri kontrol edebilir. OTB sürecinde amaç; içerik alanı, sözcük sayısı, madde türü gibi sınırlamaları karşılayan modülleri oluşturmak için madde havuzundan maddelerin seçilerek test panellerinin oluşturulmasıdır. Bu sayede madde havuzundan maddelerin modüllere seçilme süreci kolaylaşır. Bu durum, modül geliştirme sürecinin daha standart hale getirilmesini sağlar.

OTB bilgisayar yazılımı, algoritmanın geliştirilmesi için birkaç yöntem kullanır. Bu yöntemler doğrusal programlama, sezgisel yöntemler, ağ-akış işlemleri ve optimal desen yaklaşımıdır. Bu yöntemlerden ağ-akış işlemleri doğrusal programlamanın özel bir durumu olup, 1'den daha büyük bir tam sayı değer aralığını alabilen karar değişkenlerine sahiptir. Bir başka yöntem olan 'optimal desen' yaklaşımı da tahmini parametrelerinin optimal doğruluğu ile sonuçlanan deneysel bir tasarımı bulmayı amaçlayan, istatistikte geliştirilen optimal deney tasarımı teorisine dayanmaktadır. Bu nedenle, test birleştirme sırasında, optimal desen yaklaşımı, bir optimallik ölçütü verildiğinde, madde parametre değerlerinin teorik aralıklarının en iyi dağılımını hesaplar (van der Linder, 1998, 2005). BÇAT'ın test birleştirme aşamasında, aşağıda ayrıntılı olarak açıklanan doğrusal programlama ve sezgisel yöntemler tercih edilmektedir (Luecht & Nungester, 1998; van der Linder, 1998).

Doğrusal (Linear, 0-1) programlama yöntemi. Doğrusal programlama yöntemi ilk olarak Feuerman ve Weiss (1973) ve Yen (1993) tarafından önerilmiş, ancak OTB kapsamında ilk uygulaması Theunissen (1985, 1986) tarafından yapılmıştır. OTB, doğrusal programlama ile tüm test birleştirme sınırlılıklarını yerine getiren bir veya daha fazla paralel test formu için tek bir optimize edilmiş çözüm sunar. Bununla birlikte sınırlandırılmaların karmaşıklığı arttıkça, madde havuzu tüm sınırlandırmaları yerine getirmek için yeterli olmadığında sınırlandırılmaz bir optimizasyon sorununun çözülemediği ve sonuçlanamayacağı kestirilebilirlik problemi ortaya çıkabilir. Bu yöntem çoklu panellerin birleştirilmesi için en uygun çözümü sağlar.

Doğrusal programlama yöntemi ile paneller birleştirilirken (Armstrong, Jones, Koppel & Pashley 2004; Breithaupt & Hare, 2007; Luecht, Brumfield, & Breithaupt, 2006; van der Linden, 2005) içerik özellikleri gibi test kısıtlamalarının

birçoęu gerekleřtirilebilir. Bu yntem, maddelerin modllere atanması olan eřitsizlik kmesinin optimal zmlerini saęlaması aısından avantajlıdır (Adema, 1990; Armstrong, Jones, Koppel & Pashley 2004; Armstrong & Roussos, 2005; Luecht, Brumfield & Breithaupt, 2006; Luecht & Nungester, 1998). Doęrusal programlama birleřtirme yntemlerine birok BAT alıřmasında rastlanmıřtır (Adema, 1990; Armstrong, Jones, Koppel & Pashley, 2004; Armstrong & Roussos, 2005; Breithaupt & Hare, 2007; Luecht, Nungester & Hadidi, 1996).

Sezgisel (Heuristic) Yntem. Doęrusal programlama yntemleri (Adema, 1990; Armstrong, Jones, Koppel & Pashley, 2004; Breithaupt & Hare, 2007; van der Linden, 2005) tm test birleřtirme sınırlamalarını gerekleřtirerek oklu paralel panellerin birleřtirilmesi iin en uygun zm saęlar. Ancak, bu yntem karmařık test sınırlamalarının gerekleřtirilmesi iin CPLEX4 gibi pahalı ticari yazılımların kullanılmasını gerektirir. Sezgisel yntemler ise tm sınırlamaların karřılanacaęını garanti etmez, ancak ideal zme yaklařır ve bu yntem iin kullanılan bilgisayar yazılımına daha kolay eriřilebilir. Doęrusal programlama ile sezgisel programlama yntemleri karřılařtırıldıęında doęrusal programlamanın daha avantajlı olduęundan sz edilebilir.

Sezgisel test birleřtirme yntemlerinde  model kullanılır: Aęırlıklı sapma modeli (ASM; Swanson & Stocking, 1993), normalleřtirilmiř aęırlıklı mutlak sapma sezgisel (NAMSS; Luecht, 1998) ve maksimum ncelikli endeks (ME; Cheng & Chang, 2009). Alanyazına gre bu yntemlerden sadece NAMSS, BAT alıřmalarında tercih edilmiřtir (Dallas, 2014; Hambleton & Xing, 2006; Jodoin, Zenisky & Hambleton, 2006; Luecht, 2000; Luecht & Nungester, 1998; Patsula, 1999).

Luecht ve Nungester (1998), BAT panellerinin birleřtirilmesi iin iki strateji nermiřtir: Yukarıdan-ařaęıya (top-down) test birleřtirme ve ařaęıdan-yukarıya (bottom-up) test birleřtirme. Her iki strateji de ncelikle tek tek maddeleri birleřtirerek modlleri, sonrasında modlleri birleřtirerek panelleri oluřturmaya gerektirir. Ancak stratejiler arasında modllerin birleřtirilerek panellerin oluřturulması ařamasında istatistiksel farklılıklar bulunmaktadır. Ařaęıdan-yukarıya test birleřtirme stratejisi, paneller oluřturmak iin modlleri serbeste karıřtırır ve eřleřtirir. Yukarıdan-ařaęıya test birleřtirme stratejisi ise modlleri panelleri oluřturmak iin seici olarak eřleřtirmeyi gerektirir. Bu da yukarıdan-

aşağıya test birleştirme yönteminin aşağıdan-yukarıya test birleştirme yöntemine göre daha karmaşık bir yapıda olduğunun göstergesidir.

Yukarıdan-aşağıya test birleştirme stratejisi. Yukarıdan-aşağıya test birleştirme stratejisinde her aşamadaki modüller için birleştirilmiş test formları paralel değildir. Paralellik sağlanması ve istatistiksel olmayan sınırlamaların gerçekleştirilmesi için test birleştirme işlemi sonrasında panel düzeyinde ek bir optimizasyon uygulanır (Belov & Armstrong, 2005; Breithaupt & Hare, 2007; Zheng, Nozawa, Gao & Chang, 2012). Stratejiye göre, istatistiksel özellikler ve kategorik sınırlamalar (test ve madde özellikleri) yalnızca test düzeyinde sınırlandırılarak gerçekleştirilir. Bu nedenle modüller istenen test düzeyi özelliklerini üretmek için her bir panel içerisinde öngörülen yollarla birleştirilmelidir. Belirli aşamalarda yukarıdan-aşağıya test birleştirme stratejisi kullanılarak birleştirilen modüller, paneller arasında birbirinin yerine kullanılamaz. Dolayısıyla modüller paneller arasında değiştirilemez. Yukarıdan-aşağıya test birleştirme stratejisi daha fazla optimizasyon gerektirdiği için aşama düzeyindeki sınırlandırmalara kolayca bölünemeyen test düzeyinde sınırlandırmalar olduğunda bu stratejiye ihtiyaç duyulur. Kısa test uzunluğu olan bir BÇAT'ın oluşturulmasında yukarıdan-aşağıya test birleştirme stratejisinin kullanılması önerilir (Luecht & Nungester, 1998).

Aşağıdan-yukarıya test birleştirme stratejisi. Aşağıdan-yukarıya test birleştirme stratejisi, modüllerin paralel formlarını oluşturmak için uygulanır. Bir sonraki evrede bu modüller paralel paneller oluşturmak için karıştırılır ve eşleştirilir. Böylece, aşağıdan-yukarıya test birleştirme stratejisinde modüller paneller arasında yer değiştirebilir. Modüller panellerin birden fazla permütasyonunu oluşturmak için karşılaştırılabilir ve eşleştirilebilir. Her bir modül için düzensiz sınırlılıklar içeren yukarıdan-aşağıya test birleştirme stratejisine kıyasla, aşağıdan-yukarıya test birleştirme stratejisinin uygulanması daha kolay ve basittir.

Aşağıdan-yukarıya test birleştirme stratejisinin avantajı, hem istatistiksel hem de istatistiksel olmayan gereksinimleri sağlayacak şekilde tasarlanmış olmasıdır. Böylece paralel paneller oluşturmak üzere her aşamada kolayca karıştırılabilir ve eşleştirilebilir. Modüller paraleldir ve panellerin birleştirilmesi, modül formları karıştırılarak ve eşleştirilerek tamamlanabilir. Bir başka ifadeyle,

BÇAT uygulamasında bireylerin takip edecekleri yollar farklı birleştirilmiş paneller arasında paraleldir.

Her iki test birleştirme stratejisinde modüller birleştirilerek paneller oluşturulurken aşağıdaki adımlar takip edilir (Luecht & Nungester, 1998):

- a) farklı aşamalardaki test örnekleri için istatistiksel hedeflerin üretilmesi,
- b) aşamaların içerik özelliklerinin belirlenmesi,
- c) birinci ve ikinci adımların sınırlamalarını karşılayan modüllerin birleştirilmesi ile panellerin oluşturulması.

İlk adımda, test bilgi işlevleri her modül veya modüllerin birleşimi için hedef olarak kullanılır. Test bilgi fonksiyonu, yeterli kestiriminin koşullu hata varyansı karşılığıdır (Luecht & Nungester, 1998). Yukarıdan-aşağıya test birleştirme stratejisinde bir hedef bilgi fonksiyonu, bir sınav modülü boyunca belirli bir yolu temsil eden modül veya modül kombinasyonu için geliştirilebilir. Test geliştiriciler, bireylerin çoğunluğu için en olası yolları değerlendirmek ve olası her yol için bir test bilgi fonksiyonu geliştirmek isteyebilirler (Luecht & Nungester, 1998). Her olası yol için test bilgi fonksiyonunun oluşturulması yukarıdan-aşağıya test birleştirme stratejisi ile tutarlıdır. Ancak aşağıdan-yukarıya test birleştirme stratejisinde her modül için hedef bilgi işlevleri geliştirilir. Her hedef bilgi fonksiyonu, yeterli ölçeği aralığındaki hedeflenen en büyük ölçme kesinliği olan maksimum değere ulaşır (Luecht & Nungester, 1998). Hendrickson (2007), modüller için istatistiksel hedeflerin seçimini “BÇAT deseninin tasarlanmasındaki en önemli karar” olarak tanımlar. Her modül için güçlüğün ortalama düzeyi ve ranjı belirlenmelidir. Test bilgi fonksiyonunun istenilen şekli korunurken test bilgilerinin maksimum olması amaçlanır. Yönlendirmenin veya ilk aşamanın istatistiksel özellikleri bir bütün olarak testin ölçme kesinliği ile ilgilidir. Yönlendirme yöntemi ile madde güçlükleri, yetenek dağılımının ortasındaki daha iyi ölçme kesinliği olan özel bir düzey etrafında odaklanır. Ama farklılıkların ölçüsü ikinci aşamadaki modüllerin sayısına göre değişkenlik gösterir.

Puanlama ve Yönlendirme. Madde Tepki Kuramının (MTK) gelişimi, geniş ölçekli sınavlarda bireyselleştirilmiş testlere izin verilen belirli maddeler ve güçlük düzeyinin atanması için yeteneğin kestirilmesi ile gerçekleşir. MTK'ya göre puanlamada en yüksek olabilirlik kestirim yöntemi, bireylerin tepkilerini maksimize

eden yeterlik düzeyini belirlemek için tepki örüntüsünü ve madde istatistiklerini kullanır (Luecht & Nungester, 1998).

BÇAT'ın uygulanması sırasında yapılan puanlama, yönlendirme amacıyla kullanılır. *1-0 puanlama (Number-correct) veya MTK'ya göre* puanlamada bireyin uygun modüle yönlendirileceği noktada kesme puanına karar verilmelidir. Bu kesme puanlarının belirlenmesi için iki yöntem vardır. Kesme puanı yetenek ölçeğinde θ_d olarak seçildiğinde, bu kesme puanı bireylerin birinci aşamadaki performanslarına göre B ve C modülüne yönlendirilmelerinde kullanılır. A modülü için bireyin madde tepki örüntüsü, maksimum olasılık yöntemi kullanılarak yeterlik ölçeğindeki bireyin kestirilen yeteneğini bulmak için kullanılır. Eğer bireyin yetenek kestirimi (θ), θ_d 'den daha küçükse birey B modülüne, θ_d 'den daha büyükse C modülüne yönlendirilir (Luecht & Nungester, 1998). Bireylerin hangi maddelere doğru yanıt verdiklerine bakılmaksızın aynı puana sahip bireyler aynı modüle veya aşamaya yönlendirilir (Luecht, Brumfield & Breithaupt, 2006). 1-0 puanlamada ise yeterlik ölçeğinde kesme noktası olan θ_d tekrar kullanılır.

Doğru yanıt sayısı bireyselleştirme için kullanılabilirken, gerçek puanın hesaplanması ve en son yetenek kestirimi için ölçme modeli gereklidir. Luecht, Brumfield ve Breithaupt'e (2006) göre BÇAT'da puanlama üç aşamada gerçekleşir. İlki maddelerin ayrı ayrı puanlanmasıdır. Sonrasında birey için final puanı oluşturulur ve herhangi bir geçme-kalma kararı verilebilir. Puanlamanın bu son aşamasının test sürecinde gerçekleşmesi gerekli değildir (Hendrickson, 2007). BÇAT ile ilgili araştırmalarda genellikle 3 PL model (Jodoin, Zenisky & Hambleton, 2006; Luecht & Nungester, 2000) veya çok kategorili MTK modeli kullanılmıştır (Davis & Dodd, 2003; Thissen, Steinberg & Mooney, 1989). 3 PL modelin kullanımında modül içerisindeki ve modüller arasındaki maddelerin birbirlerinden şartlı bağımsız olacağı varsayılır. Çok kategorili MTK modelinde maddeler üzerinde toplanarak artan puanlama yöntemi uygulanır.

MTK modelinin seçiminden sonra en son yetenek kestirimi için Maksimum Olabilirlik Kestirimi (MOK; Jodoin, Zenisky & Hambleton, 2006; Kim & Plake, 1993), Beklenen Sonsal Dağılım (BSD; Armstrong, 2002; Hambleton & Xing, 2006; Keng, 2008; Luecht, Brumfield & Breithaupt, 2006) ve Maksimum Sonsal Dağılım (MSD; Schnipke & Reese, 1999) yöntemlerinden biri kullanılabilir.

BÇAT'ın Üstünlükleri ve Sakıncaları. Test birleştirme, test formu kalite kontrolü, veri yönetiminin kolaylaştırılması, karmaşık puanlama ve madde seçim algoritmaları ölçütlerinde BÇAT, BBT'ye göre üstünlükler sunar (Luecht, Brumfield & Breithaupt, 2006). BÇAT, testin tasarlanması aşamasında daha fazla kontrole izin verir, maddelerin düzenlenmesi ve gözden geçirilebilmesi için formlar önceden oluşturulur (Hendrickson, 2007). Ayrıca bireyler test sırasında maddelere verdikleri yanıtları gözden geçirebilirler. BÇAT'ın bu özelliği hem testi oluşturan uzmanların hem de bireylerin test kaygısını hafifletir (Vispoel, 1998). Madde kullanım kontrolü süreci testin uygulanma aşamasında gerçekleştirilir. Böylece BÇAT kapsamında BBT'de uygulanan madde seçim yöntemleri ile birlikte kullanılan özel algoritmalara ihtiyaç duyulmaz (Georgiadou, Triantafillou & Economides, 2007). BÇAT'ın bir başka üstünlüğü yerel bağımsızlığın madde düzeyindeki BBT'ye göre BÇAT'daki modül setleri arasında daha iyi kontrol edilebilmesidir (Thissen, Steinberg & Mooney, 1989). Kısaca, BÇAT'ın üstünlükleri modül içerisindeki maddelerin gözden geçirilmesi, bireyin maddeye yanıt verme süresinin BBT ile aynı olması, seçilen maddeler üzerinde test yöneticisinin kontrolünün artması şeklinde sıralanabilir.

BÇAT'ın sakıncası ise birey yeteneğini BBT'den daha düşük doğrulukta kestirmesidir. Bu nedenle BÇAT'ın BBT ile aynı ölçme kesinliğini elde edebilmesi için daha fazla maddeye ihtiyacı vardır. Ancak BBT ile BÇAT arasında ölçme kesinliği açısından büyük oranda bir farklılık yoktur. Benzer uzunluktaki kağıt-kalem testi ile karşılaştırıldığında da BÇAT'ın ölçme kesinliği daha yüksektir (Hendrickson, 2007). Aşamaların ve her aşama içerisindeki güçlük düzeyinin sayısı kestirimlerin kesinliğini etkiler. Kağıt-kalem testi, BBT ve BÇAT'ın temel özellikleri Tablo 1'de sunulmuştur.

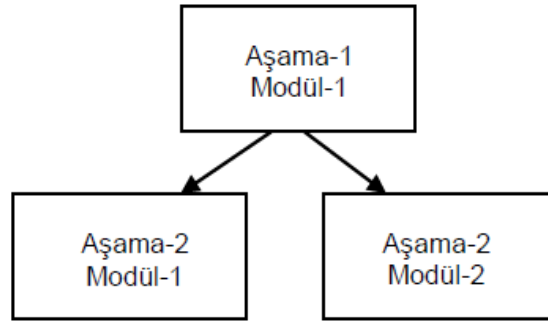
Tablo 1

Kağıt-kalem Testi, Bireyselleştirilmiş Bilgisayarlı Test ve BÇAT'ın Özellikleri

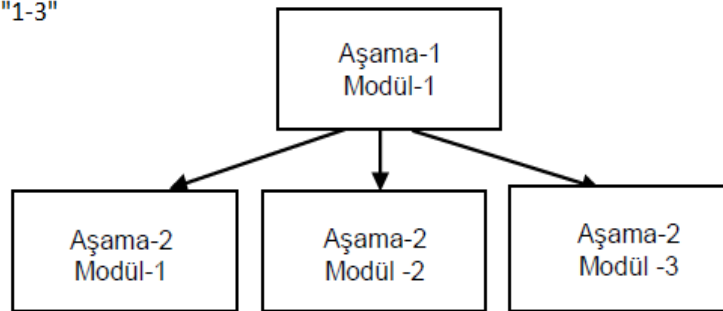
	Kağıt-kalem testi	BBT	BÇAT
Test düzeni	Aynı test	Farklı test	Aynı modül ile başlama, farklı modül ile ilerleme
Testin güçlüğü	Orta güçlükteki bireyi hedef alır	Bireyleri bireysel olarak hedef alır	Bireyleri bireysel olarak hedef alır
Test uzunluğu	Uzun-her birey için aynı uzunluk	Kısa-her birey için farklı uzunluk	Orta-her birey için aynı uzunluk
Maddeleri gözden geçirme	Yapılabilir	Yapılamaz	Yapılabilir
Test süresi	Sabit bir zaman	Öğrenci için uygun süre	Öğrenci için uygun süre
Bireyselleştirme	Yapılamaz	Madde düzeyinde	Modül düzeyinde

BÇAT yapılarının birkaç örneği ("1-2", "1-3", "1-4", "1-2-2", "1-2-3", "1-2-4", "1-3-3" ve "1-3-4") Şekil 3'te sunulmuştur:

"1-2"

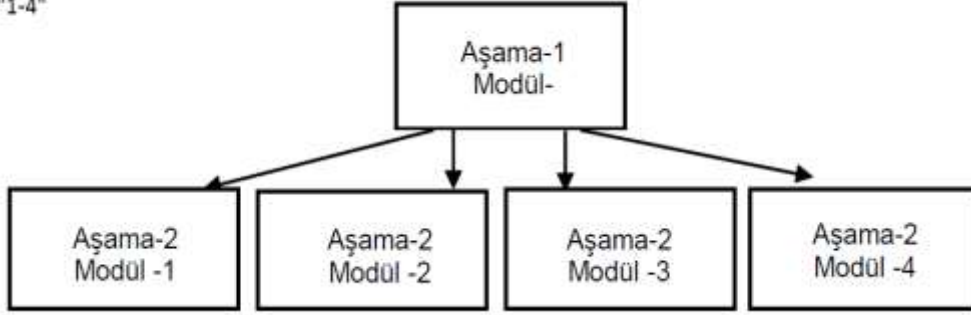


"1-3"

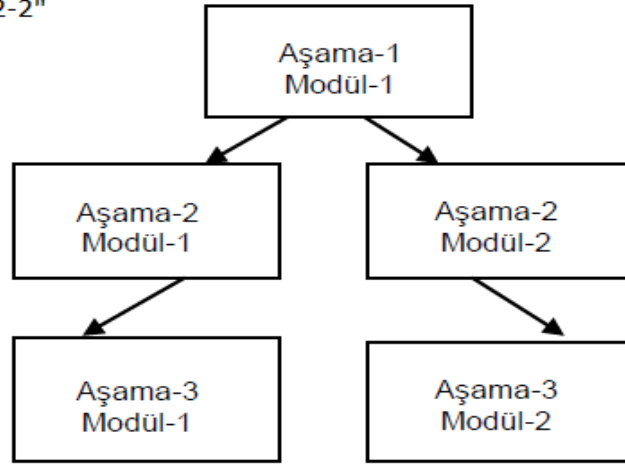


Şekil 3. Örnek BÇAT yapıları.

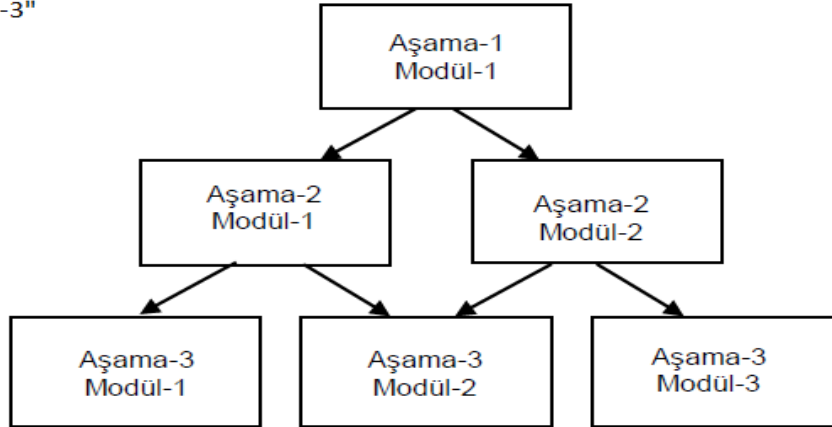
"1-4"



"1-2-2"

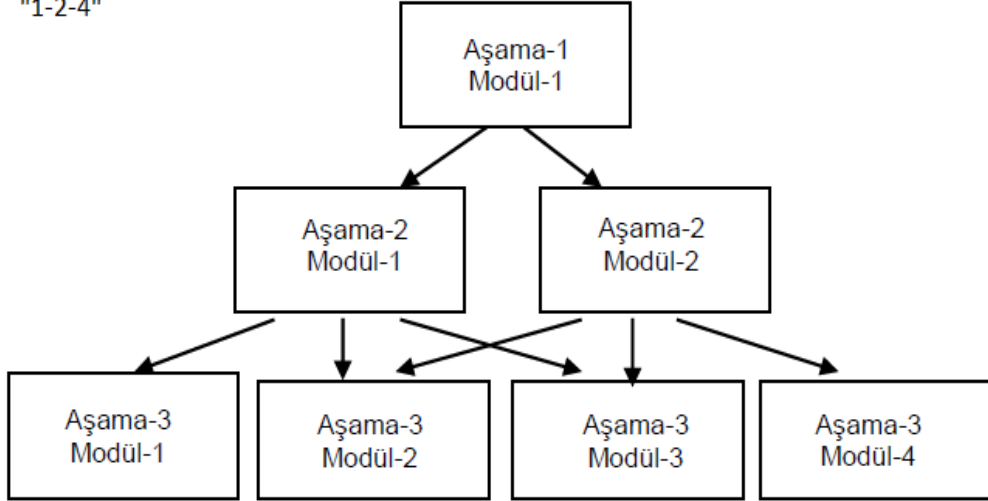


"1-2-3"

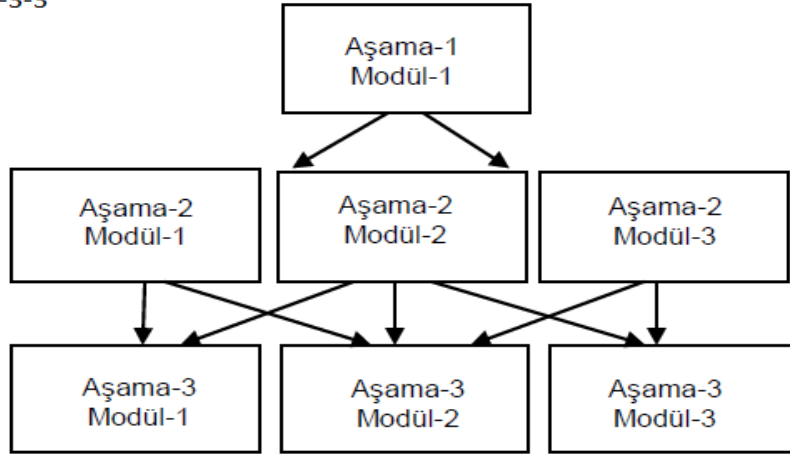


Şekil 3. Örnek BÇAT yapıları (Devam).

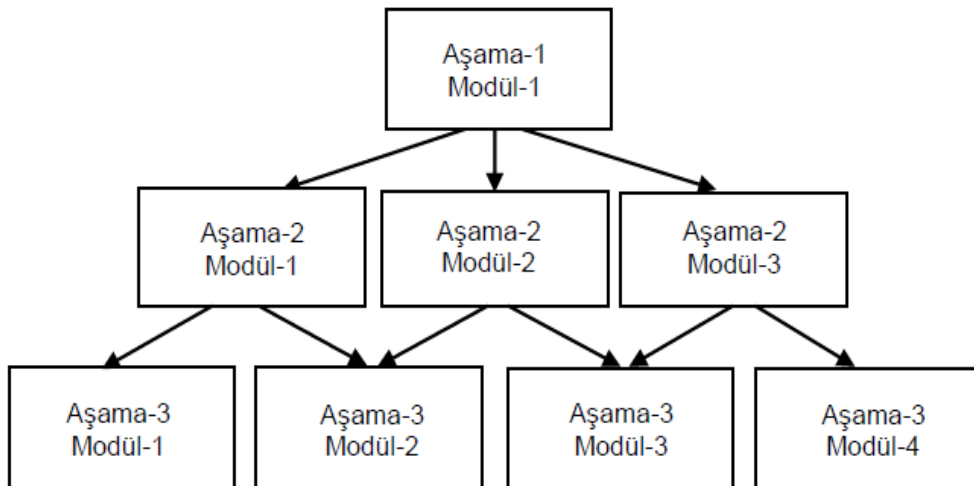
"1-2-4"



"1-3-3"



"1-3-4"



Şekil 3. Örnek BÇAT yapıları (Devam).

PISA (Program for International Student Assessment-Uluslararası Öğrenci Başarısını Belirleme Programı)

PISA, öğrencilerin zorunlu eğitimleri sonunda okulda öğrendikleri bilgi ve becerileri gerçek yaşamda kullanabilme becerilerini test etmeyi amaçlayan ve performanslarını uluslararası karşılaştırabilmeye olanak sağlayan kapsamlı bir araştırma programıdır. Bu program Ekonomik İşbirliği ve Kalkınma Örgütü (Organization of Economic Cooperation and Development -OECD) tarafından birçok ülkede uygulanmaktadır. PISA'da öğrencilerin fen okuryazarlığı, okuma becerileri ve matematik okuryazarlığı alanlarındaki konulardan elde edilen bilgi ve becerilerin ölçülmesi ve değerlendirilmesi amaçlanır. Bu doğrultuda 7. sınıf ve üzeri sınıf düzeylerinde örgün eğitime kayıtlı olan 15 yaş grubu öğrenciler, çalışmanın örneklemini oluşturur.

İlk PISA uygulaması 2000 yılında 28 üye ülke olmak üzere toplam 32 ülkenin katılımı ile gerçekleştirilmiştir. Diğer ülkeler ise ilk PISA çalışmasına, 2002 yılında tekrar uygulanan 2000 yılındaki PISA uygulaması ile katılmışlardır. İkinci PISA uygulaması; 2003 yılında 41 ülkenin katılımı ile odaklanılan alan matematik okuryazarlığı olmak üzere okuma becerileri, fen okuryazarlığı ve problem çözme becerilerinin gözlemlendiği projedir (OECD, 2005). PISA 2006 yılı uygulamasına 57 ülke, PISA 2009 yılı uygulamasına 75 ülke, PISA 2012 yılı uygulamasına 65 ülke ve PISA 2015 yılı uygulamasına 72 ülke katılmıştır.

PISA uygulamasında üçer yıllık dönemlerin her birinde, okuma becerileri, matematik ve fen okuryazarlığı alanlarından sadece birine ağırlık verilmekte; ayrıca eğitim çıktılarının, öğrencilerin başarı düzeylerindeki farklılıklarının cinsiyet yönünden ve sosyo-kültürel yönden incelenmesi de değerlendirme bölümünde yer almaktadır (MEB, 2010, s.1-8). Dokuz yıllık bir döngüde bu alanlardan her biri, bir kez temel alan olmaktadır (MEB, 2010, s.2).

PISA 2015. 2015 yılında gerçekleştirilen PISA çalışmasında odak noktası fen okuryazarlığı, yenilikçi alan ise “işbirlikçi problem çözme” olarak belirlenmiştir. Yenilikçi alan ilk kez 2012 yılında gerçekleştirilen PISA çalışmasına “yaratıcı problem çözme” alanı olarak eklenmiştir. 2012 yılından itibaren her döngüde, yenilikçi bir alanın da katkı sağlaması ile öğrencilerin temel bilgi ve becerilere ne ölçüde sahip oldukları belirlenir.

PISA arařtırmasında, öğrencilerden *bađlamsal anketler* ve *biliřsel test* aracılıđıyla veriler toplanır. *Bađlamsal anketler* öğrencilerin motivasyonları, kendileri hakkındaki görüşleri, öğrenme süreçlerine yönelik psikolojik özellikleri, okul ortamları ve aileleri ile ilgili verileri; *biliřsel test* ise fen okuryazarlıđı, okuma becerileri ve matematik okuryazarlıđı alanları ile ilgili verileri içerir. Bu alanların yanı sıra PISA 2015 uygulaması öğrencilerin finansal okuryazarlık alanını da kapsamına dahil etmiştir.

Uygulamada tüm alanların her biri için birden fazla farklı madde yanıt türü hazırlanmıştır:

- Seçeneđin işaretlenmesi
- Çoktan seçmeli maddeler için tek bir dođru seçeneđin işaretlenmesi,
- Çoktan seçmeli maddeler için birden çok dođru seçeneđin işaretlenmesi
- Karışık çoktan seçmeli maddelerin işaretlenmesi (tablo ile birlikte açıklama ve birkaç Evet/Hayır veya Dođru/Yanlıř seçenekleri içerir)
- Resmin işaretlenmesi
- Sayı giriři (sisteme sayılar, virgöl, periyot, çizgi ve ters bölü işareti girilebilir)
- Metin giriři (kađıt-kaleme dayalı maddelerde olduđu gibi öğrenci yanıtlarının uzunluklarının sınırlandırılmadıđı bir metin kutusunu içerir)
- Ařađıya dođru açılan menüden seçim yapma
- Bilgisayar imlecinin sürüklenip-bırakılması

OECD tarafından PISA 2015 uygulamasının temel alanı olan fen okuryazarlıđına ait dört üniteye yer alan maddeler kamuoyuna açıklanmıştır. Bu üniteler:

- Sürdürülebilir balık çiftliđi (3 madde)
- Kuř göçü (3 madde)
- Eđim yüzeyinin arařtırılması (2 madde)
- Meteoroid ve Kraterler (4 madde)

Bu maddelere ek olarak PISA 2015 “Pilot uygulamasında” yer alan ünitelerden ‘Sıcak havalarda koşma’ ünitesine ait 6 madde de kamuoyuna açıklanmıştır.

PISA ve BÇAT. PISA 2015 uygulaması birçok açıdan farklı bir yenilik döngüsünü başlatmıştır. Bunlardan ilki testin temel öğelerini oluşturan dağıtım türü ve değerlendirme aşamasıdır. PISA Yönetim Kurulunun rehberliğinde öncelikle ülkelerin isteklerine bağlı olarak kağıt-kaleme dayalı test türünden bilgisayara dayalı teste geçme kararı alınmıştır. Bilgisayara dayalı testi uygulayamayan az sayıda ülkeler kağıt-kaleme dayalı testi uygulamaya devam etmişlerdir. Kağıt-kaleme dayalı ölçme araçları sadece önceki döngülerde kullanılan maddeleri içerip, 2015 yılı PISA uygulaması için geliştirilen yeni maddeleri kapsamamaktadır. Bilgisayara dayalı test PISA'nın temel alanlarının yeni ve genişletilmiş yönlerini ölçmektedir. 2015 yılındaki bu yeni ve geliştirilmiş yönler fen okuryazarlığı için etkileşimli benzetim görevleri ve iş-birlikçi problem çözme için interaktif sohbete dayalı hedefleri içerir. Bu yönlerin kağıt-kaleme dayalı testler ile ölçülmesi mümkün değildir. Bir diğer önemli nokta ise, 2015 yılında uygulanmaya başlanan bilgisayara dayalı değerlendirmenin, gelecek yıllarda yapılması planlanan değerlendirmelerin temellerini oluşturmasıdır. Dijital platformun tasarımı ve geliştirilmesi, madde türleri ve temel alanlardaki madde tepki türlerinin tasarımı *Bireyselleştirilmiş Çok Aşamalı Testin* temel yapısını hazırlamaktadır. ETS, 2018 yılında gerçekleştirilecek olan PISA'da *Bireyselleştirilmiş Çok Aşamalı Test* uygulamasına geçileceğini öngörmektedir (International Consortium PISA 2015, 2014).

BÇAT ile ilgili çalışmalar

Keng (2008), madde takımı tepki modeli (Testlet Response Theory -TRT) kullanılarak ölçülen BBT ve BÇAT'ın performansını araştırmıştır. Bu amaç doğrultusunda yalnızca madde takım düzeyinde bireyselleştirilen BBT, hem madde takım hem madde düzeyinde bireyselleştirilen BBT ve BÇAT deseni karşılaştırılmıştır. Test koşulları test uzunluğu, madde havuzu büyüklüğü ve birey yetenek dağılımını içerecek şekilde manipüle edilmiştir. Üç test deseni ölçme etkinliği ve madde kullanım kontrol özelliklerine göre değerlendirilmiştir. Bu test desenlerinin benzer ve iyi ölçme doğruluğu ürettiği sonucuna ulaşılmıştır. Genel olarak madde düzeyinde BBT daha iyi ölçme kesinliği üretmiştir. Ayrıca üç test deseni madde takımı düzeyinde kabul edilebilir madde kullanım kontrol oranı üretmiştir. Yetenek dağılımındaki çarpıklık, BÇAT'ın madde kullanım kontrol özellikleri üzerinde önemli ölçüde negatif etki oluşturmuştur.

Macken-Ruiz'in (2008) çalışmasında geniş ölçekli ulusal bir sınav için simülasyon verisi kullanılmıştır. Bu çalışmada BBT ile BÇAT'ın üç deseni karşılaştırılmıştır. Her aşamadaki madde sayısı azaltılmış, artırılmış, sabit tutulmuş ve maksimum bilgi, sabit yetenek ve doğrusal programlama olmak üzere üç yönlendirme kuralı incelenmiştir. Beklenildiği gibi BBT, yeterlik kestirimi ve madde havuzu kullanımı açısından en iyi performansı göstermiştir. BÇAT'ın her aşamasında madde sayısının artması yetenek kestirimi açısından iyi bir performans göstermesini sağlamıştır. Yönlendirme kurallarından maksimum bilgi yönlendirme kuralı en iyi performansı, doğrusal puanlama yönlendirme kuralı ise en düşük performansı göstermiştir. Her BÇAT deseninde sadece bir panel oluşturulmuştur. Madde havuzunun kullanımının karşılaştırması bir panelin her BÇAT deseninde oluşturulması ile gerçekleştirilmiştir. Çalışmanın sonucunda BÇAT deseninin BBT'nin uygulanabilir bir alternatifi olduğu sonucuna ulaşılmıştır.

Chen (2010), Genelleştirilmiş Kısmi Bilgi Modeli kullanarak sekiz farklı test deseni, yönlendirme test uzunlukları (kısa ve uzun) ve toplam test uzunluklarına (kısa ve uzun) göre oluşturulan BÇAT'ın yetenek kestirimini incelemiştir. Madde havuzu ve üretilen madde tepkileri 273 kısmi bilgi madde içeren ulusal bir testten kalibre edilen maddelerden oluşur. Tüm test desenleri arasında maksimum bilgi yönlendirme yöntemi uygulanmış ve maksimum olasılık kestirimi yetenek kestirimi için kullanılmıştır. Her bir test deseninin performansı yetenek kestiriminin kesinliği,

madde kullanım sıklığı oranı, madde havuzu kullanımı ve madde örtüşmesi bakımından değerlendirilmiştir. Çalışmada tüm test desenlerinin çok benzer sonuçlar ürettiği bulunmuştur. Genel olarak önceki bulguların aksine, mevcut çalışmanın sonuçları BÇAT test yapısının yetenek kestirimini etkileyen faktör olma olasılığının Genelleştirilmiş Kısmi Puan Modeline dayalı çok kategorili puanlanan maddeler kullanıldığında daha düşük olduğunu göstermiştir.

Kim (2010) çalışmasında; BBT, BÇAT ve Sıralı Olasılık Oran Testi (SOOT) yaklaşımları arasındaki sınıflandırma kararlarının kesinliğini sistematik şekilde karşılaştırmıştır. Genelleştirilmiş kısmi puanlama modeline göre karma test formu kullanılarak üç test uzunluğu ve kesme puanı değişkenleri oluşturulmuştur. Madde kullanım (exposure) kontrol yöntemi ve sınırlandırılmış BBT içerik ağırlıklandırma yöntemi ile test ünitesi türleri de çalışmaya dahil edilmiştir. Tüm koşullar sınıflandırma kararının kesinliği ve madde kullanım kontrol yöntemi bakımından değerlendirilmiştir. Çalışmanın sonuçlarına göre üç test yaklaşımı iki kategoride bireylerin sınıflandırılması bakımından iyi performans göstermiştir. BBT ve SOOT yaklaşımları sınıflandırma kararlarının kesinliğinde az farkla BÇAT yaklaşımına göre daha iyi performans göstermiştir. Test uzunluğu arttıkça, sınıflandırma kararının doğruluğu genellikle tüm yaklaşımlarda artmıştır. Ancak BBT ve SOOT yaklaşımları daha kısa uzunluktaki testlerde daha doğru sonuçlar üretmiştir.

Lu (2010), BÇAT'daki madde takımı (testlet) maddelerinin yerel madde bağımlılığı üzerine araştırma yapmıştır. Yerel madde bağımlılığı madde takımı maddelerinde simüle edilmiştir. Sonuçlar simülasyon koşulları altında madde setinin günlük düzeyinin yetenek kestirimi ve karar doğruluğu üzerinde çok küçük etkisinin olduğunu göstermiştir. Yetenek kestirimi, madde takımı maddelerinin orta ve yüksek düzeyde yerel madde bağımlılığı olması durumunda ve kısa testlerde olumsuz olarak etkilenmiştir.

Kim, Chung, Dodd ve Park (2012), sınıflandırma testleri kapsamında karma test formatındaki BÇAT'ların çeşitli panel desenlerini karşılaştırmıştır. İlk aşamadaki test bilgi fonksiyonlarının düzeyi arttığında, simülasyonlar daha fazla sayıda doğru sınıflandırma oranı üretmiştir. Tüm BÇAT koşulları BBT'nin maksimum bilgi koşulları ile karşılaştırıldığında daha iyi sonuçlar elde etmiştir.

Wang (2013), ölçme özellikleri ve çok boyutlu BÇAT panellerinin doğruluğu ve etkililiğinde madde havuzunun (madde, modül ve panel desenlerinin sayısı) özellikleri arasındaki korelasyonun etkisini araştırmıştır. Bunun için çok boyutlu BÇAT panelleri ve onların tek boyutlu benzerleri simülasyon çalışması ile karşılaştırılmıştır. Yapı ve test uzunluğu açısından farklı olan dokuz çok boyutlu BÇAT panel deseni çeşitli korelasyon senaryoları ile simüle edilmiştir. Çalışmada farklı özelliklere sahip maddelerden madde havuzları oluşturulmuştur. Çalışmanın sonucunda çok boyutlu BÇAT'ın çok boyutlu puanlama yapıldığında tek boyutlu BÇAT benzerlerinden daha iyi performans göstermediği gözlenmiştir. Ancak çok boyutlu BÇAT desenlerinin tanılayıcı amaçlar için oluşturulacak olan BÇAT'larda kullanılabileceği belirtilmiştir. Çok boyutlu BÇAT panellerinin doğruluğu ve etkililiği madde havuzu çok daha kolay veya bilgilendirici olmadığında azalmıştır. Sonuçlara göre BÇAT uygulamaları için "1-2-3" panel deseninin kullanılması önerilmiştir.

Andrew (2014) çalışmasında, BÇAT çerçevesinde yönlendirme ve puanlama yöntemlerinin genel etkilerini araştırmıştır. Bu doğrultuda çalışmanın değişkenlerini bilginin konumu gibi madde havuzu özellikleri ve "1-3", "1-2-3", "1-2-3-4" gibi BÇAT panel desenleri oluşturmuştur. Çalışmanın sonucunda 1-0 puanlamanın her aşamadaki MTK kalibrasyonlarında kullanılabileceği, testin sonunda 3 PL model kullanılması durumunda 1-0 puanlama yönteminin geleneksel yöntemlere göre daha doğru yönlendirme yapabileceği sonucuna ulaşılmıştır.

Colvin (2014) çalışmasında, otomatik madde üretimi olduğunda, özellikle de madde setleri BÇAT'larda kullanıldığında birey yeteneğinin kestirilebilirliğini ve doğruluğunu değerlendirmiştir. Bu çalışmadaki madde setleri Sinharay ve Jonhson'nın (2008) çalışmasının sonuçlarına göre modellenmiştir. Sinharay ve Jonhson'nın araştırmasında madde setleri GRE sınavının deney bölümünde uygulanmıştır. Colvin'in çalışmasında ise daha fazla madde seti çalışmaya dahil edilmiştir. Madde setleri madde havuzunu oluşturan ana madde olarak adlandırabilen temel maddelerin birkaç özelliği değiştirilerek oluşturulmuştur. Madde setleri madde havuzundaki bu temel maddelerden büyük ölçüde çeşitlilik gösterdiğinde, bireylerin yetenekleri tahmin edildiği kadar doğru ölçülememiştir. Ancak burada olumlu sonuçların elde edildiği bazı koşulların olduğu

vurgulanmıştır. Örneğin, 600 puanlık bir ölçekte mutlak yanlılık (100 tekrardaki 'iptal etme' etkisini ortadan kaldırmak için kullanılmıştır), tüm maddeler temel maddelerinden küçük değişikliklerle bir madde seti olacak şekilde simüle edildiğinde çoğu birey için 10 puandan daha azdır. İlk aşamadaki maddeler onların temel maddelerinden orta düzeyde değişiklik olacak şekilde simüle edilmiştir. İkinci aşamadaki hiçbir madde ise madde seti olarak bir araya gelememiştir.

Hembry (2014), 3 PL Madde Takımı Tepki Teorisi (Testlet Response Theory -TRT) modelinde karma format madde takımına dayalı madde havuzu için BÇAT'ların işlevsel özelliklerini incelemiştir. Çalışmada çeşitli test koşullarında işlevsel olarak yanlılık bulunamamıştır. Tüm koşullarda BÇAT, yeterli ölçme kesinliği göstermiştir. Ancak test koşullarının bazıları arasında birkaç farklılık oluşmuştur. Ölçme kesinliği panel deseni ve test uzunluğu büyüklüğünden etkilenmiştir. Yönlendirme yöntemlerinin ölçme etkililiği üzerinde çok az etkisi olduğu sonucuna ulaşılmıştır.

Park, Kim, Chung ve Dodd (2014) çalışmalarında, genelleştirilmiş kısmi bilgi modeline (G-PCM) göre karma test formu kullanılarak oluşturulan BÇAT'ın yeni bir madde havuz kullanımını araştırmıştır. Çalışmanın sonucunda BÇAT, sınıflandırma kararlarının kesinliği açısından tüm test koşullarında eşit düzeyde iyi performans göstermiştir.

Park (2015) çalışmasında, karma test formatındaki BÇAT'ın ölçme kesinliğini etkileyen panel deseni faktörünü araştırmıştır. Uzun test uzunluğu yüksek ölçme kesinliği üretmiş; "1-3-3" panel deseni "1-2-2" panel deseninden daha iyi performans göstermiştir. Çok kategorili maddelerin büyük bir oranı, testin oluşturulması sırasında test bilgisi azaltıldığından dolayı düşük ölçme kesinliği üretmiştir. Genel olarak test uzunluğunun iki faktörü ve BÇAT yapısı yetenek kestiriminden etkilenmiş, çok kategorili maddelerin oranı ve yönlendirme yönteminin etkisi madde havuzuna yansımıştır.

Sari (2016), bireyselleştirilmiş testlerin içerik kontrolünü incelediği çalışmasında; bireyselleştirilmiş bilgisayarlı test ve bireyselleştirilmiş çok aşamalı testi örnek olarak ele almıştır. Çalışmada simülasyon yardımıyla farklı içerik alanları sayısının çeşitli test uzunluklarıyla farklılaştığı BBT ve BÇAT sonuçlarının kesinliği araştırılmıştır. BBT ve iki farklı BÇAT panel deseni ("1-3" ve "1-3-3"),

toplam test uzunluđu (24 ve 48 madde) ve kontrol edilen ierik alanları aısından eřitli kořullarda karřılařtırılmıřtır. alıřmanın sonuları test uzunluđu ve test uygulama modelinin ierik alanına gre BAT sonularını daha fazla etkilediđini gstermiřtir. BBT ile iki farklı yapıdaki BAT'ın arasında performans aısından anlamlı bir farklılık olmadığı, ancak BAT yapıları ile BBT karřılařtırıldıđında BBT'nin BAT yapılarına gre daha iyi bir performans gsterdiđi bulunmuřtur.

Sari, Yahsi-Sari ve Huggins-Manley (2016) tarafından Bireye Uyarlanmış ok Ařamalı testlerin kavramsal yapısı hakkında alıřma yapılmıřtır. Bu alıřma sonucunda bireye uyarlanmış ok ařamalı testlerin tm zellikleri, diđer testlerden farklılıkları, avantajları ve dezavantajlarının arařtırmacılarla paylařılması, arařtırmacıların bu alana olan ilgilerini artırması ve bu alanın geliřmesine katkı sađlamalarına teřvik etmek amalanmıřtır.

Wang (2017) alıřmasında, BBT ve BAT'ın performansını karřılařtırmıřtır. BAT ve BBT'ye ynelik tasarlanan madde havuzları farklı zelliklere sahip olduđundan alıřmada BAT ve BBT farklı madde havuzuna gre arařtırılmıř ve karřılařtırılmıřtır. Bu dođrultuda alıřmada; BAT ve BBT iin tasarlanmış farklı madde havuzlarında lmenin standart hatası, madde kullanım sıklıđı oranı, MTK puanlama yntemleri ve ierik zellikleri eřleřtirilerek; BBT ve BAT'ın lme kesinliđi ve ortalama test uzunluklarının karřılařtırılmasına odaklanılmıřtır. BAT; "1-2-3" ve "1-3-3" panel desenleri, OMB ve tanımlanmıř rneklem aralıkları (Defined Population Intervals) ynlendirme stratejileri, 45 ve 60 maddeden oluřan test uzunlukları ve test birleřtirme sırasında erken ařamadaki modllerin sonraki ařamalardaki modllerden sonra oluřturulması (ileriye dođru birleřtirme-forward assembly) veya nce oluřturulmasına (geriye dođru birleřtirme-backward assembly) gre tasarlanmıřtır. alıřmada ařađıdan-yukarıya test birleřtirme yntemi kullanılmıřtır. alıřmanın sonuları BBT ve BAT arasında benzer lme kesinliđi olduđunu gstermiřtir. Bunun nedeni her iki test trnn de benzer kořullara sahip test bilgilerine gre karřılařtırılması olarak aıklanmıřtır. alıřmada bu Őekilde benzer kořullara sahip test bilgilerine gre karřılařtırma yapılmasının yetenek ve ortalama test uzunluđu aısından BBT yerine BAT'ın tercih edilmesi iin bir referans sađladıđı belirtilmiřtir. Geriye dođru birleřmenin gerekleřtirildiđi BAT, sınıflandırma odaklı test iin nerilmemektedir. BBT veya BAT'ın kullanılıp kullanılmamasında BBT'nin ortalama test uzunluđu gre

karar verilmesi gerektiği sonucuna ulaşılmıştır. Test uzunluğu uzun olduğunda BÇAT yerine BBT kullanılması önerilmektedir.

Otomatik Test Birleştirme İle İlgili Çalışmalar

Luecht ve Nungester (1998) çalışmasında, bilgisayarda bireyselleştirilmiş sıralı test (BBST) olarak adlandırılan test geliştirme ve yönetimine yönelik bir yaklaşım açıklamıştır. BBST, test geliştiricilerinin üretim, kalite ve bilgisayarda uygulanan farklı test türlerinin yönetimi üzerinde daha büyük bir kontrol derecesi elde etmesini sağlamak için otomatik test birleştirme ile bireyselleştirilmiş test yöntemlerini içeren yapılandırılmış bir yaklaşım olarak tanımlanmıştır. Çalışmada, BBST çerçevesi ayrıntılı bir şekilde açıklanmış ve tıbbi uzmanlık örneği kullanılarak birçok uygulama gösterilmiştir. BBST'nin sadece BBT'nin yerine geçmesi için kullanılacak bir test olmadığı, yüksek kalitede bilgisayar kullanılarak gerçekleştirilen testlerin seri üretimleri hakkında farklı bir felsefeyi temsil ettikleri vurgulanmıştır.

Luecht ve Nungester (2000), bilgisayar ile uygulanan BBST'lerin (Luecht & Nungester, 1998; Luecht, Nungester & Hadadi, 1996) büyük ölçekli üretimi ve yönetimini açıklayan kavramsal bir çalışma yürütmüştür. BBSTy'i, test tasarımını, birleşimini, yönetimini ve veri yönetimi bileşenlerini, zaman içinde güvenli, yüksek kaliteli, paralel test formlarının seri üretimini desteklemek üzere kapsamlı bir şekilde entegre eden test türü olarak tanımlamıştır. Bu kavramın modern psikometrik ve bilgisayar teknolojilerini kullanan testlere modüler bir yaklaşım sunduğu belirtilmiştir. BBST'nin panel, aşama, modül ve izlenecek olan yol olmak üzere dört bileşenden oluştuğu ve BBST panellerini birleştiren test birleştirme yöntemleri açıklanmıştır.

Breithaupt ve Hare'in (2007) çalışmasında, BÇAT'ın birleştirilmesi ile ilgili bir vaka çalışması yürütülmüştür. Bilgisayar tabanlı lisans sınavı için test birleştirme problemlerini test eden uygulanabilir bir yol sunulmuştur. Bu vaka çalışmasıyla yüksek nitelikli sınav programları için gerekli olan psikometrik özellikler belirtilerek, maddelerin kullanım sıklığını azaltan etkin bir otomatik test birleştirme yöntemi ve özellikleri açıklanmıştır. Birincil hedef, test içeriğinin paralellik düzeyini ve istatistiksel özelliklerini zaman içinde sağlayarak, test süresi içerisinde test maddelerinin kabul edilebilir ve uygun bir şekilde bulunmasını sağlayan otomatik eş zamanlı test birleştirmenin yapılması olarak açıklanmıştır. Bu

vaka çalışmasının geliştirilmesiyle aynı metodoloji farklı psikometrik modeller ve içerik bileşimi gerektiren yapılandırılmış testlerde başarıyla kullanılmıştır. Çalışmanın sonucunda bu vaka çalışmasının; var olan yazılımın, test problemleri veya test formları için istenen psikometrik özellikler korunarak, test birleştirme probleminin matematiksel bir ifadesini çözmek için kullanıldığı sonucuna ulaşılmıştır.

Edmonds ve Armstrong (2009), BÇAT test birleştirme problemleri için karışık tam sayı programlama modellerini önerdikleri bir çalışma yürütmüştür. Çalışmada ticari optimizasyon yazılımı ile hesaplama sonuçları sunulmuş ve değerlendirilen modellerin avantajları incelenmiştir. Çoklu Aşamalı Bireyselleştirilmiş Test Deseninden (Multiple Adaptive Test Design-MSTD) birden çok BÇAT birleştirmek için modeller tanıtılmıştır. Çalışmada, BÇAT birleştirme modellerinden karma tam sayı programlama (KTP) problemine çözümler üretmek için ILOG CPLEX kodu kullanılmıştır. Çalışmanın sonucunda ele alınan modellerin oldukça pratik olduğu, BÇAT'ları bir araya getirme ve gözden geçirme sürecinin, kağıt-kaleme dayalı testlerin birleştirilmesi ve gözden geçirilmesinde kullanılan süreç ile benzer olduğu belirtilmiştir. BÇAT'ın uygulamadan birkaç ay önce birleştirilebileceği ve bir birim olarak önceden test edilebileceği, bunun da hem MTK'ya göre kestirilen parametreleri hem de testin geçerliğini iyileştirme potansiyeline sahip olduğu belirlenmiştir.

Zheng, Nozawa, Gao ve Chang'ın (2012) çalışmasında, otomatik sezgisel yaklaşım temelinde yukarıdan-aşağıya test birleştirme yöntemine göre oluşturulan BÇAT; kağıt-kalem testi ve BBT ile karşılaştırılmış ve sınıflandırma doğruluğu incelenmiştir. Çalışmanın veri setini gerçek madde havuzundan seçilen 600 çoktan seçmeli madde oluşturmuştur. BÇAT ve BBT için toplam test uzunluğu 21 madde, kağıt-kalem testleri için toplam test uzunluğu 30 madde olarak belirlenmiştir. Çalışmada BÇAT'ın farklı panel desenleri, modül uzunlukları ve yönlendirme yöntemleri karşılaştırılmıştır. Üç ve dört aşamalı panel desenleri, farklı modül uzunlukları ve yönlendirme yöntemleri arasında anlamlı bir farklılık bulunamamıştır. Ayrıca çalışmanın sonucunda BÇAT'ın kağıt-kalem testi ve BBT'ye göre madde havuzunu daha etkili kullandığı ve daha doğru bir sınıflandırma yaptığı sonucuna ulaşılmıştır.

Zheng (2014), BBT ve BÇAT'ın avantajlarını birleştiren ve sınırlamalarını dengeleyen yeni bir uyarlanabilir test deseni olan “test sırasında test birleştirmenin yapıldığı çok aşamalı bireyselleştirilmiş testin (TSBÇAT)” incelenmesinin amaçladığı bir çalışma yürütmüştür. TSBÇAT, BÇAT ve BBT'nin karma bir oluşumdur. Çalışmada, TSBÇAT'nin hem BÇAT hem BBT ile karşılaştırıldığı simülasyonlar oluşturulmuştur. Sonuç olarak, bu çalışmada TSBÇAT'ın BÇAT ve BBT'ye göre avantajları olduğu belirtilmiş ve TSBÇAT'ın gelecekteki test çalışmalarında kullanılabilecek özellikleri sunulmuştur.

Tay (2015) tarafından yapılan bir çalışmada, yenilikçi bir test tasarımı olan test sırasında test birleştirme yöntemine göre oluşturulan bireyselleştirilmiş çok aşamalı test (TSBÇAT) incelenmiştir. Bu çalışmada hem yetenek tahmini hem de beceriyi ölçen testler (mastery test) için iki yeni yenilikçi BÇAT tasarımı önerilmiştir. Simülasyonlar, önerilen test sırasında test birleştirilmesinin yapıldığı BÇAT'ın ve karma BÇAT tasarımlarının BBT tarafından karşı karşıya kalınan düşük ve aşırı tahmin problemlerinin üstesinden gelebildiğini göstermiştir. Ayrıca standart BÇAT'ın geleneksel BBT ve test sırasında test birleştirilmesinin yapıldığı BÇAT tasarımları ile kıyaslandığında test güvenilirliğini sağladığı görülmüştür.

Belov (2016) tarafından hazırlanan Hukuk Fakültesine Kabul Komisyonu Araştırma Raporunca Hukuk Fakültesi Kabul Testinin bilgisayarlaştırılmış ve çok aşamalı olması durumundaki olası uygulamaları tartışılmış, test birleştirme yöntemlerinin geliştirilmesinde uygulanan başlıca yaklaşımlar gözden geçirilmiştir. Çalışmada, karma tam sayı programlama (KTP) ve tek biçimli test birleştirme (uniform test assembly) yaklaşımları başarılı bir şekilde uygulanmıştır. KTP yaklaşımının, testin kısıtlamalar ve optimize edilecek nesnel bir işlevle tanımlandığı programların test edilmesi; tek biçimli test birleştirme yaklaşımının, testin yalnızca kısıtlamalarla tanımlandığı programların test edilmesi için uygulanabilecek bir yöntem olduğu sonucuna ulaşılmıştır.

Zheng, Nozawa, Zhu ve Gao (2016) tarafından yapılan bir çalışmada, BÇAT'ın oluşturulmasında kullanılan yukarıdan-aşağıya test birleştirme yöntemi incelenmiştir. Çalışmada üç aşamalı “1-2-4” ve dört aşamalı “1-2-3-4” panel desenleri; 15, 16 ve 21 maddeden oluşan test uzunlukları çalışılmıştır. Ayrıca örtüşmenin olmadığı, yani bir maddenin yalnızca bir modül formunda maksimum görünmesine izin verildiği ve örtüşmenin olduğu, yani bir maddenin iki farklı modül

formunda görülmesine izin verildiği iki madde örtüşme koşulu ele alınmıştır. Çalışmanın sonucunda üç ve dört aşamalı panel desenlerinin modül bilgi eğrileri arasında fark olmadığı, test birleştirme sonuçlarının BÇAT'ın tüm panellerinde tutarlı olduğu, sezgisel algoritmanın içerik kısıtlamalarını kontrol etmede genellikle etkili olduğu bulgularına ulaşılmıştır. Bu bulgulara ek olarak BÇAT'ın doğrusal testlere göre daha avantajlı olduğu, çakışan koşulların madde havuzunun kullanımından daha az yararlandığı, TBF hedefleri madde havuzundan optimize edildiğinde daha fazla ölçme kesinliği ürettiği belirtilmiştir.

İlgili Araştırmalar Özet

Alanyazın incelendiğinde BÇAT ile ilgili yurt dışında yapılmış birçok çalışma bulunmasına karşın, ülkemizde bir tane kuramsal çalışmanın bulunduğu görülmektedir. Bahsedilen araştırmalarda ilk yıllarda BÇAT'ın BBT ile karşılaştırılması yapılmış (Keng, 2008; Kim, 2010; Macken-Ruiz, 2008), sonraki yıllarda ise BÇAT'ın psikometrik özellikleri araştırılmıştır (Colvin, 2014; Hembry, 2014; Park, 2015). Bu özellikler madde kullanım kontrol özellikleri, madde havuzu kullanımı ve test sırasında test birleştirmenin yapılmasıdır. Odak noktası test birleştirme yöntemi olan araştırmalar genel olarak incelendiğinde bu çalışmalar arasında test birleştirme yöntemlerinden yukarıdan-aşağıya test birleştirme yönteminin ve test sırasında gerçekleştirilen test birleştirme yönteminin araştırıldığı çalışmaların olduğu görülmektedir. Aşağıdan-yukarıya test birleştirme yöntemine göre yukarıdan-aşağıya test birleştirme yönteminin daha az tercih edilmesi ve kısa testlerin uygulanmasında daha avantajlı olması, yukarıdan-aşağıya test birleştirme yöntemini araştırmaların odak noktası haline getirmiştir. Yine otomatik test birleştirme yöntemi ile uygulanan test sırasında test birleştirme yöntemiyle de yeni bir BÇAT deseni oluşturulmuş ve gelecekte BÇAT ve BBT'nin yerine kullanılacağı belirtilmiştir. Ayrıca araştırmalarda panel deseni, modül uzunluğu ve aşama sayısı gibi BÇAT'ın oluşmasını sağlayan ögelerin karşılaştırılması yapılmıştır.

Bölüm 3

Yöntem

Bu bölümde araştırmanın yöntemine ilişkin bilgiler bulunmaktadır. Bu bölüm araştırma yöntemi, çalışma grubu ve verilerin analizi olmak üzere üç bölümden oluşmaktadır. Araştırma yöntemi bölümünde araştırmanın niteliği ile ilgili bilgiler yer almaktadır. Çalışma grubu bölümünde 2015 yılında uygulanan PISA'ya ait genel bilgilere ve bu uygulamaya katılan öğrencilerin bu çalışmada yer alma nedenlerine değinilmiştir. Son bölüm olan verilerin analizi bölümünde ise MTK'ya dayalı madde analizi ve post-hoc simülasyon yöntemine ilişkin BÇAT değişkenleri açıklanmıştır.

Araştırma Modeli

Bu araştırmada, 2015 yılında gerçekleştirilen PISA uygulamasına katılan bireylerin fen okuryazarlığı yeteneklerinin MTK'ya dayalı olarak kestirilmesi ve post-hoc simülasyon yöntemi ile farklı özelliklere sahip BÇAT desenlerine göre test birleştirme yöntemlerinin performanslarının karşılaştırılması amaçlanmıştır. Post-hoc simülasyonunda gerçek maddelerden elde edilen yanıtlara göre eğer bu öğrenciler bu maddelere dayalı olarak oluşturulan bir BÇAT'ı yanıtlasalar ne olurdu?" sorusuna yanıt aranır (Nydyck & Weiss, 2009). Araştırmada tek boyutlu MTK'ya dayalı olarak oluşturulan BÇAT'dan elde edilen ortalama hata, yanlışlık ve gözlenen kestirimler ile gerçek parametrelerin korelasyonları incelenmiştir. Böylece yetenek kestiriminin kesinliği (precision of measurement) ve sınıflandırma kararlarının doğruluğu bilgilerine ulaşılabacağından araştırma, simülasyona dayalı betimsel araştırma niteliği taşımaktadır.

Çalışma Grubu

Araştırmanın çalışma grubunu PISA 2015 uygulamasına katılan bireyler oluşturmaktadır. PISA 2015 uygulamasına 35'i OECD üyesi olmak üzere 72 ülkeden yaklaşık 29 milyon öğrenciyi temsilen 540.000'e yakın öğrenci katılmıştır (OECD, 2015).

PISA 2015 yılı uygulamasında "kağıt-kaleme dayalı test (paper based assessment)" ve "bilgisayara dayalı test (computer based assessment)" olmak üzere iki test türü kullanılmıştır. Kağıt-kaleme dayalı testi tercih eden ülke sayısı 15, bilgisayara dayalı testi tercih eden ülke sayısı 57'dir. Türkiye, PISA uygulamasını bilgisayara dayalı olarak gerçekleştirmiştir. PISA'nın kapsamını

okuma-yazma becerileri, matematik okuryazarlığı ve fen okuryazarlığı olmak üzere üç temel alan oluşturmaktadır. 2015 yılında gerçekleştirilen uygulamada ağırlıklı alan fen okuryazarlığı olarak belirlenmiş, bu alanda yeni maddeler geliştirilmiş ve bu maddeler sadece bilgisayara dayalı testin kapsamında yer almıştır. Her öğrenci için test süresi iki saattir ve test 30 dakikalık dört bölümden oluşmaktadır. Kağıt-kaleme dayalı test için 30 farklı kitapçık (Form 1-30), bilgisayara dayalı test için 66 farklı kitapçık (Form 31-96) oluşturulmuştur. Özel gereksinimleri olan öğrenciler için de “UH (Une Heure)” kitapçığı (Form 99) hazırlanmıştır. UH kitapçığındaki maddeler testin her alanında bulunan kolay güçlük düzeyindeki ve çok daha kısa sürede yanıtlanabilen maddelerden seçilmiştir (OECD, 2015).

Araştırmada BÇAT'ın yapısal özelliklerinin daha uygun olması nedeniyle çalışma grubunu, yukarıda da bahsedildiği gibi 2015 yılında gerçekleştirilen PISA uygulamasında fen okuryazarlığı alanına ait bilgisayara dayalı testin 91 numaralı kitapçığını alan tüm bireyler oluşturmaktadır. Fen okuryazarlığı alanının seçilmesinin nedeni, 2015 yılı PISA uygulamasının ağırlıklı alanını oluşturmasıdır. Bilgisayara dayalı teste göre oluşturulan toplam 66 kitapçıktan (Form 31-96), fen okuryazarlığına ait madde ve kitapçığı alan birey sayısının diğer kitapçıklara göre daha fazla olması nedeniyle, 91 numaralı kitapçık çalışmanın veri toplama aracı olarak belirlenmiştir. Bu kitapçıkta fen okuryazarlığı alanında çeşitli kategorilere (iki kategorili, çok kategorili, açık uçlu gibi) sahip toplam 501 madde yer almaktadır. Çalışmanın madde havuzunu ise bu kitapçıkta yer alan iki kategorili maddeler oluşturmaktadır. Bu bağlamda, 91 numaralı kitapçıkta bulunan toplam 501 madde içerisinde iki kategorili olan 159 madde, çalışmanın madde havuzuna seçilmiştir. Bu iki kategorili maddeleri yanıtlayan 15059 öğrenciye ilişkin veri seti üzerinden analizler gerçekleştirilmiştir.

Verilerin Analizi

Çalışma verilerinin analizi iki aşamada gerçekleştirilmiştir. Birinci aşamada 2015 yılında uygulanan PISA sınavına katılan öğrencilerden oluşturulan çalışma grubundan elde edilen veriler MTK'ya dayalı olarak 2 PL modele göre analiz edilmiş ve BÇAT için madde havuzu oluşturulmuştur. İkinci aşamada ise post-hoc simülasyon yöntemi kullanılarak birinci aşamada oluşturulan madde havuzuna göre BÇAT yapısı oluşturularak, çeşitli BÇAT yapılarına ilişkin analiz sonuçları elde edilmiştir.

Verilerin Analizinin Birinci Aşaması

Bu aşamada ilk olarak 2015 yılında gerçekleştirilen PISA uygulamasından elde edilen veri setinin MTK'ya göre tek boyutluluk, yerel bağımsızlık, model-veri uyumu, madde ve yetenek parametrelerinin değişmezliği varsayımları test edilmiştir (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985; Yıldırım, Çömlekoğlu & Berberoğlu, 2003). Veri setinin tek boyutluluğu tetrakorik korelasyona dayanan faktör analitik yöntemleri kullanan Factor 10.5 programı aracılığıyla incelenmiştir. Madde parametreleri ve bireylere ilişkin yetenek parametreleri BILOG-MG (Zimowski, Muraki, Mislevy & Bock, 1996) programında kestirilmiştir.

Tek boyutluluk. Veri setinin faktör analizi için uygunluğu Bartlett's testi ve Kaiser-Meyer-Olkin (KMO) ölçütü ile test edilmiş (Bartlett's = 1584902.1, sd = 12561, $p = 0.00$; KMO = 0.98) ve veri setinin analize uygun olduğu sonucuna ulaşılmıştır. Veri setine ait açıklanan varyans oranları Tablo 2'de sunulmuştur.

Tablo 2

Bileşenlere Ait Açıklanan Varyans Oranı

Bileşen	Özdeğer	Açıklanan Varyans Oranı	Kümülatif Varyans Oranı
1	55.73	0.35	0.35
2	12.19	0.07	0.42
3	6.44	0.04	0.46
4	5.28	0.03	0.49
5	4.67	0.02	0.51

Tablo 2'de öz değeri 2'den büyük olan ilk beş bileşene ait öz değer, açıklanan varyans ve kümülatif varyans oranları yer almaktadır. Tablo 2'deki öz değerler incelendiğinde, birinci bileşenin öz değeri ile ikinci ve üçüncü bileşenlerin öz değerleri arasında büyük farklılıkların olması testin bir bileşenli olduğuna ilişkin ipucu vermektedir. İlk bileşenin öz değerinin ikinci bileşenin öz değerine oranı 4.57'dir. Bu değer verinin tek faktörlü bir yapısı olduğunun göstergesidir (Lord, 1980). İlk bileşen toplam varyansın %35'ini, ikinci bileşen %7'sini ve üçüncü bileşen %4'ünü açıklamıştır. Bu üç bileşen toplam varyansın %46'sını açıklamaktadır. EK-A'de yer alan bileşenler matrisindeki faktör yükleri de incelendiğinde veri setini oluşturan maddelerin tek faktör altında toplanabileceği sonucuna ulaşılmıştır. Bu sonuç OECD tarafından düzenlenen "PISA 2015 Teknik

Raporu” nun sonuçlarıyla da paralellik göstermiştir (OECD, 2017).

Yerel bağımsızlık. Aynı yetenek düzeylerinde testi oluşturan maddelerin istatistiksel olarak birbirinden bağımsız olması anlamına gelen yerel bağımsızlık varsayımı, tek boyutluluk ile paraleldir (Hambleton & Swaminathan, 1989; Hambleton, Swaminathan & Rogers, 1991). Faktör analizi sonucunda veri setinin tek boyutlu özellik göstermesi, yerel bağımsızlık varsayımının karşılandığının bir kanıtı olarak sayılabilir (Baker, 2001). Bu bulguya ek olarak tüm grup ve alt-üst gruptaki maddeler arası korelasyonlar da incelenerek bu varsayım test edilmiştir. Bunun için veri seti %27’lik alt ve üst grup olmak üzere iki gruba ayrılmıştır. Alt ve üst gruptan ve grubun tamamından elde edilen maddeler arası korelasyon değerleri hesaplanmıştır. Bu korelasyon değerlerinin büyük bir çoğunluğu her iki grup için düşük düzeyde ilişki göstermiştir. Buna göre sınırlı yetenek aralığındaki maddeler arasında düşük ilişki olması ve tek boyutluluk varsayımının sağlanmasından dolayı yerel bağımsızlık varsayımının da karşılandığı kabul edilmiştir.

Model-veri uyumu. Veri setinin hangi lojistik modele uygun olduğunu incelemek amacıyla veri seti 1 PL, 2 PL ve 3 PL modele uygun olarak analiz edilmiştir. Hesaplanan -2 log (olabilirlik) değerleri Tablo 3’te sunulmuştur.

Tablo 3

Modellere İlişkin -2 Log (Olabilirlik) Değerleri

Model	-2 log (olabilirlik)	Fark
1 PL	2125726.0064	1 PL-2 PL=107928.0000
2 PL	2017798.0064	2 PL-3PL=40024.0912
3 PL	1977773.9152	

Tablo 3’ten izlenebileceği gibi 1 PL ve 2 PL modele ilişkin -2 log (olabilirlik) değerlerinin farkı 107928.0000, 2 PL ve 3 PL modele ilişkin -2 log (olabilirlik) değerlerinin farkı 40024.0912’dir. Modele eklenen parametre sayısına göre belirlenen 159 serbestlik derecesinde ve 0.05 anlamlılık düzeyinde χ^2 tablo değeri 189.424’tür. Buna göre 3 PL ile 2 PL modelin -2 log (olabilirlik) değerleri arasındaki farkın çok fazla olmadığı dikkate alınarak 2 PL model tercih edilmiştir

(Hambleton, Swaminathan & Rogers, 1991). Bu sonuç OECD tarafından açıklanan teknik raporla da uyumludur (OECD, 2017).

Bu sonuçlar dikkate alınarak çalışmada tek boyutlu olduğu belirlenen iki kategorili veri setine ait ölçümler (kalibrasyon) 2 PL modele uygun olarak kestirilmiştir. Benzer şekilde PISA 2015 uygulamasında da 2 PL modele uygun olarak kestirim yapılmıştır. 2015 yılında gerçekleştirilen PISA uygulamasına ait madde parametreleri Ek-B'de, madde ve yetenek parametrelerine ilişkin betimleyici istatistikler ise Tablo 4'te sunulmuştur.

Tablo 4

2015 Yılında Gerçekleştirilen PISA Uygulamasına Ait Madde ve Yetenek Parametrelerine İlişkin Betimleyici İstatistikler

	Ortalama	S.s.	Min	Mak
Madde Parametreleri				
Ayırt edicilik (a)	1.16	0.06	0.18	2.21
Güçlük (b)	0.07	0.30	-2.65	1.58
Yetenek Parametresi	0.00	1.02	-2.85	2.97

Tablo 4'ten izlenebileceği gibi 2015 yılında gerçekleştirilen PISA uygulamasına ait veri setinin ayırt edicilik parametresi değerinin ortalaması 1.16, standart sapması 0.06; güçlük parametresi değerinin ortalaması 0.07, standart sapması 0.30'dur. Bireylere ait en küçük yetenek parametresi -2.85; en yüksek yetenek parametresi ise 2.97 olarak hesaplanmıştır.

Madde ve Yetenek Parametrelerinin Değişmezliğinin İncelenmesi.

Madde parametrelerinin değişmezliği, testin uygulandığı gruptan bağımsız olması; diğer bir ifadeyle farklı gruplarda madde parametresi değerlerinin büyüklük sıralamasının değişmemesidir (Hambleton, Swaminathan & Rogers, 1991, s. 56). Madde parametrelerinin değişmezliğini belirleyebilmek için bireyler tesadüfi olarak seçilen 11 gruba ayrılmıştır. Ardından farklı gruplarda 2 PL modele göre madde parametreleri kestirilmiştir, Pearson Momentler Çarpımı Korelasyon tekniği ile gruplar arasındaki madde parametreleri karşılaştırılmıştır. EK-B'de verilen madde parametreleri arasındaki korelasyonlar Tablo 5'te sunulmuştur.

Tablo 5

Tesadüfi Olarak Atanan 11 Gruba Ait Madde Parametreleri Arasındaki Korelasyonlar

Grup	a	b	Grup	A	B	Grup	a	b	Grup	a	b
1-2	0.97*	0.99*	2-7	0.96*	0.99*	4-6	0.96*	0.99*	6-9	0.94*	0.99*
1-3	0.96*	0.98*	2-8	0.96*	0.99*	4-7	0.95*	0.99*	6-10	0.96*	0.99*
1-4	0.97*	0.99*	2-9	0.95*	0.99*	4-8	0.95*	0.99*	6-11	0.96*	0.99*
1-5	0.97*	0.99*	2-10	0.97*	0.99*	4-9	0.96*	0.99*	7-8	0.95*	0.99*
1-6	0.97*	0.99*	2-11	0.95*	0.99*	4-10	0.97*	0.99*	7-9	0.92*	0.99*
1-7	0.97*	0.99*	3-4	0.97*	0.99*	4-11	0.96*	0.99*	7-10	0.96*	0.99*
1-8	0.97*	0.99*	3-5	0.96*	0.99*	5-6	0.95*	0.99*	7-11	0.94*	0.98*
1-9	0.94*	0.99*	3-6	0.95*	0.99*	5-7	0.96*	0.99*	8-9	0.94*	0.99*
1-10	0.97*	0.99*	3-7	0.95*	0.99*	5-8	0.96*	0.99*	8-10	0.96*	0.99*
1-11	0.97*	0.99*	3-8	0.96*	0.99*	5-9	0.93*	0.99*	8-11	0.96*	0.99*
2-3	0.96*	0.99*	3-9	0.96*	0.99*	5-10	0.97*	0.99*	9-10	0.96*	0.99*
2-4	0.97*	0.99*	3-10	0.97*	0.99*	5-11	0.95*	0.99*	9-11	0.95*	0.99*
2-5	0.96*	0.99*	3-11	0.95*	0.99*	6-7	0.97*	0.99*	10-11	0.97*	0.99*
2-6	0.96*	0.98*	4-5	0.95*	0.99*	6-8	0.95*	0.99*			

N_{Toplam birey}: 15059, N_{1-N11}: 1369

* $p < 0.01$

Tablo 5'ten izlenebileceği gibi tesadüfi atanan, her biri 1369 bireyden oluşan 11 grup üzerinden kestirilen madde parametreleri arasında anlamlı, yüksek düzeyde ilişkiler vardır ($p < 0.01$). Sonuç olarak 2 PL model için madde parametrelerinin değişmezliği varsayımının karşılanmış olduğu tespit edilmiştir.

Yetenek parametrelerinin değişmezliği, bireylere ait yetenek parametrelerinin madde setinden bağımsız olmasıdır. Yetenek parametrelerinin değişmezliğini belirleyebilmek için 159 madde tesadüfi olarak 53 maddeden oluşan üç farklı gruba ayrılmış, birey yetenekleri bu alt gruplardan 2 PL modele göre kestirilmiştir. Ardından kestirilen yetenek parametreleri arasındaki ilişkiye "Pearson Momentler Çarpımı Korelasyon Katsayısı" ile bakılmıştır. Korelasyon değerleri Tablo 6'da sunulmuştur.

Tablo 6

Yetenek Parametreleri Arasındaki Korelasyonlar

Değişkenler	Madde Seti-1	Madde Seti-2	Madde Seti-3
Madde Seti-1	---		
Madde Seti-2	0.91*	---	
Madde Seti-3	0.90*	0.90*	---

N_{Toplam madde}: 159, N_{1-N3}: 53

$p < 0.01$

Tablo 6'dan izlenebileceği gibi tesadüfi atanan, her biri 15059 birey ve 53 maddeden oluşan üç alt grubun kestirilen yetenek parametreleri arasında anlamlı, pozitif yönde yüksek düzeyde ilişkiler bulunmuştur. Korelasyon değerlerinin bu şekilde birbirine yakın ve yüksek değerler almaları bireylerin yetenek kestirimlerinin maddelerden bağımsız olarak kestirilebilirliğini göstermektedir.

Verilerin Analizinin İkinci Aşaması

Birinci aşamada veri setinin çalışma için uygunluğuna karar verilmiş, ikinci aşamada ise her alt problem için BÇAT simülasyonu gerçekleştirilmiştir. Analizler, madde havuzundan 24 simülasyon koşuluna uygun olarak seçilmiş madde parametreleri ve yine 24 simülasyon koşuluna uygun olarak seçilen birey yetenekleri üzerinden gerçekleştirilmiştir. BÇAT'ın oluşturulması için R (R Development Core Team, 2011) yazılımı üzerinden 'xxIRT' (Luo, 2017) paket programı kullanılmıştır. Sonuçların genellenebilirliğini arttırmak amacıyla her bir koşul için 30 tekrar yapılmıştır. BÇAT simülasyonunda kullanılan BÇAT değişkenleri test birleştirme, modül uzunluğu, panel deseni ve örneklem büyüklüğüdür. Bu değişkenler Tablo 7'de sunulmuştur. Değişkenlere ilişkin bilgilere aşağıda yer verilmiştir.

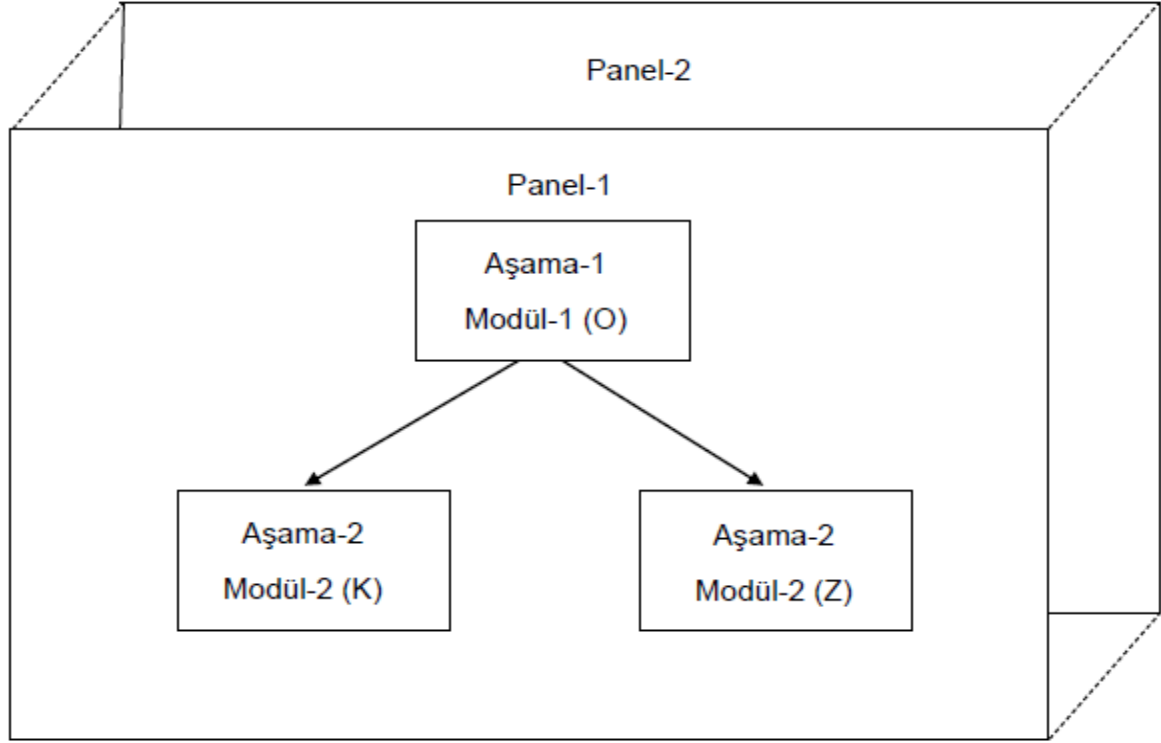
Tablo 7

BÇAT Değişkenleri

Değişken	Düzeyler
Test birleştirme	Yukarıdan-aşağıya
Panel deseni	Aşağıdan-yukarıya
Modül uzunluğu	"1-2"
Örneklem büyüklüğü	"1-2-2"
	"1-2-3"
	6
	12
	250
	2000

Panel Deseni. Paneller, BÇAT'ın test özelliklerini içeren birleştirilmiş modüllerdir. Madde ve modül kullanım sıklığının kontrol edilebilmesi için birden çok panel oluşturulmalıdır (Luecht & Nungester, 1998). Bu nedenle araştırmada her bir panel deseni için iki panel oluşturulmuştur. Paneller arasında karşılaştırma yapabilmek ve güvenilirliği artırabilmek için aşama sayısı, modül uzunluğu, modüllere giden yollar gibi panel yapıları her iki panelde de aynıdır. Araştırmada iki ("1-2") ve üç ("1-2-2" ve "1-2-3") aşamalı panel desenlerine uygun olarak BÇAT'lar oluşturulmuştur. Bu üç panel deseni, alanyazında en fazla araştırılan BÇAT panel desenleri arasında yer aldıkları için bu araştırmada kullanılmıştır

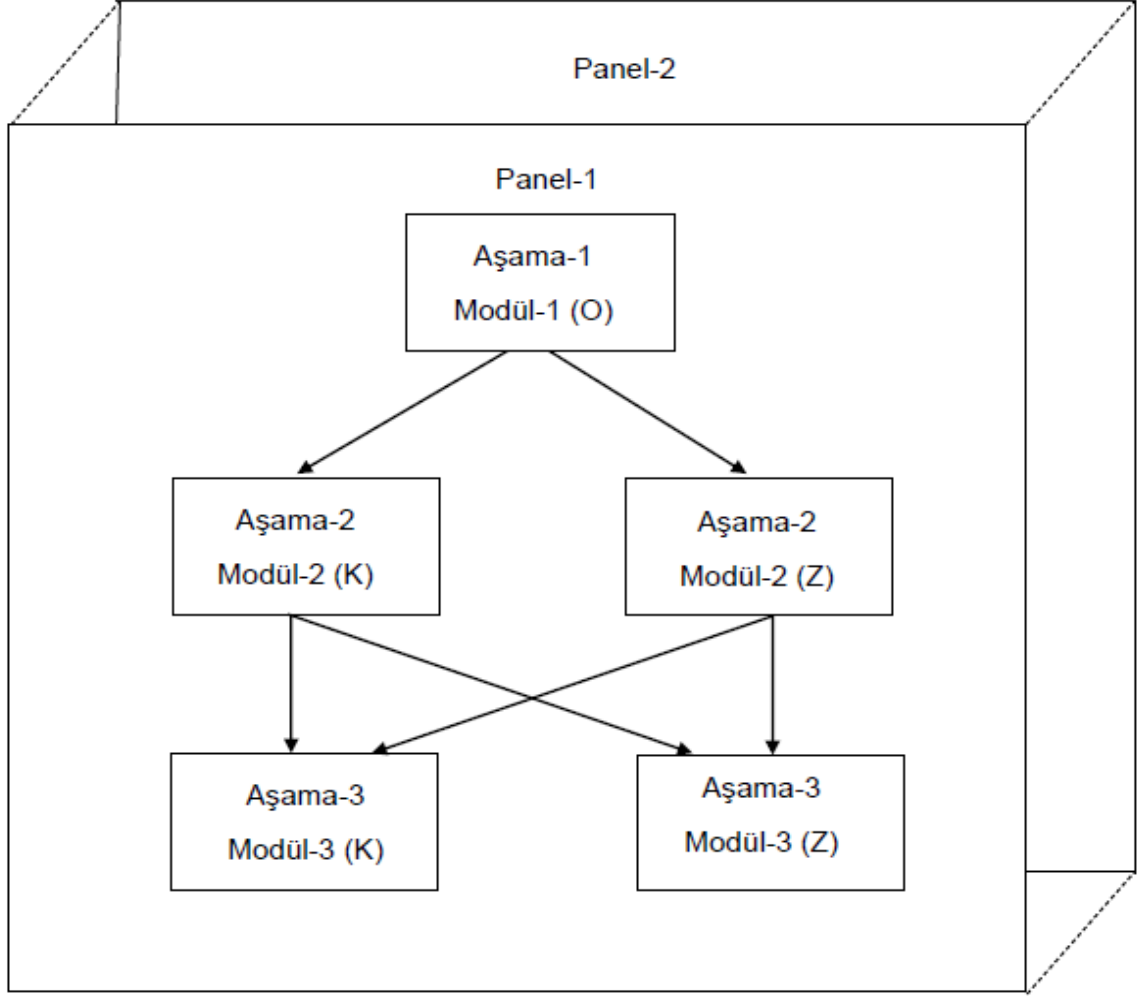
(Jodoin, Zenisky & Hambleton 2006; Luecht, Brumfield & Breithaupt, 2006; Wang, 2017; Zenisky, 2004). Her aşamada üçten fazla modül içeren paneller ve üç aşamadan daha fazla aşaması olan desenler test birleştirilmesi sırasında test karmaşıklığını arttırdığından araştırmaya dahil edilmemiştir (Luecht & Nungester, 1998; Patsula & Hambleton, 1999; Zenisky, Hambleton & Luecht, 2010). Araştırma kapsamında kullanılan “1-2” panel deseni Şekil 4’te şema ile sunulmuştur.



K: Kolay, O: Orta, Z: Zor

Şekil 4. İki panelden oluşan iki aşamalı “1-2” panel deseni.

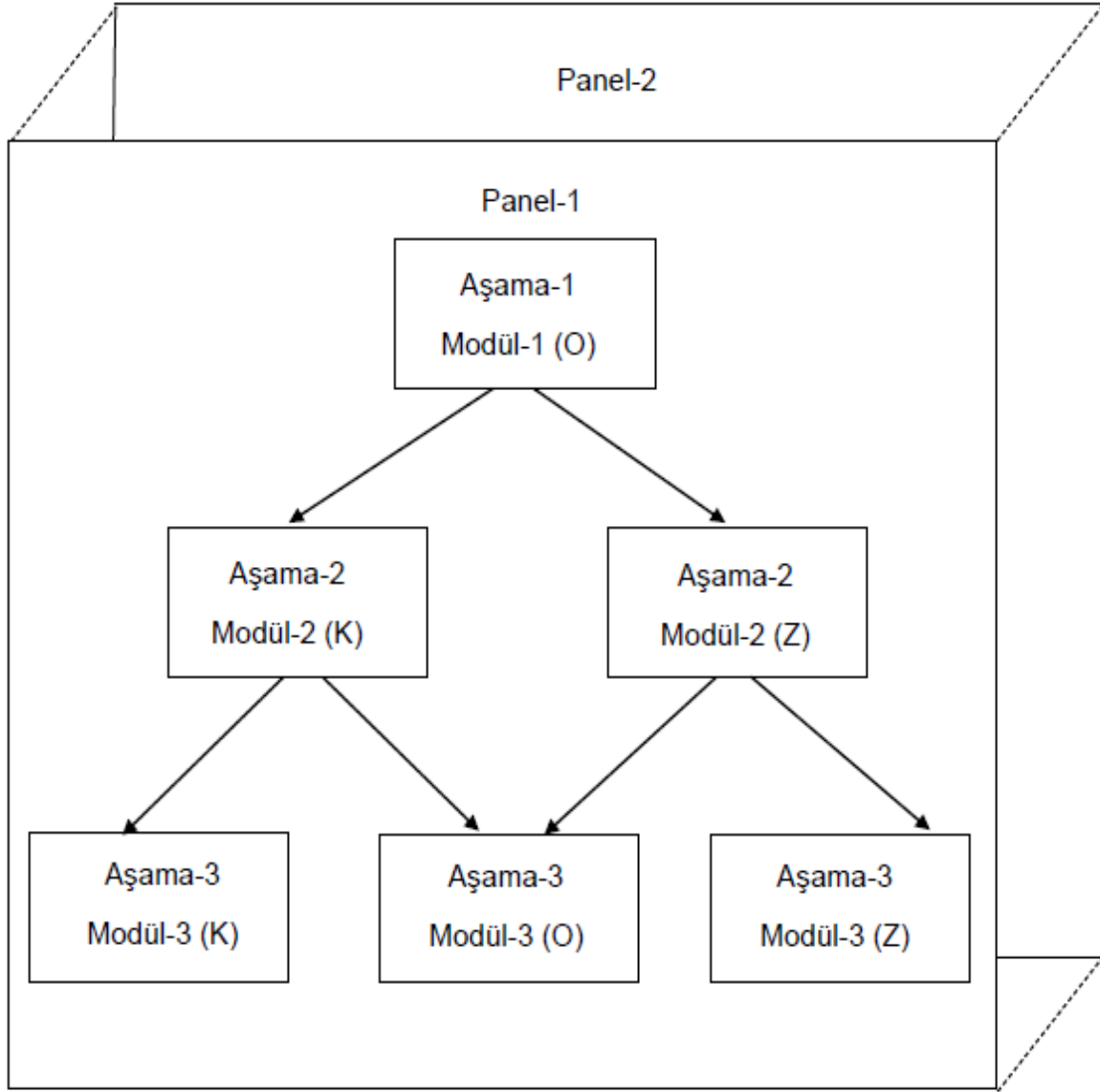
Şekil 4 incelendiğinde “1-2” panel deseninin iki aşama ve panelden oluştuğu görülmektedir. Birinci aşamada orta güçlük düzeyindeki Modül-1 (O), ikinci aşamada kolay güçlük düzeyindeki Modül-2 (K) ve zor güçlük düzeyindeki Modül-2 (Z) bulunmaktadır. Araştırma kapsamında kullanılan diğer panel deseni olan “1-2-2” panel deseni Şekil 5’te şema ile sunulmuştur.



K: Kolay, O: Orta, Z: Zor

Şekil 5. İki panelden oluşan üç aşamalı “1-2-2” panel deseni.

Şekil 5 incelendiğinde “1-2-2” panel deseninin üç aşamadan ve iki panelden oluştuğu görülmektedir. Birinci aşamada orta güçlük düzeyindeki Modül-1 (O), ikinci aşamada kolay güçlük düzeyindeki Modül-2 (K) ve zor güçlük düzeyindeki Modül-2 (Z), üçüncü aşamada kolay güçlük düzeyindeki Modül-3 (K) ve zor güçlük düzeyindeki Modül-3 (Z) bulunmaktadır. Araştırma kapsamında kullanılan bir diğer panel deseni olan “1-2-3” panel deseni Şekil 6’da şema ile sunulmuştur.



K: Kolay, O: Orta, Z: Zor

Şekil 6. İki panelden oluşan üç aşamalı “1-2-3” panel deseni.

Şekil 6 incelendiğinde “1-2-3” panel deseninin üç aşamadan ve iki panelden oluştuğu görülmektedir. Birinci aşamada orta güçlük düzeyindeki Modül-1 (O), ikinci aşamada kolay güçlük düzeyindeki Modül-2 (K) ve zor güçlük düzeyindeki Modül-2 (Z), üçüncü aşamada kolay güçlük düzeyindeki Modül-3 (K), orta güçlük düzeyindeki Modül-3 (O) ve zor güçlük düzeyindeki Modül-3 (Z) bulunmaktadır.

Modül uzunluğu. BÇAT’da modül uzunluğu, diğer bireyselleştirilmiş testler ile benzer öneme sahiptir. Bireyin yanıtladığı madde sayısının artması ölçme doğruluğunu artırır (Patsula, 1999). Ancak daha uzun testlerin oluşturulması, güçlü bir madde havuzu ve iyi bir bilgisayar yazılımı kullanımını da gerektirir.

BÇAT çalışmalarında test uzunluğunun 33 ile 60 madde arasında değiştiği tespit edilmiştir (Hambleton & Xing, 2006; Jodoin, Zenisky & Hambleton, 2006; Patsula, 1999; Zenisky, 2004). Bu araştırmada kısa test uzunluğunu temsil eden modül sayısı 6, orta test uzunluğunu temsil eden modül sayısı, kısa test uzunluğunun iki katı olan 12 olarak belirlenmiştir. Modül uzunluğunun 6 olduğu “1-2” panel deseninde bireyin testin sonunda toplam 12 maddeyi, “1-2-2” ve “1-2-3” panel desenlerinde toplam 18 maddeyi yanıtlandıracağı şekilde BÇAT’lar tasarlanmıştır. Modül uzunluğunun 12 olduğu “1-2” panel deseninde birey testin sonunda toplam 24 maddeyi, “1-2-2” ve “1-2-3” panel desenlerinde ise toplam 36 maddeyi yanıtlandırmıştır. Bir paneldeki modüllerin madde sayıları eşittir. Panel ve modülde farklı maddeler bulunmaktadır. Panel tasarımlarındaki modül ve test uzunlukları Tablo 8’de sunulmuştur.

Tablo 8

Panel Tasarımlarına Ait Modül ve Test Uzunlukları

Panel deseni	Modül sayısı	Modül uzunluğu	Her paneldeki test uzunluğu	İki paneldeki toplam madde sayısı
“1-2”	3	6	18	36
“1-2”	3	12	36	72
“1-2-2”	5	6	30	60
“1-2-2”	5	12	60	120
“1-2-3”	6	6	36	72
“1-2-3”	6	12	72	144

Madde Havuzu. 2015 yılında uygulanan PISA verilerinden elde edilen iki kategorili veri setinden 2 PL modele göre kalibre edilen madde parametreleri madde havuzunu oluşturmaktadır. Madde havuzunda toplam 159 madde bulunmaktadır. Madde parametreleri EK-B’de yer almaktadır.

Örneklem Büyüklüğü. 2015 yılında uygulanan PISA’ya katılan 15059 birey arasından tesadüfi olarak seçilen 250 ve 2000 birey araştırmanın örneklemini oluşturmaktadır. Bu tesadüfi olarak seçilen 250 ve 2000 bireye çeşitli modül uzunluğu ve panel deseni değişkenlerine göre oluşturulan BÇAT’lar uygulanmıştır. Alanyazın incelendiğinde BÇAT çalışmalarında 250 (Yan, Lewis & von Davier, 2014) ile 5000 arasındaki örneklem büyüklükleri ile çalışıldığı görülmektedir (Dallas, 2014; Sari, 2016; Wang, 2017; Xing & Hambleton, 2004; Yang, 2016). Bu araştırmada BÇAT deseninin büyük örneklere uygulanmasının yanı sıra, küçük örneklere de uygulanabilirliğinin değerlendirilmesi hedeflendiğinden,

araştırmada küçük örneklem büyüklüğünü 250 ve büyük örneklem büyüklüğünü 2000 birey temsil etmektedir.

Test Birleştirme. BÇAT çalışmalarında en çok kullanılan iki otomatik test birleştirme yöntemi vardır: Yukarıdan-aşağıya (top-down) ve aşağıdan-yukarıya (bottom-up). Bu nedenle bu iki yöntem araştırmacının odak noktası olarak seçilmiş ve bu yöntemler arasında karşılaştırma yapmak amaçlanmıştır. Bu doğrultuda araştırmada yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemlerine göre BÇAT yapıları oluşturulmuştur. Her iki yönetime uygun olarak hedef Test Bilgi Fonksiyonu (TBF) değeri Ortalama Maksimum Bilgi (OMB) (Luecht, 2000; Luecht, Brumfield & Breithaupt, 2006) stratejisi ile belirlenmiştir. Bu strateji aşağıdaki gibidir:

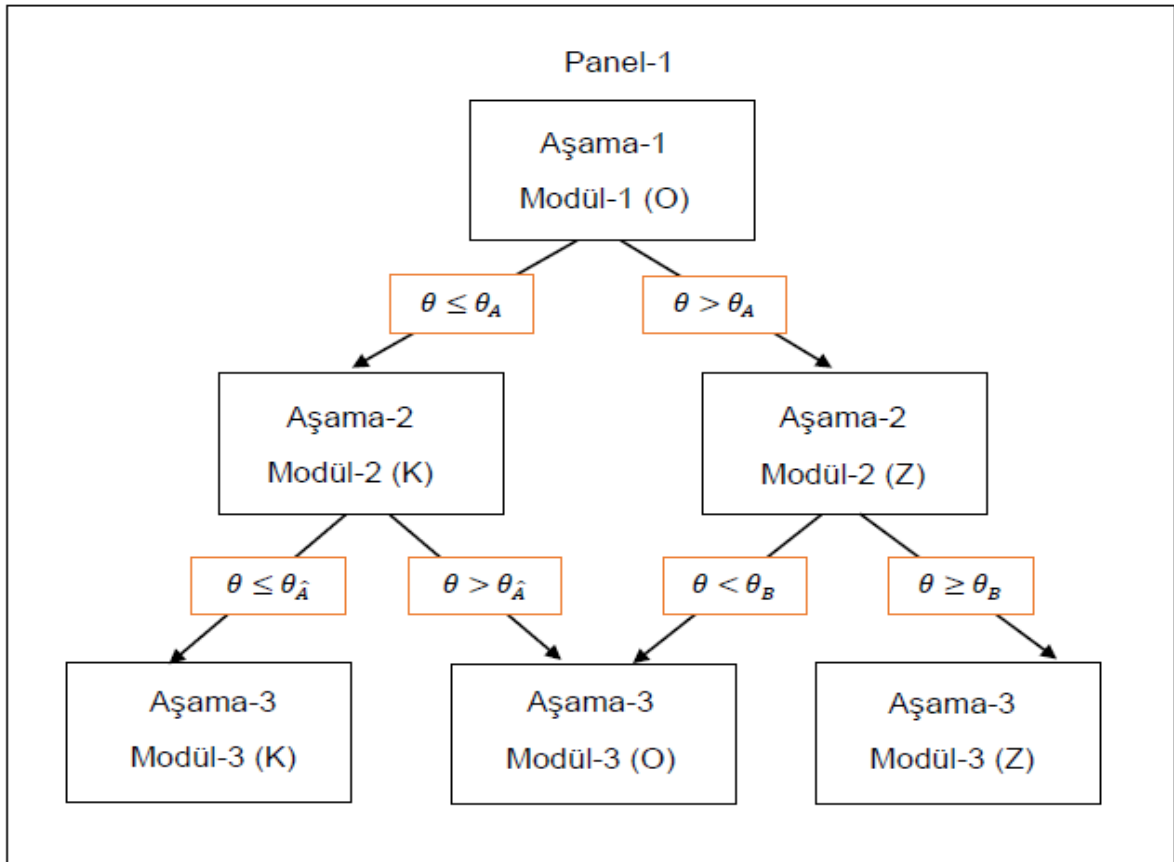
- θ ölçeğinde belirli bir nokta belirlenir. Bu nokta genellikle TBF'nin istenen maksimum bilgi (tepe) noktasına karşılık gelir. Örneğin; bir bireyin yetenek düzeyi 0.00 olduğunda 'orta' hedef TBF'nin, yetenek dağılımının ortalamasındaki maksimum bilgiye sahip olması beklenir.
- Madde havuzundaki her bir madde için maksimum bilgi istenilen yerlerin birindeki (örneğin, yetenek düzeyinin 0.00 olduğu noktadaki) madde bilgisi hesaplanır.
- Hesaplanan madde bilgileri en yüksek değerden en düşük değere doğru sıralanır.
- Modül uzunluğu 'n' ve panel sayısı (çoğaltılan modüllerin sayısı) 'm' olduğunda, daha önce belirlenen yerdeki en bilgilendirici $n \times m$ maddeler seçilir.
- Yetenek ölçeğinde seçilen noktaların her birindeki bu $n \times m$ maddenin madde bilgilerinin toplamı hesaplanır. Sonrasında TBF hedefleri bu toplam miktarın 'm'ye bölünmesi ile elde edilir ve eşitlik 3.1'de sunulmuştur:

$$TBF(\theta_k) = \frac{\sum_{i=1}^{n \times m} l_i(\theta_k)}{m} \quad (3.1)$$

Birbirleriyle çakışmayan iki paralel panel, yukarıda açıklanan OMB yönteminin aşamalarına uygun olarak aynı anda oluşturulmuştur.

Yönlendirme Stratejisi ve Puanlandırma. BÇAT'daki yönlendirme stratejisi, BBT'deki madde seçim sürecine benzer şekilde bireyin bir önceki modül

sonunda kestirilen yeteneğine göre, bir sonraki aşamadaki uygun modüle atanmasıyla gerçekleşir. Bu çalışmada yönlendirme stratejisi olarak yaygın kullanılan Ortalama Maksimum Bilgi (OMB) stratejisi seçilmiştir (Luecht, Brumfield & Breithaupt, 2006; Zenisky, Hambleton & Luecht, 2010). Bu strateji öncelikle hedef TBF değerini tanımlar, sonrasında daha önce yönetilen bir modülün ve alternatif modüllerin TBF'lerini toplayarak bitişik kümülatif TBF'lerin kesişme noktasını bulur. Yani, her bir alternatif modülün TBF'sini ayrı ayrı kümülatif TBF'ye ekler ve TBF'lerin kesişme noktalarını kesme puanı olarak tanımlar. Örneğin, üç TBF'nin iki kesişim noktası -1.0 ve 1.0 ise, kesme noktaları da -1.0 ve 1.0'dır. Testin son aşamasında birey, geçici gizil yetenek için en yüksek bilgiyi sağlayan alternatif modüllerden birisine yönlendirilir. Örneğin, bireyin yeteneği -1.0'dan küçükse birey kolay güçlükteki modüle; -1.0 ile 1.0 aralığında ise orta güçlükteki modüle; 1.0'dan büyük ise zor güçlükteki modüle yönlendirilir. Bu yöntem BBT'deki maksimum bilgi madde seçme yöntemine benzer olarak çalışır. Araştırmada incelenen panel desenlerinden birisi olan "1-2-3" panel deseninde OMB yönteminin nasıl çalıştığı Şekil 7'de sunulmuştur.



Şekil 7. "1-2-3" BÇAT panel desenine ilişkin OMB yönlendirme yöntemi.

Şekil 7'den de izlenebileceği gibi "1-2-3" panel deseninde ikinci aşama için yönlendirme noktası θ_A olarak belirlenmiştir. θ_A yönlendirme noktası birinci aşamadaki Modül-1 (O) ile ikinci aşamadaki Modül-2 (K)'nin test bilgi fonksiyonları toplanarak bitişik modüller için elde edilen toplam test bilgi eğrisi ile Modül-1 (O) ve Modül-2 (Z)'nin test bilgi fonksiyonlarının toplanmasıyla bitişik modüller için elde edilen toplam test bilgi eğrisinin kesişim noktasıdır. Birinci aşamadaki modülün (Modül-1 (O)) yanıtlandırılmasından sonra bireyin kestirilen yeteneği yönlendirme noktasından küçük veya eşit olduğunda ($\theta \leq \theta_A$), birey ikinci aşamadaki kolay güçlükteki modüle (Modül-2 (K)) yönlendirilir. Ancak bireyin kestirilen yeteneği yönlendirme noktasından büyük olduğunda ($\theta > \theta_A$), birey ikinci aşamadaki zor güçlükteki modüle (Modül-2 (Z)) yönlendirilir. İkinci aşamadan üçüncü aşamaya geçişte yönlendirme noktaları, üç aşamada toplanan test bilgi eğrileri ile belirlenir. Bu noktalar $\theta_{\hat{A}}$ ve θ_B olarak adlandırıldığında birey ikinci aşamadakine benzer şekilde üçüncü aşamadaki modüllere yönlendirilir. İkinci aşamadan üçüncü aşamaya kadar kestirilen yetenek $\theta_{\hat{A}}$ 'den küçük veya eşitse ($\theta \leq \theta_{\hat{A}}$), birey üçüncü aşamadaki kolay güçlükteki modüle (Modül-3 (K)), büyükse ($\theta > \theta_{\hat{A}}$) orta güçlükteki modüle (Modül-3 (O)) yönlendirilir. Kestirilen yetenek θ_B 'den küçükse ($\theta < \theta_B$), birey üçüncü aşamadaki orta güçlükteki modüle (Modül-3 (O)), eşit ve büyükse ($\theta \geq \theta_B$) zor güçlükteki modüle (Modül-3 (Z)) yönlendirilir. Benzer şekilde "1-2" panel deseni için bir karar noktası, "1-2-2" panel deseni için iki yönlendirme noktası belirlenmiştir. Panellerdeki her bir modülün TBF'si büyük bir ihtimalle farklı olduğundan, yönlendirme noktalarının da panellerde farklılık göstermesi olasıdır.

OMB yöntemi yönlendirme kesme noktası puanı olarak modül bilgisinin kesişim noktasını kullanır (Breithaupt & Hare, 2007). Kesme puanı yetenek düzeyinin (θ) üzerinde olduğunda zor güçlük düzeyindeki modül daha fazla bilgi sağlar. Yetenek düzeyinin (θ) altında olduğunda ise daha düşük güçlük düzeyindeki modül daha fazla bilgi sağlar. Paneller paralel olarak tasarlanmasına rağmen, gerçek modül bilgi eğrileri paneller arasında değişiklik gösterir. Böylece OMB kesme θ değerleri her panel için ayrı ayrı hesaplanır. Her aşama ve yönlendirmeden sonra bireyin geçici yetenek (θ) kestirimi elde edilir. Bu araştırmada OMB yöntemi uygulanmış, öncelikle bireyler test koşullarının her birindeki iki panelden birisine tesadüfi olarak atanmıştır. Bir panele atandıktan sonra bireyler ilk aşamada yer alan modüle yönlendirilmiştir. İlk aşamadaki

yönlendirme modülü bireye uygulandıktan sonra bireyin yeteneği Maksimum Olasılık Kestirimi (MOK) yöntemi kullanılarak kestirilmiştir. Aynı noktalarda maksimum bilgi veren bir panelin birden fazla modülü bulunduğu (örneğin, “1-2-3” panel deseninde iki modülün en yüksek bilgiyi sıfır noktasında verdiği bir durumda) önceki aşamadaki modül, sonraki aşamalardaki modüllerden önce (geriye doğru birleştirme) birleştirilir. Tüm koşullarda yönlendirme modülü (Modü-1 (O)), orta güçlük düzeyindeki maddeler seçilerek oluşturulmuştur.

Test Yönetimi. Çalışmada incelenen “1-2”, “1-2-2” ve “1-2-3” panel desenlerinin test yönetimi aşağıdaki adımlar izlenerek gerçekleştirilmiştir:

1. Birey iki farklı panel deseninden birisine tesadüfi olarak atanmıştır.
2. Birey atanmış olduğu panelin yönlendirme modülünü (orta güçlükteki modül) yanıtlamıştır.
3. Yönlendirme modülü tamamlandıktan sonra bireyin yeteneği Maksimum Olabilirlik Kestirimi (MOK) yöntemi kullanılarak kestirilmiştir.
4. Birinci aşama sonrasında bireyin kestirilen yeteneği (θ) ve önceden belirlenen yönlendirme noktaları karşılaştırılarak birey, birinci aşamadan ikinci aşamaya yönlendirilmiştir.
5. İkinci aşama sonrasında bireyin yeteneği tekrar MOK yöntemi ile kestirilmiştir. “1-2” panel yapısındaki testi alan birey için test uygulaması burada sonlandırılmıştır. “1-2-2” ve “1-2-3” panel yapısındaki testi alan bireyler için test yönetimi bireyin ikinci aşama sonrasında kestirilen yeteneği (θ) ve önceden belirlenen yönlendirme noktaları karşılaştırılarak bireyin ikinci aşamadan üçüncü aşamaya yönlendirilmesi şeklinde gerçekleştirilmiştir.
6. Üçüncü aşamanın tamamlanmasıyla bireyin nihai yetenek kestirimi gerçekleştirilerek test sonlandırılmış ve test bilgileri hesaplanmıştır.

Değerlendirme Ölçütleri. BÇAT’ın performansı yetenek kestirimine dayalı olarak ortalama RMSE ve yanlışlık ölçütlerine göre değerlendirilmiştir. Bu ölçütler, R yazılımında fonksiyonlar yazılarak hesaplanmıştır. Ayrıca birey performansı sınıflandırma doğruluğu; doğru sınıflandırma, yanlış-negatif hata, yanlış-pozitif hata ve Kappa değeri ölçütlerine göre değerlendirilmiştir. Bu ölçütlerin hesaplanması için SPSS 20 istatistik programından yararlanılmıştır.

Yetenek Kestiriminin Değerlendirilmesi. Yetenek kestirimi RMSE ve yanlılık ölçütlerine göre değerlendirilmiştir. *RMSE (Root Mean Squared Error-hataların ortalama karekökü)*, bireye uygulanan BÇAT sonucunda kestirilen yetenek düzeyi ile gerçek yetenek düzeyi arasındaki farkın karesinin ortalamasının kareköküdür. Her bir simülasyon koşulu için bireyin yetenek düzeyinde RMSE değeri eşitlik 3.2’de sunulan formül yardımıyla hesaplanır:

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (\hat{\theta}_j - \theta_j)^2}{N}} \quad (3.2)$$

Bu formülde N , toplam birey sayısı; $\hat{\theta}_j$, j . bireyin kestirilen yetenek düzeyi değeri; θ_j , j . bireyin gerçek yetenek düzeyi değeridir.

Yanlılık (Bias), BÇAT sonucunda kestirilen yetenek düzeyi ile gerçek yetenek düzeyi arasındaki ortalama farklılık istatistiğidir. Her bir simülasyon koşulu için bireyin yetenek düzeyinde yanlılık değeri eşitlik 3.3’te sunulan formül yardımıyla hesaplanır:

$$Yanlılık = \frac{\sum_{j=1}^N (\hat{\theta}_j - \theta_j)}{N} \quad (3.3)$$

Yukarıda eşitlikte $\hat{\theta}_j$ ve θ_j , j bireyinin kestirilen ve gerçek yetenek değerleri; N , toplam birey sayısıdır.

RMSE ve yanlılık değerleri her bir koşul için tekrar sayısı olan 30 kez hesaplanmış, sonrasında 30 RMSE ve yanlılık değerinin ortalaması alınarak her bir koşul için tek bir RMSE ve yanlılık değeri elde edilmiştir.

Çeşitli koşullarda yer alan BÇAT değişkenlerinin ortalama RMSE ve yanlılık değerleri üzerindeki anlamlılıklarının test edilmesinde çok yönlü varyans analizi (ANOVA) tekniği kullanılmıştır. Ortalamalar arası farkların hangi koşullar arasında olduğunu bulmak amacıyla ise ‘Bonferroni’ çoklu karşılaştırma testi yapılmıştır. ANOVA testi sonucu elde edilen etki büyüklükleri Cohen (1988) tarafından önerilen 0.01 (küçük etki), 0.06 (orta etki) ve 0.14 (büyük etki) ölçütlerine göre değerlendirilmiştir. Elde edilen bulgular, alt problemlerin sırasına göre sunulmuştur.

Sınıflandırma Doğruluğunun Değerlendirilmesi. Araştırmada bireylerin doğru sınıflandırılıp sınıflandırılmadığının belirlenmesi amacıyla ‘sınıflandırma

doğrulukları' hesaplanmıştır. Sınıflandırma doğruluğu her bir bireyin BÇAT sonucunda kestirilen yeteneği ile gerçek yeteneği karşılaştırılarak gerçekleştirilmiştir. Bunun için bireylerin gerçek yetenek düzeyleri ve BÇAT sonucunda kestirilen yetenek düzeyleri kesme puanlarının (-0.524, 0.00 ve 0.524) üzerinde olduğunda 'geçti', altında olduğunda ise 'kaldı' olarak sınıflandırılmıştır. -0.524, 0.00 ve 0.524 kesme puanları bireylerin yetenek puanlarının normal dağılım gösterdiği varsayılarak sırasıyla %70, %50 ve %30'luk geçme oranını göstermektedir. Kesme puanları ölçüt alınarak bireylerin olası sınıflandırma doğrulukları değerlendirilmiştir. Dört olası sınıflandırma kararı Tablo 9'da sunulmuştur.

Tablo 9

Olası Sınıflandırma Kararı Durumlarının Değerlendirilmesi

		Kestirilen Yetenek	
		$\hat{\theta} \geq \theta_{kesme}$	$\hat{\theta} < \theta_{kesme}$
Gerçek yetenek	$\theta \geq \theta_{kesme}$	Doğru-geçme	Yanlış-negatif
	$\theta < \theta_{kesme}$	Yanlış-pozitif	Doğru-kalma

Tablo 9'da yer alan 'doğru-geçme' kararı, bireyin gerçek yeteneği ve kestirilen yeteneği kesme puanına eşit veya büyük olduğunda verilir. 'Doğru-kalma' kararı da gerçek ve kestirilen yetenek puanları kesme puanının altında olduğunda verilir. 'Yanlış-negatif' hatalar birey gerçek yeteneğine göre geçmesi gerekirken, kestirilen yeteneğine göre kalması gerektiği durumda gerçekleşir. 'Yanlış-pozitif' hatalar da birey gerçek yeteneğine göre geçmemesi gerekirken, kestirilen yeteneğine göre geçtiğinde gerçekleşir. Sınıflandırma kararının doğruluğunun değerlendirilmesi için Cohen'in (1960) Kappa (κ) katsayısı bir karar ölçütü olarak hesaplanmıştır. Bu katsayı kararlar arasındaki şansın etki düzeyini dikkate alır. Bir diğer deyişle Kappa katsayısı, şansla beklenenin üzerinde verilen kararların oranıdır (Crocker & Algina, 1986; Swaminathan, Hambleton & Algina, 1974). Eşitlik 3.4'te sunulan formül ile hesaplanır:

$$\kappa = \frac{P - P_C}{1 - P_C} \quad (3.4)$$

Yukarıdaki eşitlikte P , gözlemlenen kararların uyum oranı, P_C şansla beklenen kararların oranını temsil eder. Swaminathan, Hambleton ve Algina

(1974), κ 'nın üst sınırı olan +1.00 deęerinin ancak her iki ölçümlerdeki marjinal oranların eşitlięi durumunda elde edileceğini belirtmiştir. κ 'nın alt sınırı ise -1.00 deęeridir. Ancak, $\kappa < 0.00$ deęerleri tutarsızlık anlamına gelir ve güvenilirlik açısından yoruma deęer bulunmaz. Sınıflama tutarlılıęı indeksleri ise verilerin nitelięinden bağımsız olarak testlerin bireyleri sınıflama tutarlılıęını yansıtmaktadır (Erkuş, 1999).



Bölüm 4

Bulgular

'Bulgular' bölümünde, alt problemlere göre verilmiş olan araştırma bulguları RMSE, yanlılık ve sınıflandırma doğruluğu ölçütlerine göre incelenmiştir.

Birinci Alt Probleme İlişkin Bulgular

"BÇAT uygulamasında yukarıdan-aşağıya test birleştirme yönteminde modül uzunlukları (6 ve 12), panel desenleri ("1-2", "1-2-2" ve "1-2-3") ve örneklem büyüklüklerine (250 ve 2000) göre

a) yetenek kestiriminin kesinliği ve

b) sınıflandırma kararının doğruluğu

nasıl bir değişim göstermektedir?"

Birinci alt probleme ilişkin bulguların yorumlanabilmesi için öncelikle problemde yer alan simülasyon koşullarına uygun olarak BÇAT'lar oluşturulmuştur. Bu koşullara göre oluşturulan bir BÇAT'ın her bir modülü için kestirilen madde parametreleri EK-C ve EK-Ç'de verilmiştir.

Çeşitli simülasyon koşullarına göre oluşturulan BÇAT'ların yetenek kestirimlerine ilişkin bulgular elde edilmiştir. Bu doğrultuda analiz sonuçlarından elde edilen ortalama RMSE ve yanlılık değerleri Tablo 10'da sunulmuştur.

Tablo 10

Yukarıdan-aşağıya Test Birleştirme Yöntemine Göre Çeşitli Simülasyon Koşullarına Göre Oluşturulan BÇAT'lara İlişkin Ortalama RMSE ve Yanlılık Değerleri

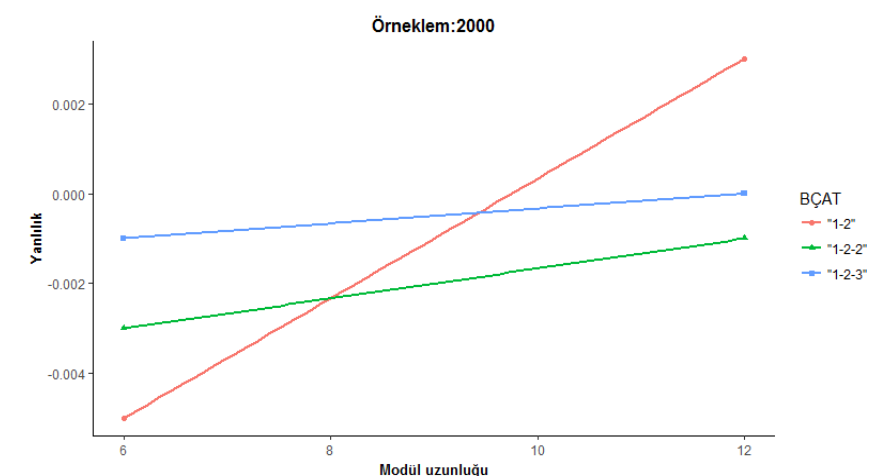
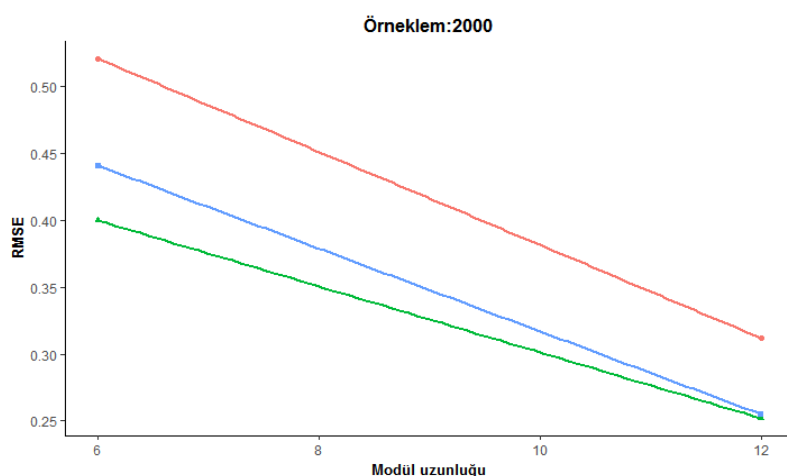
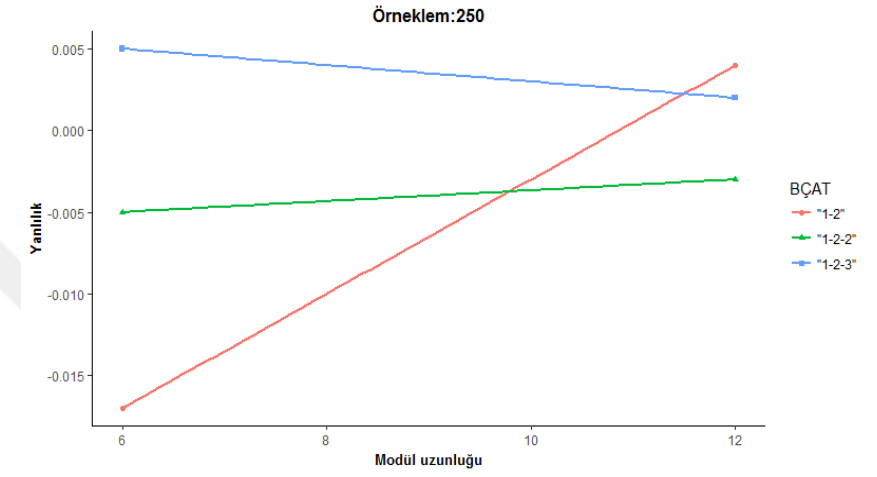
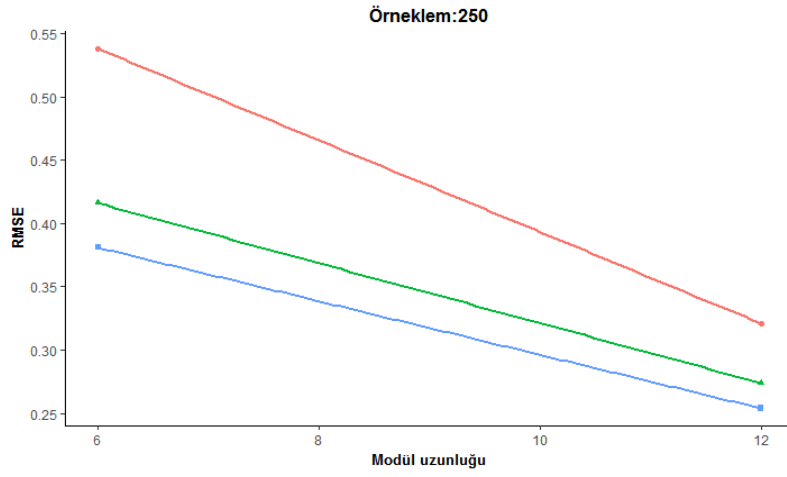
Örneklem	Panel deseni	Modül uzunluğu	RMSE	Yanlılık
250	"1-2"	6	0.538	-0.017
		12	0.321	0.004
	"1-2-2"	6	0.416	-0.005
		12	0.274	-0.003
	"1-2-3"	6	0.381	0.005
		12	0.254	0.002
2000	"1-2"	6	0.521	-0.005
		12	0.312	0.003
	"1-2-2"	6	0.400	-0.003
		12	0.252	-0.001
	"1-2-3"	6	0.441	-0.001
		12	0.255	0.000

Tablo 10'dan izlenebileceği gibi yukarıdan-aşağıya test birleştirme yöntemine göre farklı örneklem büyüklüğü, test uzunluğu ve panel desenlerinden elde edilen ortalama RMSE değerleri 0.252 ile 0.538 arasında değişmektedir. Bulgular genel hatlarıyla incelendiğinde en düşük hata kestiriminin orta uzunluktaki modülün büyük örnekleme uygulandığı "1-2-2" panel deseninde, en büyük hata kestiriminin kısa uzunluktaki modülün küçük örnekleme uygulandığı "1-2" panel deseninde kestirildiği görülmektedir. Bulgular modül uzunluğu açısından incelendiğinde her iki örneklem büyüklüğünde modül uzunluğu arttığında, ortalama RMSE değerinin azaldığı görülmektedir. Bulgular panel deseni açısından incelendiğinde küçük ve büyük örnekleme panel desenlerinin farklılaşmasının her test düzeyindeki ortalama RMSE miktarını değiştirdiği görülmektedir. "1-2" panel deseninden "1-2-2" ve "1-2-3" panel desenine doğru geçişte, ortalama RMSE değerleri düşmüştür. Ancak büyük örnekleme uygulanan "1-2-2" ve "1-2-3" panel desenlerinin her iki modül uzunluğundaki ortalama RMSE değerlerinde, "1-2-2" panel deseninden "1-2-3" panel desenine geçişte artış olduğu fark edilmiştir. Bu artış miktarı "1-2-2" panel deseninden "1-2-3" panel desenine geçişte kısa modül uzunluğu için 0.041, orta modül uzunluğu için 0.003'tür. Bulgular örneklem büyüklüğü açısından incelendiğinde örneklem büyüklüğünün artmasının "1-2-3" panel deseninin haricindeki diğer desenlerin her iki modül uzunluğunda ortalama RMSE değerlerini küçülttüğü görülmektedir. Küçük örnekleme en düşük ortalama

RMSE değeri; orta uzunluktaki modülün uygulandığı “1-2-3” panel deseninde; büyük örnekleme en düşük ortalama RMSE değeri orta uzunluktaki modülün uygulandığı “1-2-2” panel deseninde hesaplanmıştır.

Tablo 10’dan yanlılık ile ilgili bulgular incelenirse ortalama yanlılık değerlerinin genel olarak oldukça düşük değerlere sahip oldukları görülmektedir. Yukarıdan-aşağıya test birleştirme yöntemi seçildiğinde örneklem büyüklüğü, panel deseni ve modül uzunluğu simülasyon koşullarına ait ortalama yanlılık değerleri -0.017 ile 0.005 arasında değişmektedir. En yüksek ortalama yanlılık değerleri küçük örnekleme uygulanan modül uzunluğunun kısa olduğu “1-2” panel desenine aittir. Bu deseni modül uzunluğunun kısa olduğu, küçük örnekleme uygulanan “1-2-2”, “1-2-3” ve büyük örnekleme uygulanan “1-2” panel desenleri takip etmektedir. En düşük ortalama yanlılık değeri büyük örnekleme uygulanan modül uzunluğunun orta olduğu “1-2-3” panel deseninde hesaplanmıştır. Bu değer 0.000’dır, yani bu simülasyon koşulu yansız hesaplamalar yapmıştır. Bulgular modül uzunluğu açısından incelendiğinde de; modül uzunluğu arttıkça, her iki örneklem türünde panel desenlerinin yanlılığının azaldığı sonucuna varılmıştır. Bulgular panel deseni açısından incelendiğinde her iki modül uzunluğunda küçük ve büyük örnekleme uygulanan “1-2” panel deseninden “1-2-2” panel desenine geçişte ve “1-2-2” panel deseninden “1-2-3” panel desenine geçişte ortalama yanlılık değerlerinde düşüş olduğu görülmektedir. Bulgular örneklem büyüklüğü açısından incelendiğinde örneklem büyüklüğü arttıkça, ortalama yanlılık değerlerinde küçük bir miktar düşüş olduğu görülmektedir.

Tablo 10’da yer alan istatistiklere ait grafikler Şekil 8’de sunulmuştur.



Şekil 8. Yukarıdan-aşağıya test birleştirme yöntemine göre elde edilen ortalama RMSE ve yanlılık değerlerine ilişkin grafikler.

Şekil 8'den izlenebileceği gibi küçük ve büyük örneklerde modül uzunluğu arttıkça ortalama RMSE değeri azalmaktadır. Benzer şekilde küçük ve büyük örneklerde "1-2" panel deseninden "1-2-2" ve "1-2-3" panel desenine doğru geçişte, ortalama RMSE değerleri modül uzunluğunun kısa ve orta olduğu her iki durumda da düşmektedir. Ortalama yanlılık değerleri ise küçük ve büyük örneklerde modül uzunluğu arttıkça ve "1-2" panel deseninden "1-2-2" ve "1-2-3" panel desenine doğru geçişte 0.000 değerine yaklaşmaktadır.

Araştırmanın bu alt problemi kapsamında yukarıdan-aşağıya test birleştirme yöntemine göre elde edilen ortalama RMSE ve yanlılık bulguları üzerindeki modül uzunluğu, panel deseni ve örneklem büyüklüğünün istatistiksel olarak anlamlı etkisinin olup olmadığı çok yönlü ANOVA testi ile sınanmıştır. ANOVA testinden elde edilen F değeri ve hesaplanan etki büyüklükleri (η^2) Tablo 11'de sunulmuştur.

Tablo 11

Yukarıdan-aşağıya Test Birleştirme Yöntemi Seçildiğinde Elde Edilen Ortalama RMSE ve Yanlılık Değerlerine Ait ANOVA Sonuçları

Test Özellikleri	Değerlendirme Ölçütleri					
	sd	RMSE F	η^2	sd	Yanlılık F	η^2
Modül uzunluğu (M)	1	3379.332*	0.051	1	20.662*	0.049
Panel deseni (P)	2	404.320*	0.012	2	1.841	0.007
Örneklem (Ö)	1	0.034	0.052	1	3.753	0.007
P*M	2	50.648*	0.015	2	2.51	0.014
P*Ö	2	27.878*	0.008	2	8.395*	0.042
M*Ö	1	10.489*	0.001	1	0.686	0.014
P*M*Ö	2	12.019*	0.005	2	6.059*	0.028

* $p < 0.05$

Tablo 11'den izlenebileceği gibi yukarıdan-aşağıya test birleştirme yöntemine göre elde edilen ortalama RMSE değeri modül uzunluğu ve panel desenine göre anlamlı bir şekilde farklılaşmaktadır ($F_{1-358}(\text{Modül uzunluğu}) = 3379.332$, $p < 0.05$; $F_{2-357}(\text{Panel deseni}) = 404.320$, $p < 0.05$). Eta-kare değerleri de modül uzunluğu ve panel deseninin etkililiğinin ortalama RMSE değeri üzerinde orta düzeyde olduğunu göstermektedir ($\eta^2_{(\text{Modül uzunluğu})} = 0.051$, $\eta^2_{(\text{Panel deseni})} = 0.012$). Panel desenleri arasındaki farklılığın hangi panel desenleri arasında olduğunu tespit etmek için Bonferroni ikili karşılaştırma testi yapılmıştır. Testin sonuçlarına göre "1-2-3" panel deseninin ($\bar{X} = 0.423$), "1-2-2" panel desenine ($\bar{X} = 0.335$) ve "1-

2" panel desenine ($\bar{X} = 0.333$) göre ortalama RMSE değeri üzerinde daha fazla etkili olduğu belirlenmiştir. Bununla birlikte panel deseni-modül uzunluğu ($F_{4-355(P*M)} = 50.648$, $p < 0.05$), panel deseni-örneklem büyüklüğü ($F_{4-355(P*Ö)} = 27.878$, $p < 0.05$), modül uzunluğu-örneklem büyüklüğü ($F_{3-356(M*Ö)} = 10.489$, $p < 0.05$) ve panel deseni-modül uzunluğu-örneklem büyüklüğü ($F_{6-353(P*M*Ö)} = 12.019$, $p < 0.05$) etkileşimlerinin ortalama RMSE değeri üzerindeki etkisi anlamlı çıkmıştır. Ayrıca panel deseni-modül uzunluğu ($\eta^2_{(P*M)} = 0.015$), panel deseni-örneklem büyüklüğü ($\eta^2_{(P*Ö)} = 0.008$), modül uzunluğu-örneklem büyüklüğü ($\eta^2_{(M*Ö)} = 0.001$) ve panel deseni-modül uzunluğu-örneklem büyüklüğü ($\eta^2_{(P*M*Ö)} = 0.005$) ortalama RMSE değerini küçük düzeyde etkilemiştir. Ancak, örneklem büyüklüğü ortalama RMSE değeri üzerinde anlamlı bir şekilde farklılaşmamaktadır.

Tablo 11'den izlenebileceği gibi yukarıdan-aşağıya test birleştirme yöntemine göre elde edilen ortalama yanlılık değeri üzerinde modül uzunluğu, panel deseni ve örneklem büyüklüğünün etkilerine bakıldığında, ortalama yanlılık değerinin yalnızca modül uzunluğuna göre anlamlı bir şekilde farklılaştığı görülmektedir ($F_{1-358(\text{Modül uzunluğu})} = 20.662$, $p < 0.05$). Bu bulguyu eta-kare değeri de desteklemektedir ($\eta^2_{(\text{Modül uzunluğu})} = 0.049$). Ancak, panel deseni ve örneklem büyüklüğü ortalama yanlılık değeri üzerinde anlamlı bir şekilde farklılaşmamaktadır. Bu üç değişkenin birbirleri arasındaki etkileşimlerinin ortalama yanlılık değeri üzerindeki anlamlı etkisine bakıldığında ise panel deseni-örneklem büyüklüğü ($F_{4-355(P*Ö)} = 8.395$, $p < 0.05$) ve panel deseni-modül uzunluğu-örneklem büyüklüğü ($F_{6-353(P*M*Ö)} = 6.059$, $p < 0.05$) etkileşiminin ortalama yanlılık değeri üzerinde anlamlı bir farklılık gösterdiği görülmektedir. Ayrıca bu değişkenlerin ortalama yanlılık üzerindeki etkisi de orta düzeydedir ($\eta^2_{(P*Ö)} = 0.042$, $\eta^2_{(P*M*Ö)} = 0.028$).

Yukarıdan-aşağıya test birleştirme yöntemine göre çeşitli modül uzunluğu, panel deseni ve örneklem büyüklüklerine göre oluşturulan BÇAT sonuçlarından elde edilen bulgulara göre %30, %50 ve %70 geçme oranı ölçütüne dayalı olarak hesaplanan doğru sınıflandırma, yanlış-pozitif hata ve yanlış-negatif hata oranları ve Kappa değerleri Tablo 12'de sunulmuştur.

Tablo 12

Yukarıdan-aşağıya Test Birleştirme Yöntemine Göre Modül Uzunluğu, Panel Deseni ve Örneklem Büyüklüklerine Ait Sınıflandırma Doğruluğunun Karşılaştırılması

Örneklem	250						2000					
	"1-2"		"1-2-2"		"1-2-3"		"1-2"		"1-2-2"		"1-2-3"	
Panel deseni												
Modül uzunluğu	6	12	6	12	6	12	6	12	6	12	6	12
Geçme oranı												
%70												
DS	97.60	97.60	96.80	98.00	98.00	97.60	98.05	98.20	97.90	98.85	95.75	98.85
YPH	0.40	2.00	2.00	0.40	2.00	0.40	0.60	0.55	1.10	0.15	3.05	0.25
YNH	2.00	0.40	1.20	1.60	-	2.00	1.35	1.25	1.00	1.00	1.20	0.90
κ	0.93	0.93	0.91	0.94	0.94	0.93	0.94	0.95	0.94	0.96	0.88	0.96
%50												
DS	96.80	98.80	98.00	98.40	98.40	98.00	98.65	98.85	98.95	98.95	98.60	99.10
YPH	3.20	1.20	0.80	-	-	-	-	1.15	-	-	-	-
YNH	-	-	1.20	1.60	1.60	2.00	1.35	-	1.05	1.05	1.40	0.90
κ	0.93	0.97	0.95	0.96	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.98
%30												
DS	97.60	98.40	96.80	98.00	98.40	97.60	96.90	98.05	97.00	98.25	95.50	98.50
YPH	0.80	0.80	0.40	-	-	-	0.20	0.10	0.20	0.05	0.10	1.50
YNH	1.60	0.80	2.80	2.00	1.60	2.40	2.90	1.85	2.80	1.70	4.40	-
κ	0.92	0.94	0.90	0.93	0.95	0.92	0.91	0.94	0.91	0.95	0.87	0.95

DS=Doğru Sınıflandırma

YPH=Yanlış-Pozitif Hata

YNH=Yanlış-Negatif Hata

κ = Kappa

Tablo 12'den izlenebileceği gibi BÇAT'ı yanıtlayan bireylerin normal dağılımındaki -0.524, 0.00 ve 0.524 yetenek noktalarında belirlenen kesme puanı sırasıyla %70, %50 ve %30 geçme oranları ile karşılaştırılmıştır. Bulgular genel hatlarıyla incelendiğinde %70, %50 ve %30 geçme oranı ölçütleri için doğru sınıflandırma oranları 95.50 ile 99.10 değerleri arasında değiştiği için bu değerlerin yüksek olduğu söylenebilir. %70 geçme oranı ölçütüne göre en düşük doğru sınıflandırma oranı 95.75 değeri ile büyük örnekleme uygulanan modül uzunluğunun kısa olduğu "1-2-3" panel deseninde, en yüksek doğru sınıflandırma oranı 98.85 değeri ile büyük örnekleme uygulanan modül uzunluğunun orta olduğu "1-2-2" ve "1-2-3" panel desenlerinde hesaplanmıştır. %50 geçme oranı ölçütüne göre en düşük doğru sınıflandırma oranı 96.80 değeri ile küçük örnekleme uygulanan modül uzunluğunun kısa olduğu "1-2" panel deseninde, en yüksek doğru sınıflandırma oranı 99.10 değeri ile büyük örnekleme uygulanan modül uzunluğunun orta olduğu "1-2-3" panel deseninde hesaplanmıştır. %30 geçme oranı ölçütüne göre en düşük doğru sınıflandırma oranı 95.50 değeri ile büyük örnekleme uygulanan modül uzunluğunun kısa olduğu "1-2-3" panel deseninde, en yüksek doğru sınıflandırma oranı 98.50 değeri ile büyük örnekleme uygulanan modül uzunluğunun orta olduğu "1-2-3" panel deseninde hesaplanmıştır. Son olarak Kappa katsayısı geçme oranlarında ve tüm koşullarda 0.87 ve 0.98 arasında, yüksek değerler almıştır. Kappa değerleri %50 geçme oranında yüksek değerler alırken, geçme oranı %70 olduğunda birkaç değerde düşüşler meydana gelmiştir. Geçme oranı %30 olduğunda ise Kappa katsayısının bazı değerleri çok az daha düşük değerler üretmiştir.

İkinci Alt Probleme İlişkin Bulgular

"BÇAT uygulamasında aşağıdan-yukarıya test birleştirme yönteminde modül uzunlukları (6 ve 12), panel desenleri ("1-2", "1-2-2" ve "1-2-3") ve örneklem büyüklüklerine (250 ve 2000) göre

a) yetenek kestiriminin kesinliği ve

b) sınıflandırma kararının doğruluğu

nasıl bir değişim göstermektedir?"

İkinci alt probleme ilişkin bulguların yorumlanmasından önce problemde yer alan simülasyon koşullarına göre BÇAT'lar oluşturulmuştur. Bu koşullara göre

oluşturulan BÇAT'ların her bir modülü için kestirilen madde parametreleri EK-D ve EK-E'de verilmiştir.

İkinci alt probleme ilişkin çeşitli simülasyon koşullarına göre oluşturulan BÇAT'ların yetenek kestirimlerine ilişkin bulgular elde edilmiştir. Bu doğrultuda gerçekleştirilen analiz sonuçlarından elde edilen ortalama RMSE ve yanlılık değerleri Tablo 13'te sunulmuştur.

Tablo 13

Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Çeşitli Simülasyon Koşullarına Göre Oluşturulan BÇAT'lara İlişkin Ortalama RMSE ve Yanlılık Değerleri

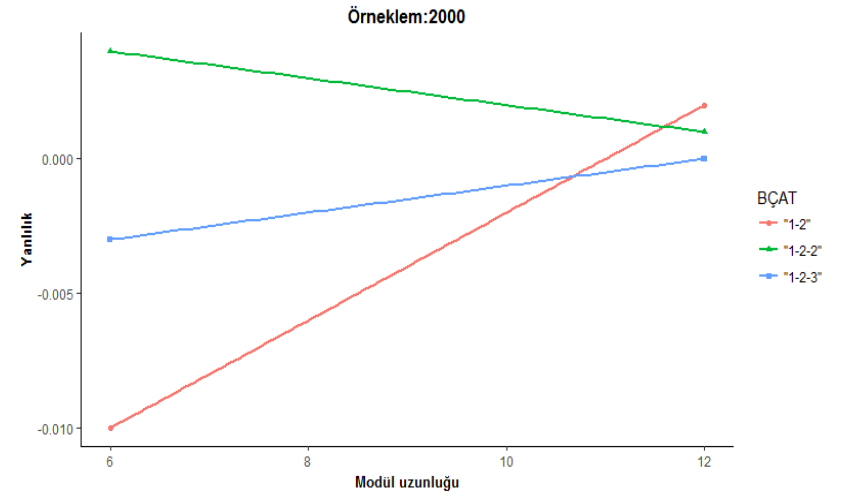
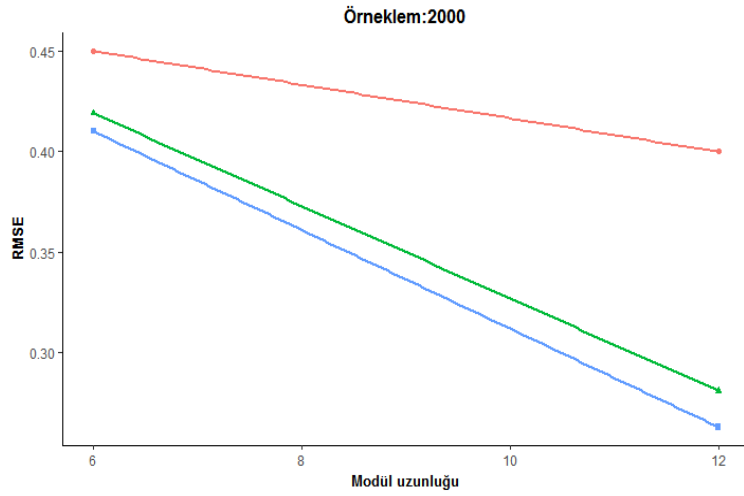
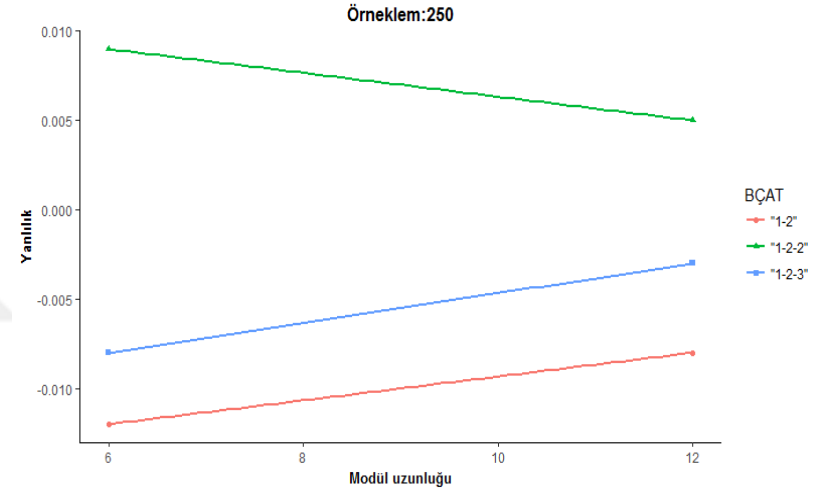
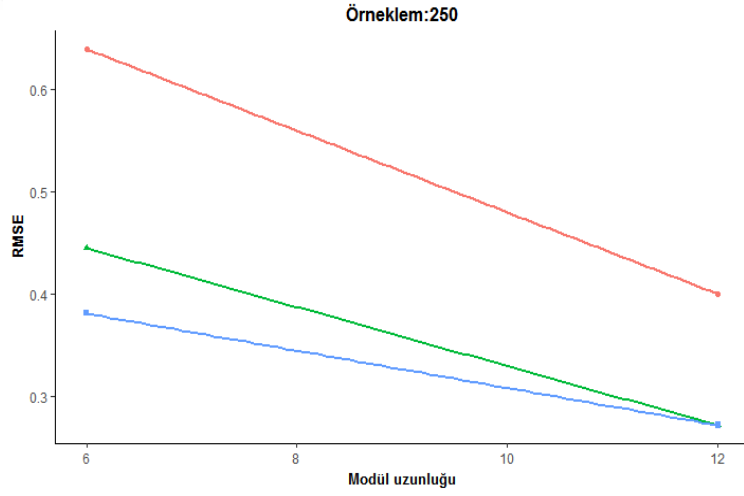
Örneklem	Panel deseni	Modül uzunluğu	RMSE	Yanlılık
250	"1-2"	6	0.639	-0.012
		12	0.400	-0.008
	"1-2-2"	6	0.445	0.009
		12	0.272	0.005
	"1-2-3"	6	0.381	-0.008
		12	0.272	-0.003
2000	"1-2"	6	0.450	-0.010
		12	0.400	0.002
	"1-2-2"	6	0.419	0.004
		12	0.281	0.001
	"1-2-3"	6	0.410	-0.003
		12	0.263	0.000

Tablo 13'ten de izlenebileceği gibi aşağıdan-yukarıya test birleştirme yöntemine göre farklı modül uzunluğu, panel deseni ve örneklem büyüklüğünden elde edilen ortalama RMSE değerleri 0.263 ile 0.639 arasında değişmektedir. En düşük ortalama hata kestirimi orta uzunluktaki modülün büyük örnekleme uygulandığı "1-2-3" panel deseninde, en yüksek ortalama hata kestirimi kısa uzunluktaki modülün küçük örnekleme uygulandığı "1-2" panel deseninde kestirilmiştir. Bulgular modül uzunluğu açısından incelendiğinde de her iki örneklem büyüklüğünde modül uzunluğu arttığında, ortalama RMSE değerinin azaldığı görülmektedir. Bulgular panel deseni açısından incelendiğinde küçük ve büyük örnekleme panel desenlerinin farklılaşmasının her test düzeyindeki ortalama RMSE miktarını değiştirdiği görülmektedir. "1-2" panel deseninden "1-2-2" ve "1-2-3" panel desenine doğru geçişte, ortalama RMSE değerleri düşmüştür. Bulgular örneklem büyüklüğü açısından incelendiğinde örneklem büyüklüğünün

artmasının ortalama RMSE deęerlerini birok kořul iin kltdę grlmektedir. Yalnızca kk rneklemeye uygulanan modl uzunluęunun orta olduęu “1-2-2” panel deseni ile aynı modl uzunluęu ve panel deseninin byk rneklemeye uygulandıęı kořulda 0.09 oranında bir artış gzlenmektedir. Bununla birlikte kk ve byk rneklemeye uygulanan “1-2-3” panel deseninin modl uzunluęunun kısa olduęu kořulda da 0.29’luk bir artış olduęu fark edilmektedir. Kk rneklemde en dřk ortalama RMSE deęeri orta uzunluktaki modln uygulandıęı “1-2-2 ve “1-2-3” panel desenlerinde; byk rneklemde en dřk ortalama RMSE deęeri orta uzunluktaki modln uygulandıęı “1-2-3” panel deseninde hesaplanmıřtır.

Tablo 13’ten ařaęıdan-yukarıya test birleřtirme yntemine gre elde edilen yanlılık ile ilgili bulgular incelendięinde ortalama yanlılık deęerlerinin genel olarak olduka dřk deęerlere sahip oldukları grlmektedir. Ařaęıdan-yukarıya test birleřtirme yntemi seildięinde modl uzunluęu, panel deseni ve rneklem byklę simlasyon kořullarına ait ortalama yanlılık deęerleri -0.012 ile 0.009 arasında deęiřmektedir. En yksek ortalama yanlılık deęerleri kk rneklemeye uygulanan modl uzunluęunun kısa olduęu “1-2” panel desenine aittir. Bu deseni modl uzunluęunun kısa olduęu, byk rneklemeye uygulanan “1-2”, sonrasında modl uzunluęunun kısa olduęu, kk rneklemeye uygulanan “1-2-2” panel desenleri takip etmektedir. En dřk ortalama yanlılık deęeri byk rneklemeye uygulanan modl uzunluęunun orta olduęu “1-2-3” panel deseninde hesaplanmıřtır. Bu kořulda hesaplanan 0.000 ortalama yanlılık deęeri, yansız hesaplama yapıldıęını belirtir. Bulgular modl uzunluęu aısından incelendięinde de modl uzunluęu arttıa, her iki rneklem trnde panel desenlerinin yanlılıęının azaldıęı grlmektedir. Bulgular panel deseni aısından incelendięinde her iki modl uzunluęunda kk ve byk rneklemeye uygulanan “1-2” panel deseninden “1-2-2” ve “1-2-2” panel deseninden “1-2-3” panel desenine geiřte ortalama yanlılık deęerlerinde dřř olduęu grlmektedir. Bulgular rneklem byklę aısından incelendięinde rneklem byklę arttıa, ortalama yanlılık deęerlerinde dřř olduęu grlmektedir.

Tablo 13’te yer alan istatistiklere ait grafikler Őekil 9’da sunulmuřtur.



Şekil 9. Aşağıdan-yukarıya test birleştirme yöntemine göre elde edilen ortalama RMSE ve yanlılık değerlerine ilişkin grafikler.

Şekil 9'dan izlenebileceği gibi küçük ve büyük örneklerde modül uzunluğu arttıkça ortalama RMSE değeri düşmektedir. Büyük örnekleme uygulanan "1-2" panel deseninin ortalama RMSE değerinde en az miktarda düşüşün meydana geldiği görülmektedir. Küçük ve büyük örneklerde "1-2" panel deseninden "1-2-2" ve "1-2-3" panel desenine doğru geçişte, ortalama RMSE değerleri modül uzunluğunun kısa ve orta olduğu her iki durumda da düşmektedir. Ortalama yanlılık değerleri ise küçük ve büyük örnekleme modül uzunluğu arttıkça ve "1-2" panel deseninden, "1-2-2" ve "1-2-3" panel desenlerine doğru geçişte 0.000 değerine yaklaşmaktadır.

Araştırmanın bu alt problemi kapsamında aşağıdan-yukarıya test birleştirme yöntemine göre elde edilen ortalama RMSE ve yanlılık bulguları üzerindeki modül uzunluğu, panel deseni ve örneklem büyüklüğünün istatistiksel olarak anlamlı etkisinin olup olmadığı çok yönlü ANOVA testi ile sınanmıştır. ANOVA testinden elde edilen F değeri ve hesaplanan etki büyüklükleri (η^2) Tablo 14'te sunulmuştur.

Tablo 14

Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Çeşitli Simülasyon Koşulları İçin Oluşturulan BÇAT'lara İlişkin Ortalama RMSE ve Yanlılık Değerlerine Ait ANOVA Sonuçları

Test Özellikleri	sd	Değerlendirme Ölçütleri			Sd	Yanlılık	
		RMSE				F	η^2
		F	η^2		F	η^2	
Modül uzunluğu (M)	1	2721.284*	0.032	1	6.400*	0.016	
Panel deseni (P)	2	1000.355*	0.023	2	22.277*	0.105	
Örneklem (Ö)	1	119.354*	0.001	1	1.786	0.005	
P*M	2	10.654*	0.002	2	7.451*	0.033	
P*Ö	2	140.741*	0.003	2	0.324	0.005	
M*Ö	1	117.107*	0.001	1	0.592	0.005	
P*M*Ö	2	149.044*	0.003	2	4.561*	0.022	

* $p < 0.05$

Tablo 14'ten izlenebileceği gibi aşağıdan-yukarıya test birleştirme yöntemine göre elde edilen ortalama RMSE değeri; modül uzunluğu, panel deseni ve örneklem büyüklüğüne göre anlamlı bir şekilde farklılaşmaktadır ($F_{1-358}(\text{Modül uzunluğu}) = 2721.284$, $p < 0.05$; $F_{2-357}(\text{Panel deseni}) = 1000.355$, $p < 0.05$; $F_{1-358}(\text{Örneklem}) = 119.354$, $p < 0.05$). Ortalama RMSE üzerinde modül uzunluğu ve panel deseninin orta, örneklem büyüklüğünün küçük etkisi vardır ($\eta^2_{(\text{Modül uzunluğu})} = 0.032$, $\eta^2_{(\text{Panel deseni})} = 0.023$, $\eta^2_{(\text{Örneklem})} = 0.001$). Panel desenleri arasındaki farklılığın hangi panel desenleri arasında olduğunu tespit etmek için Bonferroni ikili karşılaştırma

testi yapılmıştır. Testin sonuçlarına göre “1-2-3” panel deseninin ($\bar{X} = 0.472$), “1-2-2” panel desenine ($\bar{X} = 0.356$) ve “1-2” panel desenine ($\bar{X} = 0.332$) göre ortalama RMSE değeri üzerinde daha fazla etkili olduğu belirlenmiştir. Bununla birlikte panel deseni-modül uzunluğu ($F_{4-355(P^*M)} = 10.654$, $p < 0.05$), panel deseni-örneklem ($F_{4-355(P^*Ö)} = 140.741$, $p < 0.05$), modül uzunluğu-örneklem ($F_{3-356(M^*Ö)} = 117.107$, $p < 0.05$) ve panel deseni-modül uzunluğu-örneklem ($F_{6-353(P^*M^*Ö)} = 149.044$, $p < 0.05$) etkileşimlerinin ortalama RMSE değeri üzerindeki etkisi anlamlı çıkmıştır. Bu değişkenlerin ortalama RMSE üzerindeki etkisi küçüktür ($\eta^2_{(P^*M)} = 0.002$, $\eta^2_{(P^*Ö)} = 0.003$, $\eta^2_{(M^*Ö)} = 0.001$, $\eta^2_{(P^*M^*Ö)} = 0.003$).

Tablo 14'ten izlenebileceği gibi aşağıdan-yukarıya test birleştirme yöntemine göre elde edilen ortalama yanlılık değeri üzerinde modül uzunluğu, panel deseni ve örneklem büyüklüğünün etkilerine bakıldığında, ortalama yanlılık değerinin panel deseni ve modül uzunluğuna göre anlamlı bir şekilde farklılaştığı görülmektedir ($F_{1-358}(\text{Modül uzunluğu}) = 6.400$, $p < 0.05$; $F_{2-357}(\text{Panel deseni}) = 22.277$, $p < 0.05$). Ortalama yanlılık üzerinde modül uzunluğunun etkisi küçük ($\eta^2_{(\text{Modül uzunluğu})} = 0.016$), panel deseninin etkisi orta düzeydedir ($\eta^2_{(\text{Panel deseni})} = 0.105$). Panel desenleri arasındaki farklılığın hangi panel desenleri arasında olduğunu tespit etmek için Bonferroni ikili karşılaştırma testi yapılmıştır. Testin sonuçlarına göre “1-2-2” panel deseninin ($\bar{X} = 0.008$), “1-2” panel desenine ($\bar{X} = 0.006$) göre ortalama RMSE değeri üzerinde daha fazla etkili olduğu belirlenmiştir. Ancak, örneklem büyüklüğü ortalama yanlılık değeri üzerinde anlamlı bir şekilde farklılaşmamaktadır. Panel deseni-modül uzunluğu ve panel deseni-modül uzunluğu-örneklem büyüklüğü etkileşimlerinin ortalama yanlılık değeri üzerinde anlamlı bir farklılık gösterdiği görülmektedir ($F_{4-355(P^*M)} = 7.451$, $p < 0.05$; $F_{6-353(P^*M^*Ö)} = 4.561$, $p < 0.05$). Bu değişkenlerin ortalama yanlılık üzerindeki etkisi orta düzeydedir ($\eta^2_{(P^*M)} = 0.033$, $\eta^2_{(P^*M^*Ö)} = 0.022$).

Aşağıdan-yukarıya test birleştirme yöntemine göre çeşitli modül uzunluğu, panel deseni ve örneklem büyüklüklerine göre oluşturulan BÇAT sonuçlarından elde edilen bulgulara göre %30, %50 ve %70 geçme oranı ölçütüne dayalı olarak hesaplanan doğru sınıflandırma, yanlış-pozitif hata ve yanlış-negatif hata oranları ve Kappa değerleri Tablo 15'te sunulmuştur.

Tablo 15

Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Modül Uzunluğu, Panel Deseni ve Örneklem Büyüklüklerine Ait Sınıflandırma Doğruluğunun Karşılaştırılması

Örneklem	250						2000					
	"1-2"		"1-2-2"		"1-2-3"		"1-2"		"1-2-2"		"1-2-3"	
Panel Deseni												
Modül uzunluğu	6	12	6	12	6	12	6	12	6	12	6	12
Geçme oranı												
%70												
DS	94.80	97.20	98.40	97.60	97.20	96.40	97.45	97.45	98.00	98.85	97.80	98.75
YPH	-	2.80	-	1.20	0.80	2.80	1.00	1.15	0.85	0.30	0.45	0.40
YNH	5.20	-	1.60	1.20	2.00	0.80	1.55	1.40	1.15	0.85	1.75	0.85
κ	0.85	0.93	0.95	0.93	0.92	0.90	0.92	0.92	0.94	0.96	0.93	0.96
%50												
DS	99.20	98.00	98.40	96.80	98.00	98.40	98.50	98.35	99.00	99.20	98.20	99.10
YPH	0.80	2.00	-	1.20	-	0.80	-	-	-	-	-	-
YNH	-	-	1.60	2.00	2.00	0.80	1.50	1.65	1.00	0.80	1.80	0.90
κ	0.98	0.95	0.96	0.93	0.96	0.96	0.97	0.96	0.98	0.98	0.96	0.98
%30												
DS	95.20	98.00	96.00	98.00	97.60	98.40	96.20	96.25	97.00	98.50	95.95	98.10
YPH	4.40	1.20	-	-	-	0.40	0.05	-	0.05	0.05	0.25	-
YNH	0.40	0.80	4.00	2.00	2.40	1.20	3.75	3.75	2.95	1.45	3.80	1.90
κ	0.88	0.93	0.88	0.93	0.92	0.95	0.89	0.89	0.91	0.95	0.88	0.94

DS=Doğru Sınıflandırma

YPH=Yanlış-Pozitif Hata

YNH=Yanlış-Negatif Hata

κ = Kappa

Tablo 15'ten izlenebileceği gibi BÇAT'ı yanıtlayan bireylerin normal dağılımındaki -0.524, 0.00 ve 0.524 yetenek noktalarında belirlenen kesme puanı sırasıyla %70, %50 ve %30 geçme oranları ile karşılaştırılmıştır. Bulgular genel hatlarıyla incelendiğinde %70, %50 ve %30 geçme oranı ölçütleri için doğru sınıflandırma oranları 94.80 ile 99.20 değerleri arasında değiştiği için bu değerlerin yüksek olduğu söylenebilir. %70 geçme oranı ölçütüne göre en düşük doğru sınıflandırma oranı 94.80 değeri ile küçük örnekleme uygulanan modül uzunluğunun kısa olduğu "1-2" panel deseninde, en yüksek doğru sınıflandırma oranı 98.85 değeri ile büyük örnekleme uygulanan modül uzunluğunun orta olduğu "1-2-2" panel deseninde hesaplanmıştır. %50 geçme oranı ölçütüne göre en düşük doğru sınıflandırma oranı 96.80 değeri ile küçük örnekleme uygulanan modül uzunluğunun orta olduğu "1-2-2" panel deseninde, en yüksek doğru sınıflandırma oranı 99.20 değeri ile küçük örnekleme uygulanan modül uzunluğunun kısa olduğu "1-2" ve büyük örnekleme uygulanan modül uzunluğunun orta olduğu "1-2-2" panel deseninde hesaplanmıştır. %30 geçme oranı ölçütüne göre en düşük doğru sınıflandırma oranı 95.20 değeri ile küçük örnekleme uygulanan modül uzunluğunun kısa olduğu "1-2" panel deseninde, en yüksek doğru sınıflandırma oranı 98.50 değeri ile büyük örnekleme uygulanan modül uzunluğunun orta olduğu "1-2-2" panel deseninde hesaplanmıştır. Son olarak Kappa katsayısı geçme oranlarında ve tüm koşullarda 0.85 ve 0.98 arasında, yüksek değerler almıştır. Kappa değerleri %50 geçme oranında yüksek değerler alırken, geçme oranı %70 olduğunda birkaç değerde düşüşler meydana gelmiştir. Geçme oranı %30 olduğunda ise Kappa katsayısının bazı değerleri biraz daha düşük değerler üretmiştir.

Bölüm 5

Tartışma, Sonuç ve Öneriler

Bu bölümde araştırmanın bulgularına dayalı olarak yorumlara ve bu yorumlardan yola çıkarak sonuçların özetine ve geliştirilen önerilere yer verilmiştir.

Tartışma

Son yıllarda GRE, CPA (Uniform Certified Public Accountancy) gibi uluslararası standartlara uygun şekilde yapılan geniş ölçekli sınavlarda adını sıklıkla duyduğumuz BÇAT, diğer test türlerine göre farklılık göstermektedir. Bu farklılık BÇAT oluşturma ve uygulama aşamasında ortaya çıkmaktadır. BÇAT, kağıt-kalem testleri ve BBT'den farklı olarak panel, modül, aşama, yol, test birleştirme, puanlama ve yönlendirme gibi birtakım alt bileşenlerden oluşmaktadır. Uygulama aşamasında ise genel olarak bilgisayarın kullanılması ve bireyselleştirilmiş bir test olması BÇAT'ı kağıt-kalem testinden ayırır. Testin amacına uygun olarak bireylerin daha fazla maddeyi yanıtladığının gerekli olması, bireyin yanıtladığı maddelere geri dönebilmesi ve test yürütücüsünün her aşamada testi kontrol edebilmesi de BÇAT'ı BBT'den ayıran özelliklerinin birkaçıdır. Kısaca BÇAT, uyarlanabilirlik, pratiklik, ölçme doğruluğu ve test formları üzerindeki kontrol arasında bir denge oluşturmaktadır (Zenisky, Hambleton & Luecht, 2010, s. 369).

Bu çalışmada BÇAT'ı oluşturan temel bileşenlerin test birleştirme yöntemlerine göre incelenmesine odaklanılmıştır. BÇAT bileşenleri modül uzunluğu, panel deseni ve örneklem büyüklüğüdür. Ancak çalışmada bu bileşenlerin test birleştirme yöntemi ile ilişkisi tespit edilmeye çalışılmıştır. Bu nedenle Luecht ve Nungester'in (1998) alanyazına kazandırdığı ve BÇAT panellerinin birleştirilmesi için önerdiği yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemleri çalışmanın odak noktasını oluşturmuştur. Çalışmanın yöntem bölümünde de ayrıntılı olarak açıklandığı gibi her iki test birleştirme yönteminde öncelikle maddeler tek tek seçilerek modüller oluşturulur. Sonrasında ise modüller birleştirilerek paneller oluşturulur. Ancak iki yöntem arasında modüllerden panellerin oluşturulması aşamasında farklılık bulunmaktadır. Yukarıdan-aşağıya test birleştirme yönteminde modüller seçici olarak eşleştirilerek paneller oluşturulurken, aşağıdan-yukarıya test birleştirme yönteminde modüller serbestçe karıştırılarak ve eşleştirilerek paneller oluşturulur. Yukarıdan-aşağıya

test birleřtirme yöntemi uygulama ařamasında oldukça pratiktir, ancak ikinci ařamada gerekleřtirilen modüllerin seilmesi iřleminden dolayı ařađıdan-yukarıya test birleřtirme yöntemine göre daha karmařıktır. Alanyazın yukarıdan-ařađıya test birleřtirme yöntemi için testin ierik özellikleri gibi kısıtlamalar getirilmesi gerektiđini, ařađıdan-yukarıya test birleřtirme yöntemi için ise bir modülün birleřtirilen formlarının paralel olması gerektiđini önermektedir (Luecht & Nungester, 1998).

Arařtırma bulguları öncelikli olarak yukarıdan-ařađıya ve ařađıdan-yukarıya test birleřtirme yöntemlerinde modül uzunluđunun ortalama hata ve yanlılık deđerlerini etkilediđini göstermiřtir. Her iki test birleřtirme yönteminde de orta modül uzunluđu kısa modül uzunluđuna göre daha düşük ortalama hata ve yanlılık deđerleri üretmiřtir. Kısa ve orta modül uzunluklarına ait hesaplanan ortalama hata ve yanlılık deđerleri arasındaki farkın muhtemel bir nedeni, toplam madde sayısı olabilir. Bu durum, testteki toplam madde sayısı arttıka, ortalama hata ve yanlılık deđerlerinin azaldıđı řeklinde yorumlanabilir. Bu bulgu Sari'nin (2016) ařađıdan-yukarıya test birleřtirme yöntemini tercih ederek test yönetimi, ierik sayısı, test uzunluđu deđiřkenlerine göre BAT ve BBT oluřturduđu; bu iki test türünün performanslarını karřılařtırdıđı alıřma bulguları ile paraleldir. Söz konusu alıřmada ortalama RMSE deđerleri üzerinde sadece test uzunluđunun anlamlı bir etkisi olduđu bulgusuna ulařılmıřtır. Bu bulgu yine Yang'ın (2016) alıřması ile desteklenmektedir. Anılan alıřmada ařađıdan-yukarıya test birleřtirme yöntemi uygulanmıř; test uzunluđu arttıđında, RMSE ve standart hata azalmıřtır. Test uzunluđu 60 olduđunda yanlılık en düşük, test uzunluđu 20 olduđunda ise en yüksek deđer almıřtır. Bu alıřmada da ařađıdan-yukarıya test birleřtirme yöntemine göre elde edilen ortalama yanlılık deđerinin modül uzunluđuna göre anlamlı bir řekilde farklılařtıđı görülmüřtür. Her iki örneklem için modül uzunluđunun kısa olduđu panel desenlerinde en büyük ortalama yanlılık, modül uzunluđunun orta olduđu panel desenlerinde de en küçük ortalama yanlılık deđerleri elde edilmiřtir. Hembry'nin (2014) alıřmasında da ařađıdan-yukarıya test birleřtirme yöntemi kullanılarak oluřturulan BAT'lardaki kısa ve orta olmak üzere iki test uzunluđunun etkisi alıřılmıřtır. Anılan alıřmada ortalama yanlılık ölçümleri sıfıra çok yakındır ve kısa test uzunluđundaki panel desenlerinde ortalama RMSE ve yanlılık deđerleri azalmıřtır. Bu bulgu arařtırma bulgusuyla

paraleldir. Bir diğ er benzer bulgu da Kim, Chung, Park ve Dodd'un (2013) test birleřtirme yöntemi olarak OTB programını kullandıđı, Lynn Chen'in (2010) test birleřtirme yöntemi olarak yukarıdan-ařađıya test birleřtirme yöntemini kullandıđı, Lu'nun (2010) ařađıdan-yukarıya test birleřtirme yöntemini kullandıđı çalıřmalara aittir. Zheng'in (2014) yukarıdan-ařađıya test birleřtirme yöntemini uyguladıđı çalıřmasında da farklı modül uzunlukları arasında tutarlı bir farklılık bulunamamıřtır.

Diğ er yandan yukarıdan-ařađıya ve ařađıdan-yukarıya test birleřtirme yöntemlerinin yanı sıra az sayıda da olsa sezgisel otomatik birleřtirme yöntemlerinden birisi olan NAMSS'in kullanıldıđı BÇAT çalıřmaları da vardır. Bu çalıřmalardan birisi olan Dallas'ın (2014) çalıřmasında 10 ve 20 modül uzunlukları dikkate alınarak oluřturulan BÇAT'larda yönlendirme ve puanlama etkileri çalıřılmıřtır. Çalıřmanın sonucu bu çalıřmada gerç ekleřtirilen yukarıdan-ařađıya ve ařađıdan-yukarıya test birleřtirme yöntemlerine göre incelenen BÇAT'ları etkileyen modül uzunluklarından elde edilen sonuçlarla benzerlik göstermektedir.

Yukarıda sözü edilen çalıřmalarca desteklendiđi gibi ortalama hata ve yanlılık deđerlerinin modül uzunluđundan etkilenmesi ve modül uzunluđu arttıka ortalama hata ve yanlılık deđerlerinin düřmesinin nedeni, kısa testlerden oluřan bir BÇAT'ın daha düřük ölçme hassaslıđına sahip olmasından kaynaklanıyor olabilir. Daha uzun testler daha yüksek sınıflandırma dođruluđu ve tutarlılıđı sađlamaktadır (Crocker & Algina, 1986).

Arařtırmanın bir diğ er bulgusu yukarıdan-ařađıya ve ařađıdan-yukarıya test birleřtirme yöntemlerinde panel deseninin ortalama hata ve yanlılık deđerlerini etkilemesidir. Yukarıdan-ařađıya test birleřtirme yöntemine göre panel deseninin "1-2" den "1-2-2" ve "1-2-3" panel desenine dođru deđiřmesi birç ok kořulda, ařađıdan-yukarıya test birleřtirme yönteminde tüm kořullarda ortalama hata ve yanlılık deđerlerini azaltmıřtır. Bu bulgu, BÇAT panelindeki ařama sayısının artmasının ortalama hata ve yanlılıđı azalttıđı řeklinde yorumlanabilir. Bu bulguyu, Patsula'nın (1999) üç ařamalı ve iki ařamalı BÇAT'ı karřılařtırdıđı ve üç ařamalı BÇAT'ın iki ařamalı BÇAT'dan daha az ölçme hatası ürettiđi çalıřmanın bulgusu desteklemektedir. Ayrıca bu çalıřmadan elde edilen bulgular, Zenisky'nin (2004) 3 PL modele dayanan çalıřmasından elde ettiđi sonuçlarla da tutarlıdır. Bir bařka benzer bulguya Hembry'nin (2014) çalıřmasında rastlanmıřtır. Söz konusu

çalışmada aşağıdan-yukarıya test birleştirme yöntemi kullanılarak oluşturulan BÇAT'larda "1-3", "1-5", "1-3-3" ve "1-5-5" olmak üzere dört panel deseni incelenmiştir. Dört panel deseninin kestirilen yetenekleri ve ortalama yanlılık değerleri arasında oldukça küçük farklılıklar elde edilmiştir. Genel olarak ortalama yanlılık ölçümleri bu çalışmada olduğu gibi sifıra çok yakın değerler almıştır. RMSE değerleri ise bu çalışmanın bulgularından farklı olarak iki aşamalı testlerde daha düşük çıkmıştır. Ancak çalışmada panel desenleri arasındaki farklılığın çok az olduğu bildirilmiştir. Ayrıca bu araştırmada her iki yöntemde göre "1-2", "1-2-2" ve "1-2-3" panel desenleri arasında anlamlı bir farklılık bulunmuş, "1-2-3" panel deseninin diğer desenlere göre ortalama hata ve yanlılık değerlerini daha fazla etkilediği sonucuna ulaşılmıştır. Sari'nın (2016) aşağıdan-yukarıya test birleştirme yöntemine göre gerçekleştirdiği çalışmasında ise farklı bir bulgu elde edilmiştir. Anılan çalışmada iki aşamalı "1-3" ve üç aşamalı "1-3-3" panel desenlerinin RMSE üzerindeki etkisi incelenmiş ve aralarında anlamlı bir farklılığın bulunmadığı rapor edilmiştir. Yang'ın (2016) aşağıdan-yukarıya test birleştirme yöntemini uyguladığı araştırmasında da Sari (2016) ile benzer sonuçlar elde edilmiştir. Bahsedilen bu çalışmada dört panel yapısı ("1-3", "1-5", "1-3-3" ve "1-5-5") incelenmiş ve aralarında anlamlı bir farklılık bulunamamıştır. Söz konusu çalışmalara bir diğer paralel bulgu ise Jodoin, Zenisky ve Hambleton'ın (2006) çalışmasında yer almıştır. Zheng, Nozawa, Gao ve Chang'ın (2012) ve Zheng'in (2014) çalışmasında da yukarıdan-aşağıya test birleştirme yöntemi kullanılmış ve dört aşamalı modeller ve üç aşamalı modeller arasında anlamlı bir farklılığın bulunmadığı bildirilmiştir. Görüldüğü gibi panel desenlerinin BÇAT çalışmaları üzerindeki etkisi hakkında iki farklı sonuç bulunmaktadır. Ortaya çıkan farklı sonuçların muhtemel nedeni her iki çalışmada sabit tutulan diğer değişkenler olabilir. Nitekim panel yapısındaki aşamaların sayısının artması bireyin daha fazla sayıda maddeyi yanıtlamasıyla doğru orantılı olduğundan daha iyi bir ölçme duyarlılığı sağlanabileceği de göz ardı edilmemelidir.

Diğer yandan araştırmaların bulgularına göre yukarıdan-aşağıya test birleştirme yönteminde kısa ve orta modül uzunluğu için "1-2" ve "1-2-2" panel desenlerinde örneklem büyüklüğünün artması ortalama hata ve yanlılık değerlerini düşürmüştür. Aşağıdan-yukarıya test birleştirme yönteminde ise "1-2" panel deseninin kısa ve orta modül uzunluğunda, "1-2-2" panel deseninin kısa ve "1-2-3"

panel deseninin orta modül uzunluğunda örneklem büyüklüğünün artması ortalama hata ve yanlılık değerlerini düşürmüştür. Ayrıca aşağıdan-yukarıya test birleştirme yönteminde örneklem büyüklüğünün ortalama hata ve yanlılık değerleri üzerinde istatistiksel olarak anlamlı bir etkisinin olduğu, yukarıdan-aşağıya test birleştirme yönteminde ise anlamlı bir etkisinin olmadığı sonucuna ulaşılmıştır. Bu bağlamda test birleştirme yöntemleri arasında küçük veya büyük örneklem büyüklüklerinden hangisinin tercih edilmesi gerektiği konusunda kesin bir sonuca ulaşılamamıştır. Aslında BÇAT'ın uluslararası çalışmalarda geniş ölçekli testlerde kullanılması, büyük örnekleme uygulanabilirliğinin bir göstergesidir. Araştırma bulgularından yola çıkılarak aşağıdan-yukarıya test birleştirme yöntemine göre "1-2" panel deseninin kısa ve orta modül uzunluğunda ve "1-2-2" panel deseninin kısa modül uzunluğunda ve "1-2-3" panel deseninin orta modül uzunluğunda büyük örnekleme uygulanabilirliği konusunda bir yorum yapılabilir. Örneklem büyüklüğünün araştırmanın değişkenlerinden birisi olarak seçilmesinin nedeni ise örneklerine geniş ölçekli uluslararası testlerde rastlanılan BÇAT'ın, ileriki zamanlarda uygun bilgisayar alt yapısı gerçekleştirildiği takdirde ülkemizde daha küçük örneklem için gerçekleştirilen müfettiş, uzman, hakimlik gibi kurum sınavları hatta ortaokul ve liselerde uygulanan ders seçme sınavlarında uygulandığında olası sonuçların görülebilmesini sağlamaktır. Küçük örneklem büyüklüğünün incelendiği benzer bir çalışmaya Yan, von Davier ve Lewis (2014) "ağaç tabanlı (tree-based)" yaklaşıma uygun olarak BÇAT'ı inceledikleri çalışmada rastlanmıştır. Çalışmada küçük örnekleme uygulanan BÇAT'ın iyi bir performans gösterdiği sonucuna ulaşılmıştır.

Araştırmanın ortalama hata ve yanlılıktan sonraki bir diğer değerlendirme ölçütü sınıflandırma doğruluğudur. Bu doğrultuda yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemlerinde örneklem büyüklüğü, panel deseni ve modül uzunluğuna göre ayrı ayrı sınıflandırma doğrulukları hesaplanmıştır. Genel olarak tüm koşullarda yüksek sınıflandırma doğrulukları elde edilmiştir. Landis ve Koch (1977), Kappa değerlerini yorumlamak için ' $0 < \kappa < 0.20$, çok zayıf uyum; $0.21 < \kappa < 0.40$, zayıf uyum; $0.41 < \kappa < 0.60$ orta uyum; $0.61 < \kappa < 0.80$ iyi uyum; $0.81 < \kappa < 0.99$ hemen hemen mükemmel uyum' şeklinde bir ölçüt sunmuştur. Bu ölçüte de dayanarak araştırmanın bulguları değerlendirildiğinde Kappa değerlerinin her iki test birleştirme yöntemi seçildiğinde tüm simülasyon koşullarına

göre hemen hemen mükemmel uyum gösterdikleri görülmektedir. Test uzunluğunun artması, birçok simülasyon koşulunda doğru sınıflandırma oranları ve ortalama Kappa değerlerini artırırken; toplam hata oranlarını azaltmıştır. Bu bulgu NAMSS test birleştirme yönteminin kullanıldığı, üç kesme puanına (-0.524, 0.00 ve 0.524) göre BÇAT'dan elde edilen verilerin sınıflandırma doğruluklarının incelendiği Kim'in (2010) araştırma bulgularını desteklemektedir. Bununla birlikte araştırmada genel olarak toplam hata oranlarının, %50 geçme oranı, yani 0.00 kesme puanında en yüksek değerlere ulaştığı saptanmıştır. Bunun nedeni, %50 geçme oranının %30 ve %70 geçme oranları, yani -0.524 ve 0.524 kesme puanları ile karşılaştırıldığında, normal dağılımdan dolayı bireylerin çoğunluğunun 0.00 kesme puanının yakınlarında yığılmasından kaynaklanıyor olabilir. Park, Kim, Chung ve Dodd'un (2017) çalışmasında %30, %50 ve %70 geçme oranına göre BÇAT'ın performansı incelenmiş ve BÇAT test bilgisi olarak adlandırılan yeni yöntemin sınıflandırma doğruluğu incelenmiştir. Söz konusu çalışmada da benzer bulgulara ulaşılmıştır. Bu bulgu; Hambleton ve Xing (2006), Jodoin (2003), Jodoin, Zenisky ve Hambleton (2006), Zenisky (2004), Brossman ve Guile'nin (2014) çalışma sonuçlarıyla da uyumludur.

Sonuç

Madde havuzunun gerçek veri setine göre oluşturulduğu, modül uzunluğu, panel deseni ve örneklem büyüklüklerinin test edildiği BÇAT'ların yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemlerine göre performansları araştırma kapsamında karşılaştırılmıştır. Bunun için değerlendirme ölçütü olarak ortalama RMSE (hata), yanlılık ve sınıflandırma doğruluğu oranları belirlenmiştir. Bu amaç doğrultusunda elde edilen araştırma bulgularına dayalı sonuçlar alt problemlere ilişkin başlıklar halinde sunulmuştur.

Birinci Alt Probleme İlişkin Sonuçlar. Araştırmanın birinci alt probleminde yukarıdan-aşağıya test birleştirme yöntemine göre oluşturulan BÇAT'ların modül uzunluğu, panel deseni ve örneklem büyüklüğü test edilmiştir.

1. Modül uzunluğu arttıkça, ortalama hata değeri azalmıştır.
2. Büyük örnekleme uygulanan kısa ve orta modül uzunluğuna sahip "1-2-2" panel deseninden "1-2-3" panel desenine geçişin dışındaki tüm koşullarda "1-2" panel deseninden "1-2-2" ve "1-2-3" panel desenlerine doğru geçişte ortalama hata değeri düşmüştür.

3. Örneklem büyüklüğünün artması, “1-2” ve “1-2-2” panel desenlerinin kısa ve orta modül uzunluğunda ortalama hata değerlerinde bir miktar düşüş oluşturmuştur.
4. Modül uzunluğunun artması, her iki örneklem türü ve panel deseninde ortalama yanlılığı azaltmıştır.
5. “1-2” panel deseninden “1-2-2” ve “1-2-3” panel desenlerine doğru geçiş ortalama yanlılık değerini düşürmüştür.
6. Örneklem büyüklüğü arttıkça, kestirimlerin ortalama yanlılığında azalma görülmüştür.

Genel olarak tüm koşullarda yüksek sınıflandırma oranları elde edilmiştir. Her iki örneklem büyüklüğü ve modül uzunluğunda “1-2-2” ve “1-2-3” panel desenlerinin tüm geçme oranlarında doğru sınıflandırma oranı oldukça yüksektir. Kappa katsayısı geçme oranlarında ve tüm koşullarda yüksek değerler almıştır. En yüksek Kappa katsayıları geçme oranı ölçütü %50 olduğunda hesaplanmıştır.

İkinci Alt Probleme İlişkin Sonuçlar. Araştırmanın ikinci alt probleminde aşağıdan-yukarıya test birleştirme yöntemine göre oluşturulan BÇAT'ların modül uzunluğu, panel deseni ve örneklem büyüklüğü test edilmiştir.

1. Modül uzunluğu arttıkça ortalama hata azalmıştır.
2. “1-2” panel deseninden “1-2-2” ve “1-2-3” panel desenlerine doğru geçiş ortalama hata değerini düşürmüştür.
3. Örneklem büyüklüğünün artması “1-2” panel deseninin kısa ve orta modül uzunluğunda ve “1-2-2” panel deseninin kısa modül uzunluğunda ve “1-2-3” panel deseninin orta modül uzunluğunda ortalama hata değerini düşürmüştür.
4. Modül uzunluğu arttıkça ortalama yanlılık azalmıştır.
5. “1-2” panel deseninden “1-2-2” ve “1-2-3” panel desenlerine doğru geçiş ortalama yanlılık değerini düşürmüştür.
6. Örneklem büyüklüğü arttıkça ortalama yanlılık değeri azalmıştır.
7. Genel olarak doğru sınıflandırma oranlarının yüksek düzeyde olduğu sonucuna ulaşılmıştır. Geçme oranları ve tüm koşullarda Kappa değerleri,

yüksek değerler almıştır. Kappa değerleri hemen hemen mükemmel uyum olduğuna işaret etmektedir. Bu değerlerin arasında en yüksek Kappa katsayıları geçme oranı ölçütü %50 olduğunda elde edilmiştir.

Öneriler

Bu bölümde araştırma sonuçları doğrultusunda geliştirilen önerilere yer verilmiştir.

Araştırmaya Dönük Öneriler

1. Araştırmada sezgisel test birleştirme yöntemlerinden yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemleri irdelenmiştir. Bundan sonraki çalışmalarda bu yöntemlere ek olarak ASM, NAMSS ve maksimum öncelikli endeks gibi diğer sezgisel yöntemler, doğrusal programlama yöntemleri ve sınav anında test birleştirmenin yapıldığı alanyazında adı “on-the-fly” test birleştirme yöntemi olarak geçen test birleştirme yönteminin çalışılması önerilebilir.
2. Araştırmada madde havuzu gerçek veri setine göre oluşturulduğu için sadece 2 PL modele uygun olarak madde ve yetenek parametreleri kestirilmiş, buna göre BÇAT’lar oluşturulmuştur. Bunda sonraki çalışmalarda 2 PL ve 3 PL modele göre parametreler kestirilerek lojistik modellerin BÇAT’ın performansı üzerindeki etkisi araştırılabilir.
3. Araştırmada kullanılan madde havuzu tek boyutludur, dolayısıyla tek boyutlu BÇAT’lar oluşturulmuştur. Bundan sonraki çalışmalarda alanyazında da sayısının çok az olduğu çok boyutlu BÇAT çalışılabilir.
4. Araştırmada madde havuzu gerçek veri setine göre oluşturulmuş ve toplamda 159 madde ile sınırlı kalmıştır. Bundan sonraki çalışmalarda çok daha büyük bir madde havuzu ile çalışılarak madde kullanım sıklığı da kontrol edilebilir.
5. Araştırmanın bağımsız değişkenlerini örneklem büyüklüğü, panel deseni ve modül uzunluğu oluşturmaktadır. Ancak yönlendirme yöntemi, madde kullanım sıklığı gibi BÇAT bileşenleri de araştırmaya dahil edilerek test birleştirme yöntemleri üzerinde bir etki oluşturup oluşturmadıkları araştırılabilir.
6. Araştırmada iki aşamalı panel desenine örnek olarak “1-2” panel deseni, üç aşamalı panel desenine örnek olarak da “1-2-2” ve “1-2-3” panel desenleri

çalışılmıştır. Bundan sonraki çalışmalarda alanyazında rastlanılan “1-3”, “1-3-3”, “1-3-5” gibi panel desenleri de çalışılabilir.

7. Araştırmada BÇAT modüllerinde yer alan maddeler iki kategorili maddelerden oluşmaktadır. Bundan sonraki çalışmalarda çok kategorili maddelerden ve hatta karma testlerden oluşturulan BÇAT’lar çalışılabilir.
8. Araştırmada sınıflandırma doğruluğu öğrencilerin ‘geçme/kalma’ durumlarını bildiren iki kategoriye göre incelenmiştir. Dolayısıyla bundan sonraki çalışmalarda BÇAT’lardan elde edilen bulgular ikiden fazla sınıflandırma kategorisi kullanılarak da sınıflandırılabilir.
9. Araştırma bulguları bir diğer bireyselleştirilmiş test olan BBT’den elde edilecek bulgular ile karşılaştırılabilir.

Uygulamaya Dönük Öneriler

1. Araştırmada yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemlerinin benzer bulgular ürettiği ve her iki yöntemin de BÇAT’ın oluşturulmasında kullanılabileceği önerilmektedir.
2. Araştırmada örneklem büyüklüğünün birkaç simülasyon koşulunda yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemlerinin ortalama hata değerlerinde düşüş oluşturduğu, yalnızca aşağıdan-yukarıya test birleştirme yöntemi üzerinde anlamlı bir etkisinin olduğu bulgusuna ulaşılmıştır. Benzer araştırma sonucu da dikkate alınarak, bundan sonraki çalışmalarda örneklem büyüklüğü doğrudan hata ve yanlılığı etkileyebilecek bir değişken olarak düşünülüyorsa BÇAT değişkenlerinden birisi olarak ele alınmayabilir.
3. Araştırmada yukarıdan-aşağıya ve aşağıdan-yukarıya test birleştirme yöntemlerinde modül uzunluğu ve panel deseninin ortalama hata ve yanlılık değerlerini etkilediği görülmüştür. Bu nedenle bu iki test birleştirme yöntemlerinden birisine göre oluşturulacak BÇAT’ın yapısında en az hata ve yanlılık değerlerini üreten “1-2-3” panel deseninin ve “12” modül uzunluğunun kullanılması önerilmektedir. Diğer yandan uygulaması yapılacak olan bir BÇAT için ortalama hata ve yanlılık değerlerinin düşük olması sebebiyle 1-2” ve “1-2-2” panel desenlerinin ve “6 ve 12” modül uzunluklarının da kullanılması önerilmektedir.

Kaynaklar

- Adema, J. J. (1990). The construction of customized two-stage tests. *Journal of Educational Measurement, 27*(3), 241-253.
- Andrew, D. (2014). *The effects of routing and scoring within a computer adaptive multi-stage framework*. (Unpublished Doctoral Dissertation). The University of North Carolina.
- Ariel, A., Veldkmap, Bernard P., & Breithaupt, K. (2006). Optimal testlet pool assembly for multistage testing designs. *Applied Psychological Measurement, 30*, 204-215.
- Armstrong, R. D. (2002). *Routing rules for multiple-form structures*. (Computerized Testing Report 02-08). Newtown: Law School Admission Council.
- Armstrong, R. D., Jones, D. H., Koppel, N. B., & Pashley, P. J. (2004). Computerized adaptive testing with multiple-form structures. *Applied Psychological Measurement, 28*(3), 147-164.
- Armstrong, R. D., & Roussos, L. (2005). *A method to determine targets for multi-stage adaptive tests*. (Research Report 02-07). Newtown, PA: Law School Admissions Council.
- Baker, F. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.
- Belov, D. I. (2016). *Review of modern methods for automated test assembly and item pool analysis*. (Research Report 16-01). Newtown, PA: Law School Admissions Council.
- Belov, D. I., & Armstrong, R. D. (2005). Monte Carlo test assembly for item pool analysis and extension. *Applied Psychological Measurement, 29*, 239-61.
- Berger, M. P. F. (1994). A general approach to algorithmic design of fixed-form tests, adaptive tests and testlets. *Applied Psychological Measurement, 18*, 141-153.
- Bergstrom, B. A., & Lunz, M. E. (1999). CAT for certification and licensure. In F. Drasgow and J. B. Olson-Buchanan (Eds.). *Innovations in computerized assessment*, 67-91. Mahwah, NJ: Lawrence Erlbaum Associates.

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.). *Statistical theories of mental scores*, 397-479. Reading MA: Addison-Wesley.
- Breithaupt, K., Ariel, A., & Veldkamp, B. (2005). Automated simultaneous assembly for multistage testing. *International Journal of Testing*, 5(3), 319-330.
- Breithaupt, K., & Hare, D. R. (2007). Automated simultaneous assembly of multistage testlets for a high-stakes licensing examination. *Educational and Psychological Measurement*, 64, 1-16.
- Brossman, B. G., & Guille, R. A. (2014). A comparison of multi-stage and linear test designs for medium-size licensure and certification examinations. *Journal of Computerized Adaptive Testing*, 2(2). Doi:10.7333/1402-020218.
- Chalhoub-Deville, M. (Ed.). (1999). *Issues in computer adaptive testing of reading proficiency: Studies in Language Testing*. Cambridge: Cambridge University Press.
- Chen, L. Y. (2010). *An investigation of the optimal test design for multistage test using the generalized partial credit model*. (Unpublished Doctoral Dissertation). The University of Texas at Austin.
- Cheng, Y., & Chang, H. H. (2009). The maximum priority index method for severely constrained item selection in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 62, 369-383.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.
- Colvin, K. F. (2014). Effect of automatic item generation on ability estimates in multistage test. (Unpublished Doctoral Dissertation). University of Massachusetts-Amherst.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: CBS College Publishing.

- Cronbach, L. J., & Glaser, G. C. (1965). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Dallas, A. (2014). *The effects of routing and scoring within a computer adaptive multi-stage framework*. (Unpublished Doctoral Dissertation). The University of North Carolina.
- Davis, L. L., & Dodd, B. G. (2003). Item exposure constraints for testlets in the verbal reasoning section of the MCAT. *Applied Psychological Measurement*, 27(5), 335-356.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- Edmonds, J., & Armstrong, R. (2009). A mixed integer programming model for multiple stage adaptive testing. *European Journal of Operational Research*, 193(1), 342-350.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Erkuş, A. (1999). *Ölçme araçlarının tutarlı ölçme ve sınıflama yapıp yapmadığını belirlemeye ilişkin bir çalışma*. (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Feuerman, M., & Weiss, H. (1973). A mathematical programming model for test construction and scoring. *Management Science*, 19, 961-66.
- Georgiadou, E., Triantafillou, E. & Economides, A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *Journal of Technology, Learning and Assessment*, 5(8).
[Çevrim-içi:
<https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1647/1482>, Erişim tarihi:10 Haziran 2016.]
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff Publishing.

- Hambleton, R. K., & Swaminathan, H. (1989). *Item response theory: Principles and applications*. USA: Kluwer Nijhoff Publishing.
- Hambleton, R. K., Swaminathan H., & Rogers H. J. (1991). Fundamentals of item response theory. *Measurement Methods for the Social Sciences Series*. Sage Publications, Inc.
- Hambleton, R. K., & Xing, D. (2006). Optimal and nonoptimal computer-based test designs for making pass-fail decisions. *Applied Measurement in Education*, 19(3), 221-239.
- Hembry, I. F. (2014). *Operational characteristics of mixed format multistage tests using the 3PL testlet response theory model*. (Unpublished Doctoral Dissertation). The University of Texas at Austin.
- Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice*, 26(2), 44-52.
- International Consortium PISA 2015 (2014). *PISA 2015 field trial analysis plan for questionnaire data: EDU/PISA/GB (2014) 3*. Paris: OECD. Unpublished report.
- Jodoin, M. G. (2003). *Psychometric properties of several computer-based test designs with ideal and constrained item pool*. (Unpublished Doctoral Dissertation). University of Massachusetts-Amherst.
- Jodoin, M. G., Zenisky, A., & Hambleton, R. K. (2006). Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education*, 19(3), 203-220.
- Kelecioğlu, H. (2001). Örtük özellikler teorisindeki b ve a parametreleri ile klasik test teorisindeki p ve r istatistikleri arasındaki ilişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20, 104-110.
- Keng, L. (2008). *A comparison of the performance of testlet based computer adaptive tests and multistage tests*. (Unpublished Doctoral Dissertation). The University of Texas at Austin.

- Kim, J. (2010). *A comparison of computer-based classification testing approaches using mixed-format tests with the generalized partial credit model*. (Unpublished Doctoral Dissertation). The University of Texas at Austin.
- Kim, J., Chung, H., Dodd, B. G., & Park, R. (2012). Panel design variations in the multistage test using the mixed-format tests. *Educational and Psychological Measurement, 72*(4), 574-588. Doi:10.1177/0013164411428977.
- Kim, J., Chung, H., Park, R., & Dodd, B. G. (2013). A comparison of panel designs with routing methods in the multistage test with the partial credit model. *Behaviour Research, 45*, 1087-1098. Doi:10.3758/s13428-013-0316-3.
- Kim, H., & Plake, B. S. (1993). *Monte Carlo simulation comparison of two-stage testing and computerized adaptive testing*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Atlanta, GA.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.
- Lewis, C., & Sheehan, K. M. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement, 14*, 367-386.
- Lord, F. M. (1971). A theoretical study of two-stage testing. *Psychometrika, 36*, 227-242.
- Lord, F. M. (1974). Practical methods for redesigning a homogeneous test, also for designing a multilevel test. *Educational Testing Service RB, 74*-30.
- Lord, F. M. (1977). Practical applications of item characteristic curve theory. *Journal of Educational Measurement, 14*, 117-38.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Loyd, B. H. (1984). *Efficiency and precision in two-stage adaptive testing*. Paper presented at the Annual Meeting of the Eastern Educational Research Association, West Palm Beach, Florida.

- Lu, R. (2010). *Impacts of local item dependence of testlet items with the multistage tests for pass-fail decisions*. (Unpublished Doctoral Dissertation). University of Maryland.
- Luecht, R. M. (1998). Computer-assisted test assembly using optimization heuristics. *Applied Psychological Measurement, 22*(3), 224-236.
- Luecht, R. (2000). *Implementing the CAST framework to mass produce high quality computeradaptive and mastery tests*. Paper presented at the Annual Meeting of the National Council on Measurement in Education (NCME), New Orleans, LA.
- Luecht, R. (2003). *Exposure control using adaptive multi-stage item bundles*. Paper presented at the Annual Conference of Meeting of the National Council on Measurement in Education (NCME), Chicago, IL.
- Luecht, R. M. (2006). *Designing tests for pass-fail decisions using item response theory*. In S. Downing & T. Haladyna (Eds.), *Handbook of test development*, 575-596. Mahwah, NJ: Lawrence Erlbaum Associates.
- Luecht, R., Brumfield, T., & Breithaupt, K. (2006). A testlet assembly design for adaptive multistage tests. *Applied Measurement in Education, 19*(3), 189-202.
- Luecht, R. M., & Nungester, R. J. (1998). Some practical examples of computer-adaptive sequential testing. *Journal of Educational Measurement, 35*(3), 229-249.
- Luecht, R. M., & Nungester, R. J. (2000). Computer-adaptive sequential testing. In C. Glas & W. J. van der Linden (Eds.), *Computer-Adaptive Testing*, 117-128. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Luecht, R. M., Nungester, R. J., & Hadidi, A. (1996). *Heuristic-based CAT: Balancing item information, content and exposure*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.
- Luo, X. (2017). *Package 'xxIRT'*. (Version 2.0.3). [Çevrim-içi: <https://cran.r-project.org/web/packages/xxIRT/xxIRT.pdf>, Erişim tarihi: 15 Ocak 2017.]

- Lynn Chen, L. Y. (2010). *An investigation of the optimal test design for multi-stage test using the generalized partial credit model*. (Unpublished Doctoral Dissertation). The University of Texas at Austin.
- Macken-Ruiz, C. L. (2008). *A comparison of multi-stage and computerized adaptive tests based on the generalized partial credit model*. (Unpublished Doctoral dissertation). The University of Texas at Austin.
- Magis, D., Yan, D., & von Davier, A. A. (2016). *Computerized adaptive and multistage testing with R*. New York: Springer.
- Mead, A. D. (2006). An introduction to multistage testing. *Applied Measurement in Education*, 19(3), 185-187.
- MEB. (2010). "PISA 2009 ulusal ön raporu", *Milli Eğitim Bakanlığı-EARGED*. [Çevrim-içi: <http://earged.meb.gov.tr/dosyalar/pisa/pisa2009rapor.pdf>, Erişim tarihi: 18 Kasım 2016.]
- Nydick, S. W., & Weiss, J. W. (2009). A hybrid simulation procedure for the development of CATs. In D. J. Weiss (Ed.). *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*, 1-23. [Çevrim-içi: <http://iacat.org/content/hybrid-simulation-procedure-development-cats>, Erişim tarihi: 28 Mayıs 2016.]
- OECD (2005). *PISA 2003 technical report*. [Çevrim-içi: <http://www.oecd.org/dataoecd/49/60/35188570.pdf>, Erişim tarihi: 15 Eylül 2016.]
- OECD (2015). *PISA 2015 technical report*. [Çevrim-içi: <http://www.oecd.org/pisa/sitedocument/PISA-2015-Technical-Report-Chapter-1-Programme-for-International-Student-Assessment-an-Overview.pdf>, Erişim tarihi: 08 Temmuz 2016.]
- OECD (2017). *PISA 2015 Technical Report*. [Çevrim-içi: <http://www.oecd.org/pisa/sitedocument/PISA-2015-Technical-Report-Chapter-9-Scaling-PISA-Data.pdf>, Erişim tarihi: 16 Eylül 2016.]
- Papadimitriou, C. H., & Steiglitz, K. (1982). *Combinatorial optimization: Algorithms and complexity*. Englewood Cliffs, NJ: Prentice-Hall.

- Park, R. (2015). *Investigating the impact of a mixed-format item pool on optimal test designs for multistage testing*. (Unpublished Doctoral Dissertation). The University of Texas at Austin.
- Park, R., Kim, J., Chung, H., & Dodd, B. G. (2014). Enhancing pool utilization in constructing the multistage test using mixed-format tests. *Applied Psychological Measurement*, 38(4), 268-280.
- Park, R., Kim, J., Chung, H., & Dodd, B. G. (2017). The development of MST test information for the prediction of test performances. *Applied Psychological Measurement*, 77(4), 570-586.
- Parshall, C. G., Spray, J. C., Kalohn, J., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.
- Patsula, L. N. (1999). *A comparison of computerized-adaptive testing and multi-stage testing*. (Unpublished doctoral dissertation). University of Massachusetts at Amherst.
- Patsula, L. N., & R. K. Hambleton. (1999). *A comparative study of ability estimates from computer-adaptive testing and multi-stage testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Quebec.
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. (Version 3.0.1). Vienna, Austria: R Foundation for Statistical Computing. [Çevrim-içi: <http://www.R-project.org/>, Erişim tarihi: 14 Ocak 2015.]
- Rasch, G. (1960). *Probabilistic Model for Some Intelligence and Achievement Tests*. Copenhagen: Danish Institute for Educational Research.
- Rudner, L. M. (1998). Item banking. *Practical Assessment, Research & Evaluation*, 6. [Çevrim-içi: <http://ericae.net/pare/getvn.asp?v=6&n=4>, Erişim tarihi: 15 Temmuz 2017.]
- Sari, H. İ. (2016). *Examining content control in adaptive tests: Computerized adaptive testing vs. computerized multistage testing*. (Unpublished doctoral dissertation). University of Florida.

- Sari, H. I., Yahsi-Sari, H., & Huggins-Manley, A. C. (2016). Computer adaptive multistage testing: Practical issues, challenges and principles. *Journal of Measurement and Evaluation in Education and Psychology*, 7, 388-406.
- Schnipke, D. L., & Reese, L. M. (1997). *A comparison of testlet-based test designs for computerized adaptive testing*. Paper presented at the meeting of American Educational Research Association, Chicago, IL.
- Sinharay, S., & Johnson, M. S. (2008). Use of item models in a large-scale admissions test: A case study. *International Journal of Testing*, 8(3), 209-236
- Slater, S. C. (2001). *Pretest item calibration within the computerized adaptive testing environment*. (Unpublished Doctoral Dissertation). Graduate School of the University Massachusetts-Amherst.
- Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion referenced tests: A decision-theoretic formulation. *Journal of Educational Measurement*, 11, 263-267.
- Swanson, L., & Stocking, M. L. (1993). A model and heuristic for solving very large item selection problems. *Applied Psychological Measurement*, 17, 151-166.
- Tay, P. H. (2015). *On-the-fly assembled multistage adaptive testing*. (Unpublished Doctoral Dissertation). University of Illinois at Urbana-Champaign.
- Theunissen, T. J. J. M. (1985). Binary programming and test design. *Psychometrika*, 50(4), 411-420. [Çevrim içi:<https://link.springer.com/article/10.1007/BF02296260>, Erişim tarihi: 24 Aralık 2017.]
- Theunissen, T. J. J. M. (1986). Some applications of optimization algorithms in test design and adaptive testing. *Applied Psychological Measurement*, 10, 381-389.
- Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement*, 26(3), 247-260.
- Thompson, N. A. (2009). Item selection in computerized classification testing. *Educational and Psychological Measurement*, 69, 778-93.

- van der Linden, W. J. (1998). Optimal assembly of psychological and educational tests. *Applied Psychological Measurement, 22*, 195-211.
- van der Linden, W. J. (2005). *Linear models of optimal test design*. New York, NY: Springer.
- van der Linden, W. J., & Boekkooi-Timminga, E. (1989). A maximin model for IRT-based test design with practical constraints. *Psychometrika, 54*(2), 237-247. [Çevrim-içi: 10.1007/bf02294518, Erişim tarihi: 01 Aralık 2017.]
- Veldkamp, B. P. (1999). Multiple-objective test assembly problems. *Journal of Educational Measurement, 36*, 253-66.
- Vispoel, W. P. (1998). Psychometric characteristics of computer-adaptive and self-adaptive vocabulary tests: The role of answer feedback and anxiety. *Journal of Educational Measurement, 35*, 155-67.
- Wainer, H., & Kiely, G. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24*, 185-202.
- Wang, X. (2013). *An investigation on computer-adaptive multistage testing panels for multidimensional assessment*. (Unpublished Doctoral Dissertation). The University of North Carolina at Greensboro.
- Wang, K. (2017). *Fair comparison of the performance of computerized adaptive testing and multistage adaptive testing*. (Unpublished Doctoral Dissertation). Michigan State University.
- Weiss, D. J. (1973). *The stratified adaptive computerized ability test (Research Report No. 73-3)*. Minneapolis, MN: University of Minnesota, Department of Psychology. Psychometric Methods Program.
- Weissman, A. (2003). *Assessing the Efficiency of Item Selection in Computerized Adaptive Testing*. Paper presented at the Annual Meeting of the American Educational Research Association, April, Chicago. [Çevrim-içi: <https://eric.ed.gov/?id=ED478985>, Erişim tarihi: 3 Ağustos 2016.]
- Wendler, C., & Walker, M. E. (2006). Practical issues in designing and maintaining multiple test forms for large-scale programs. In S. Downing & T. Haladyna (Eds.). *Handbook of test development*, 445-467. Hillsdale, NJ: Erlbaum.

- Wise, S. L., & Kingsbury, G. G. (2000). Practical issues in developing and maintaining a computerized adaptive testing program. *Psicologica*, 21, 135-155.
- Xing, D., & Hambleton, R. K. (2004). Impact of test design, item quality, and item bank size on the psychometric properties of computer-based credentialing examinations. *Educational and Psychological Measurement*, 64, 5-21.
- Yan, D., Lewis, C., & von Davier, A. (2014). Overview of computerized multistage tests. In D. Yan, A. A. von Davier, & C. Lewis (Eds.). *Computerized Multistage Testing: Theory and Applications*, 3-20. London, England: Chapman & Hall.
- Yan, D., von Davier, A. A., & Lewis, C. (2014). *Computerized Multistage Testing: Theory and Applications*. London, England: Chapman & Hall.
- Yang, L. (2016). *Enhancing item pool utilization when designing multistage computerized adaptive tests*. (Unpublished Doctoral Dissertation). Michigan State University.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187-214.
- Yıldırım, H. H., Çömlekoğlu, G., & Berberoğlu, G. (2003). Milli Eğitim Bakanlığı özel okullar sınavı verilerinin madde tepki kuramı modellerine uyumu. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 24, 159-168.
- Zenisky, A. (2004). *Evaluating the effects of several multistage testing design variables on selected psychometric outcomes for certification and licensure assessment*. (Unpublished doctoral dissertation). University of Massachusetts at Amherst.
- Zenisky, A., & Hambleton, R. (2014). Multistage test designs: Moving research results into practice. In Yan, D., Von Davier, A., & Lewis, C. (Eds.), *Computerized Multistage Testing: Theory and Applications*, 21-36. London, England: Chapman & Hall.
- Zenisky, A., Hambleton, R. K. & Luecht, R. M. (2010). Multistage testing: Issues, designs and research. In: der Linden, W. J. & Glas, C. A. W. (Eds.). *Elements of adaptive testing*. New York: Springer, 355-372.

- Zheng, Y. (2014). *New methods of online calibration for item bank replenishment*. (Unpublished Doctoral Dissertation). University of Illinois at Urbana-Champaign.
- Zheng, Y., Nozawa, Y., Gao, X., & Chang, H. (2012). *Multistage adaptive testing for a large-scale classification test: the designs, heuristic assembly, and comparison with other testing modes*. Paper presented at the Annual Meeting of the National Council on Measurement in Education (NCME) (ACT Research Reports 2012-6). Vancouver, British Columbia, Canada.
- Zheng, Y., Nozawa, Y., Zhu, R., & Gao, X. (2016) Automated top-down heuristic assembly of a classification multistage test. *Int. J. Quantitative Research in Education*, 3(4), 242–265.
- Zheng, Y., Wang, C., Culbertson, M., & Chang, H-H. (2014). Test assembly in computerized multistage testing. In Yan, D., von Davier, A. A. and Lewis, C. (Eds.). *Computerized Multistage Testing: Theory and Applications*, 87-98. London, England: Chapman & Hall.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items*. [Computer software]. Chicago: Scientific Software International.

EK-A: PISA 2015 Yılı Uygulamasına Ait Veri Setinin Faktör Yükleri

Madde	Faktör Yükü	Madde	Faktör Yükü	Madde	Faktör Yükü	Madde	Faktör Yükü	Madde	Faktör Yükü
M1	0.622	M34	0.596	M67	0.632	M100	0.407	M133	0.743
M2	0.708	M35	0.642	M68	0.548	M101	0.648	M134	0.518
M3	0.718	M36	0.691	M69	0.516	M102	0.769	M135	0.664
M4	0.598	M37	0.419	M70	0.557	M103	0.740	M136	0.815
M5	0.791	M38	0.586	M71	0.554	M104	0.512	M137	0.580
M6	0.755	M39	0.720	M72	0.527	M105	0.506	M138	0.610
M7	0.785	M40	0.723	M73	0.771	M106	0.793	M139	0.599
M8	0.545	M41	0.700	M74	0.576	M107	0.490	M140	0.620
M9	0.781	M42	0.715	M75	0.467	M108	0.596	M141	0.820
M10	0.538	M43	0.725	M76	0.598	M109	0.541	M142	0.807
M11	0.538	M44	0.664	M77	0.727	M110	0.418	M143	0.612
M12	0.784	M45	0.704	M78	0.792	M111	0.767	M144	0.594
M13	0.794	M46	0.464	M79	0.761	M112	0.798	M145	0.814
M14	0.412	M47	0.652	M80	0.624	M113	0.605	M146	0.509
M15	0.388	M48	0.722	M81	0.724	M114	0.795	M147	0.419
M16	0.794	M49	0.502	M82	0.456	M115	0.238	M148	0.094
M17	0.742	M50	0.358	M83	0.455	M116	0.509	M149	0.285
M18	0.737	M51	0.417	M84	0.645	M117	0.457	M150	0.565
M19	0.752	M52	0.539	M85	0.443	M118	0.696	M151	0.231
M20	0.640	M53	0.641	M86	0.406	M119	0.503	M152	0.232
M21	0.571	M54	0.523	M87	0.328	M120	0.404	M153	0.434
M22	0.723	M55	0.616	M88	0.418	M121	0.798	M154	0.239
M23	0.764	M56	0.484	M89	0.252	M122	0.800	M155	0.369
M24	0.620	M57	0.611	M90	0.166	M123	0.475	M156	0.554
M25	0.608	M58	0.642	M91	0.348	M124	0.770	M157	0.448
M26	0.587	M59	0.653	M92	0.662	M125	0.412	M158	0.109
M27	0.559	M60	0.555	M93	0.207	M126	0.597	M159	0.576
M28	0.612	M61	0.648	M94	0.317	M127	0.689		
M29	0.736	M62	0.498	M95	0.196	M128	0.492		
M30	0.377	M63	0.360	M96	0.537	M129	0.272		
M31	0.428	M64	0.359	M97	0.578	M130	0.277		
M32	0.649	M65	0.325	M98	0.666	M131	0.662		
M33	0.342	M66	0.671	M99	0.347	M132	0.597		

**EK-B. PISA 2015 Yılı Uygulamasına Ait Veri Setinin 2 PL Modelde Kestirilen
Madde Parametreleri**

Madde	A	b	Madde	a	b	Madde	a	B
M1	1.047	0.565	M32	1.355	-0.437	M63	0.485	0.814
M2	1.665	0.625	M33	0.489	0.964	M64	0.467	1.191
M3	1.820	-0.4366	M34	1.165	0.761	M65	0.532	1.526
M4	1.009	0.644	M35	1.330	-0.467	M66	1.222	0.279
M5	1.854	-0.014	M36	1.363	0.403	M67	1.652	-0.666
M6	1.775	0.434	M37	1.048	-1.294	M68	1.317	0.986
M7	1.843	-0.091	M38	1.247	0.855	M69	0.964	0.830
M8	0.927	-0.714	M39	1.403	0.163	M70	1.178	0.865
M9	1.736	-0.091	M40	1.591	0.555	M71	1.014	0.690
M10	0.940	0.937	M41	1.365	0.249	M72	1.367	-0.947
M11	0.933	0.901	M42	1.436	0.216	M73	1.757	0.160
M12	1.833	-0.111	M43	1.490	0.340	M74	0.990	0.762
M13	1.938	-0.059	M44	1.374	-0.348	M75	1.133	1.314
M14	1.184	-1.366	M45	1.419	-0.104	M76	1.492	-0.742
M15	1.119	-1.438	M46	0.770	-0.833	M77	1.635	0.434
M16	1.903	-0.068	M47	1.343	-0.373	M78	1.976	-0.007
M17	1.466	-0.201	M48	1.473	-0.072	M79	1.806	0.315
M18	1.510	-0.232	M49	0.830	0.808	M80	1.611	0.836
M19	1.487	-0.158	M50	0.765	-1.463	M81	1.644	-0.244
M20	1.078	0.486	M51	0.744	-0.977	M82	0.677	0.973
M21	0.899	-0.281	M52	1.047	-0.740	M83	0.808	1.329
M22	1.344	0.409	M53	1.146	0.411	M84	0.944	0.191
M23	1.484	0.079	M54	0.988	0.884	M85	0.547	-0.108
M24	1.268	-0.640	M55	1.210	-0.520	M86	0.485	-0.473
M25	1.304	0.911	M56	0.964	1.052	M87	0.435	1.418
M26	1.190	0.938	M57	1.137	-0.466	M88	0.597	1.062
M27	1.195	-0.827	M58	1.082	0.034	M89	0.312	1.132
M28	1.230	0.870	M59	1.131	0.002	M90	0.221	-0.420
M29	1.520	-0.257	M60	1.176	-0.835	M91	0.644	-1.639
M30	0.899	-1.467	M61	1.103	-0.189	M92	1.013	-0.205
M31	1.001	-1.247	M62	0.771	-0.370	M93	0.296	-1.903

**EK-B. PISA 2015 Yılı Uygulamasına Ait Veri Setinin 2 PL Modelde
Kestirilen Madde Parametreleri (Devam)**

Madde	A	b	Madde	a	b	Madde	a	B
M94	0.444	-1.183	M116	0.768	0.656	M138	1.059	0.658
M95	0.315	-1.444	M117	0.873	-0.861	M139	1.013	0.631
M96	0.841	0.927	M118	1.594	0.639	M140	1.433	0.925
M97	0.884	0.692	M119	1.363	1.322	M141	1.894	0.090
M98	1.027	-0.104	M120	0.591	-0.116	M142	1.786	0.117
M99	0.730	-1.754	M121	2.101	0.225	M143	1.721	1.001
M100	0.745	1.559	M122	2.049	0.307	M144	1.494	1.028
M101	1.014	0.406	M123	0.692	0.523	M145	1.789	0.137
M102	1.870	0.298	M124	2.218	-0.336	M146	1.434	1.308
M103	1.858	-0.283	M125	0.911	1.425	M147	0.664	-0.343
M104	0.874	-0.671	M126	1.471	0.850	M148	0.192	-1.632
M105	0.861	-0.688	M127	1.525	-0.518	M149	0.481	-1.060
M106	1.986	-0.058	M128	1.090	1.255	M150	0.924	0.178
M107	0.836	-0.814	M129	0.300	-1.506	M151	0.332	1.546
M108	1.641	-0.783	M130	0.317	-1.486	M152	0.324	1.264
M109	1.167	1.089	M131	1.525	0.802	M153	0.696	0.833
M110	0.819	-1.079	M132	1.630	1.050	M154	0.356	1.580
M111	1.724	0.092	M133	1.805	-0.406	M155	0.537	-0.110
M112	2.057	0.150	M134	1.484	-1.109	M156	0.969	0.547
M113	1.451	0.846	M135	1.678	-0.679	M157	0.966	1.227
M114	1.944	0.037	M136	1.860	0.072	M158	0.183	-1.654
M115	0.303	-1.852	M137	1.393	1.038	M159	0.962	0.132

EK-C. Yukarıdan-aşağıya Test Birleştirme Yöntemine Göre Kestirilen Madde Ayırt Edicilik Parametrelerinin Ortalama ve Standart Sapma Değerleri

	Modül uzunluğu	Örneklem Büyüklüğü	Modül	Panel-1		Panel-2	
				\bar{x}_{a1}	Ss	\bar{x}_{a2}	Ss
İki aşamalı panel deseni ("1-2")	6	250	Modül-1 (O)	1.525	0.169	1.738	0.334
			Modül-2 (K)	1.523	0.184	1.381	0.245
			Modül-2 (Z)	1.481	0.107	1.551	0.122
		2000	Modül-1 (O)	1.515	0.332	1.668	0.310
			Modül-2 (K)	1.328	0.210	1.505	0.244
			Modül-2 (Z)	1.621	0.067	1.493	0.139
	12	250	Modül-1 (O)	1.455	0.392	1.359	0.215
			Modül-2 (K)	1.466	0.311	1.452	0.324
			Modül-2 (Z)	1.480	0.325	1.431	0.279
		2000	Modül-1 (O)	1.465	0.305	1.467	0.317
			Modül-2 (K)	1.159	0.356	1.367	0.432
			Modül-2 (Z)	1.468	0.214	1.406	0.161
Üç aşamalı panel deseni ("1-2-2")	6	250	Modül-1 (O)	1.190	0.240	1.820	0.219
			Modül-2 (K)	1.840	0.249	1.770	0.352
			Modül-2 (Z)	1.460	0.198	1.830	0.249
		2000	Modül-3 (K)	1.180	0.279	1.260	0.376
			Modül-3 (Z)	1.660	0.098	1.450	0.111
			Modül-1 (O)	1.533	0.238	1.670	0.324
	12	250	Modül-2 (K)	1.485	0.209	1.406	0.169
			Modül-2 (Z)	1.090	0.601	1.061	0.572
			Modül-3 (K)	1.300	0.344	1.290	0.155
		2000	Modül-3 (Z)	1.478	0.193	1.415	0.119
			Modül-1 (O)	1.245	0.397	1.416	0.311
			Modül-2 (K)	1.196	0.476	1.269	0.432
12	250	Modül-2 (Z)	1.189	0.510	1.235	0.580	
		Modül-3 (K)	1.371	0.392	1.470	0.329	
		Modül-3 (Z)	1.318	0.389	1.412	0.269	
	2000	Modül-1 (O)	1.257	0.350	1.109	0.298	
		Modül-2 (K)	1.499	0.304	1.679	0.343	
		Modül-2 (Z)	0.896	0.237	0.903	0.204	
			Modül-3 (K)	1.078	0.437	1.436	0.321
			Modül-3 (Z)	1.445	0.228	0.414	0.182

**EK-C. Yukarıdan-Aşağıya Test Birleştirme Yöntemine Göre Kestirilen Madde
Ayırt Edicilik Parametrelerinin Ortalama ve Standart Sapma Değerleri
(Devam)**

Modül uzunluğu	Örneklem Büyüküğü	Modül	Panel-1		Panel-2		
			\bar{x}_{a1}	Ss	\bar{x}_{a2}	Ss	
Üç aşamalı panel deseni ("1-2-3")	6	Modül-1 (O)	1.225	0.340	1.493	0.276	
		Modül-2 (K)	1.568	0.412	1.418	0.176	
		250	Modül-2 (Z)	1.521	0.153	1.606	0.323
		Modül-3 (K)	0.538	0.246	1.203	0.182	
		Modül-3 (O)	1.361	0.458	1.410	0.148	
		Modül-3 (Z)	1.181	0.902	0.815	0.585	
	2000	Modül-1 (O)	1.533	0.385	1.490	0.228	
		Modül-2 (K)	1.535	0.194	1.441	0.159	
		Modül-2 (Z)	1.031	0.416	0.738	0.326	
		Modül-3 (K)	1.535	0.178	1.528	0.199	
		Modül-3 (O)	0.733	0.303	0.916	0.110	
		Modül-3 (Z)	0.753	0.234	0.620	0.339	
	250	Modül-1 (O)	1.283	0.357	1.169	0.403	
		Modül-2 (K)	1.093	0.331	1.338	0.464	
		Modül-2 (Z)	1.212	0.501	1.185	0.418	
		250	Modül-3 (K)	1.677	0.228	1.498	0.314
		Modül-3 (O)	1.299	0.442	1.010	0.593	
		Modül-3 (Z)	0.731	0.472	1.016	0.220	
	12	2000	Modül-1 (O)	1.420	0.424	1.306	0.336
			Modül-2 (K)	1.131	0.397	1.255	0.424
			Modül-2 (Z)	1.210	0.622	1.115	0.417
			Modül-3 (K)	1.481	0.327	1.463	0.424
			Modül-3 (O)	1.300	0.494	1.111	0.426
			Modül-3 (Z)	1.019	0.469	0.843	0.404

EK-Ç. Yukarıdan-aşağıya Test Birleştirme Yöntemine Göre Kestirilen Madde Güçlük Parametrelerinin Ortalama ve Standart Sapma Değerleri

			Panel-1		Panel-2		
Modül uzunluğu	Örneklem Büyüklüğü	Modül	\bar{x}_{b1}	Ss	\bar{x}_{b2}	Ss	
İki aşamalı panel deseni ("1-2")	6	Modül-1 (O)	-0.040	0.450	-0.258	0.275	
		250	Modül-2 (K)	-0.851	0.316	-0.737	0.179
		Modül-2 (Z)	0.787	0.165	0.963	0.097	
		2000	Modül-1 (O)	-0.322	0.308	-0.013	0.567
		Modül-2 (K)	-0.907	0.272	-0.828	0.344	
		Modül-2 (Z)	0.821	0.179	0.924	0.286	
	12	250	Modül-1 (O)	0.659	0.318	0.672	0.336
		Modül-2 (K)	-0.598	0.415	-0.640	0.336	
		Modül-2 (Z)	0.702	0.417	0.690	0.336	
		2000	Modül-1 (O)	-0.272	0.243	-0.238	0.145
		Modül-2 (K)	-0.904	0.309	-0.800	0.377	
		Modül-2 (Z)	0.849	0.269	0.890	0.181	
Üç aşamalı panel deseni ("1-2-2")	6	250	Modül-1 (O)	-0.827	0.448	-0.436	0.567
		Modül-2 (K)	-0.991	0.486	-0.834	0.468	
		Modül-2 (Z)	0.201	1.119	0.111	0.411	
		Modül-3 (K)	-1.360	0.404	-0.640	0.208	
		Modül-3 (Z)	0.620	0.186	0.840	0.169	
		2000	Modül-1 (O)	-0.043	0.249	-0.040	0.603
	12	250	Modül-2 (K)	-0.084	0.643	-0.098	0.651
		Modül-2 (Z)	0.113	0.891	0.724	0.620	
		Modül-3 (K)	-1.001	0.379	-0.841	0.363	
		Modül-3 (Z)	0.971	0.193	1.031	0.142	
		2000	Modül-1 (O)	-0.484	0.501	-0.537	0.293
		Modül-2 (K)	-0.685	0.616	0.788	0.211	
12	250	Modül-2 (Z)	-0.334	0.737	0.153	0.505	
	Modül-3 (K)	-0.804	0.355	-0.373	0.340		
	Modül-3 (Z)	0.616	0.800	0.941	0.313		
	2000	Modül-1 (O)	-0.336	0.430	-0.131	0.313	
	Modül-2 (K)	-0.540	0.455	-0.448	0.413		
	Modül-2 (Z)	0.670	0.451	0.652	0.441		
		Modül-3 (K)	-0.943	0.459	-0.655	0.292	
		Modül-3 (Z)	0.894	0.229	0.855	0.228	

EK-Ç. Yukarıdan-Aşağıya Test Birleştirme Yöntemine Göre Kestirilen Madde Güçlük Parametrelerinin Ortalama ve Standart Sapma Değerleri (Devam)

	Modül uzunluğu	Örneklem Büyüklüğü	Modül	Panel-1		Panel-2	
				\bar{x}_{b1}	Ss	\bar{x}_{b2}	Ss
Üç aşamalı panel deseni ("1-2-3")	6	250	Modül-1 (O)	0.684	0.544	0.424	0.151
			Modül-2 (K)	-0.415	0.265	-0.388	0.207
			Modül-2 (Z)	0.784	0.220	0.746	0.351
			Modül-3 (K)	-0.681	0.368	-0.911	0.374
			Modül-3 (O)	0.543	0.546	0.676	0.277
			Modül-3 (Z)	0.718	0.761	0.826	0.701
	2000	Modül-1 (O)	0.152	0.917	0.078	0.586	
		Modül-2 (K)	-0.207	0.878	-0.090	0.763	
		Modül-2 (Z)	0.416	0.743	0.548	1.195	
		Modül-3 (K)	-0.029	0.954	-0.170	0.799	
		Modül-3 (O)	0.296	1.082	0.647	0.842	
		Modül-3 (Z)	0.478	1.207	0.826	0.963	
	250	Modül-1 (O)	-0.441	0.342	-0.373	0.414	
		Modül-2 (K)	-0.843	0.632	-0.496	0.452	
		Modül-2 (Z)	0.363	1.057	0.689	0.299	
		Modül-3 (K)	-0.112	0.307	-0.287	0.427	
		Modül-3 (O)	0.257	0.443	0.171	0.989	
		Modül-3 (Z)	0.604	0.918	0.588	0.615	
	12	2000	Modül-1 (O)	-0.308	0.487	-0.365	0.506
			Modül-2 (K)	-0.732	0.508	-0.607	0.397
			Modül-2 (Z)	0.507	0.443	0.414	0.769
			Modül-3 (K)	-0.371	0.520	-0.168	0.292
			Modül-3 (O)	0.565	0.703	0.049	0.278
			Modül-3 (Z)	0.874	0.983	0.631	0.790

**EK-D. Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Kestirilen Madde
Ayırt Edicilik Parametrelerinin Ortalama ve Standart Sapma Değerleri**

			Panel-1		Panel-2		
Modül uzunluğu	Örneklem Büyüküğü	Modül	\bar{x}_{a1}	Ss	\bar{x}_{a2}	Ss	
İki aşamalı panel deseni ("1-2")	6	Modül-1 (O)	1.263	0.622	1.591	0.399	
		250	Modül-2 (K)	1.561	0.357	1.505	0.347
		Modül-2 (Z)	1.515	0.369	1.521	0.249	
		2000	Modül-1 (O)	1.515	0.314	1.366	0.291
		Modül-2 (K)	1.355	0.280	1.471	0.293	
		Modül-2 (Z)	1.432	0.335	1.356	0.277	
	12	250	Modül-1 (O)	1.385	0.458	1.355	0.320
		Modül-2 (K)	1.440	0.338	1.464	0.226	
		Modül-2 (Z)	1.459	0.356	1.479	0.297	
		2000	Modül-1 (O)	1.390	0.446	1.357	0.265
		Modül-2 (K)	1.433	0.320	1.461	0.365	
		Modül-2 (Z)	1.483	0.354	1.525	0.277	
Üç aşamalı panel deseni ("1-2-2")	6	250	Modül-1 (O)	1.376	0.225	1.403	0.305
		Modül-2 (K)	1.595	0.377	1.491	0.201	
		Modül-2 (Z)	1.513	0.429	1.581	0.460	
		Modül-3 (K)	1.216	0.390	1.368	0.246	
		Modül-3 (Z)	0.321	0.016	0.378	0.201	
		2000	Modül-1 (O)	1.383	0.184	1.531	0.353
	12	250	Modül-2 (K)	1.511	0.317	1.420	0.252
		Modül-2 (Z)	1.603	0.220	1.580	0.273	
		2000	Modül-3 (K)	1.546	0.418	1.453	0.442
		Modül-3 (Z)	0.728	0.288	0.786	0.385	
		Modül-1 (O)	1.374	0.296	1.375	0.488	
		Modül-2 (K)	1.215	0.351	1.400	0.321	
2000	250	Modül-2 (Z)	1.256	0.495	1.254	0.480	
	Modül-3 (K)	1.139	0.512	1.130	0.524		
	Modül-3 (Z)	0.805	0.543	0.687	0.340		
	Modül-1 (O)	1.456	0.370	1.135	0.365		
	Modül-2 (K)	1.397	0.352	1.324	0.385		
	Modül-2 (Z)	1.342	0.324	1.301	0.413		
2000	Modül-3 (K)	1.185	0.663	1.271	0.549		
	Modül-3 (Z)	0.849	0.265	0.623	0.395		

**EK-D. Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Kestirilen Madde
Ayırt Edicilik Parametrelerinin Ortalama ve Standart Sapma Değerleri
(Devam)**

Modül uzunluğu	Örneklem Büyüküğü	Modül	Panel-1		Panel-2		
			\bar{x}_{a1}	Ss	\bar{x}_{a2}	Ss	
Üç aşamalı panel deseni ("1-2-3")	6	Modül-1 (O)	1.406	0.293	1.588	0.366	
		Modül-2 (K)	1.241	0.627	1.071	0.731	
		250	Modül-2 (Z)	1.265	0.292	1.358	0.327
		Modül-3 (K)	1.093	0.717	1.130	0.706	
		Modül-3 (O)	1.473	0.181	1.398	0.284	
		Modül-3 (Z)	1.208	0.453	1.276	0.388	
	2000	Modül-1 (O)	1.356	0.277	1.638	0.348	
		Modül-2 (K)	1.323	0.750	1.316	0.638	
		Modül-2 (Z)	1.241	0.526	1.403	0.339	
		Modül-3 (K)	1.276	0.719	1.196	0.727	
		Modül-3 (O)	1.356	0.172	1.583	0.298	
		Modül-3 (Z)	1.181	0.537	1.296	0.258	
	250	Modül-1 (O)	1.125	0.363	1.060	0.348	
		Modül-2 (K)	1.086	0.355	1.153	0.396	
		Modül-2 (Z)	1.235	0.391	1.390	0.420	
		Modül-3 (K)	1.350	0.420	1.210	0.468	
		Modül-3 (O)	1.657	0.183	1.455	0.206	
		Modül-3 (Z)	0.930	0.687	0.499	0.486	
	12	2000	Modül-1 (O)	1.377	0.342	1.400	0.436
			Modül-2 (K)	1.106	0.597	1.071	0.546
			Modül-2 (Z)	1.187	0.401	1.119	0.646
		Modül-3 (K)	1.047	0.609	1.003	0.488	
		Modül-3 (O)	1.270	0.352	1.456	0.311	
		Modül-3 (Z)	1.116	0.342	1.047	0.375	

**EK-E. Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Kestirilen Madde
Güçlük Parametrelerinin Ortalama ve Standart Sapma Değerleri**

			Panel-1		Panel-2		
Modül uzunluğu	Örneklem Büyüküğü	Modül	\bar{x}_{b1}	Ss	\bar{x}_{b2}	Ss	
İki aşamalı panel deseni ("1-2")	6	Modül-1 (O)	-0.203	0.426	-0.331	0.353	
		250	Modül-2 (K)	-0.669	0.392	-0.618	0.394
		Modül-2 (Z)	0.563	0.328	0.586	0.416	
		Modül-1 (O)	0.666	0.354	0.729	0.253	
		2000	Modül-2 (K)	-0.650	0.209	-0.818	0.414
		Modül-2 (Z)	0.735	0.349	0.792	0.320	
	12	250	Modül-1 (O)	0.543	0.617	0.791	0.362
		Modül-2 (K)	-0.637	0.371	-0.619	0.393	
		Modül-2 (Z)	0.623	0.333	0.874	0.401	
		2000	Modül-1 (O)	0.527	0.365	0.537	0.284
		Modül-2 (K)	-0.588	0.314	-0.728	0.399	
		Modül-2 (Z)	0.890	0.371	0.634	0.449	
Üç aşamalı panel deseni ("1-2-2")	6	250	Modül-1 (O)	0.168	0.425	0.106	0.338
		Modül-2 (K)	-0.510	0.199	-0.576	0.320	
		Modül-2 (Z)	0.724	0.463	0.252	0.933	
		Modül-3 (K)	0.470	0.820	0.353	0.947	
		6	Modül-3 (Z)	0.631	1.475	0.561	1.317
		Modül-1 (O)	0.500	0.218	0.433	0.920	
	2000	Modül-2 (K)	-0.811	0.421	-0.513	0.110	
		Modül-2 (Z)	0.616	0.454	0.527	0.391	
		Modül-3 (K)	0.145	0.356	0.221	0.385	
		Modül-3 (Z)	0.662	0.848	0.654	0.957	
		Modül-1 (O)	0.250	0.445	0.162	0.410	
		Modül-2 (K)	-0.659	0.388	-0.532	0.429	
12	250	Modül-2 (Z)	0.632	0.903	0.684	0.851	
	Modül-3 (K)	0.156	0.822	0.129	0.808		
	Modül-3 (Z)	0.461	1.212	0.263	1.243		
	Modül-1 (O)	0.324	0.429	0.319	0.462		
	2000	Modül-2 (K)	-0.426	0.308	-0.644	0.390	
	Modül-2 (Z)	0.555	0.334	0.540	0.474		
		Modül-3 (K)	-0.460	0.790	-0.147	0.962	
		Modül-3 (Z)	0.623	0.468	0.254	1.314	

EK-E. Aşağıdan-yukarıya Test Birleştirme Yöntemine Göre Kestirilen Madde Güçlük Parametrelerinin Ortalama ve Standart Sapma Değerleri (Devam)

	Modül uzunluğu	Örneklem Büyüklüğü	Modül	Panel-1		Panel-2	
				\bar{x}_{b1}	Ss	\bar{x}_{b2}	Ss
Üç aşamalı panel deseni ("1-2-3")	6	250	Modül-1 (O)	0.407	0.257	0.609	0.294
			Modül-2 (K)	-0.078	1.183	0.208	0.709
			Modül-2 (Z)	0.605	0.827	1.000	0.290
			Modül-3 (K)	0.026	1.070	0.018	0.479
			Modül-3 (O)	0.515	0.262	0.830	0.449
			Modül-3 (Z)	0.797	0.543	1.042	0.334
	2000	Modül-1 (O)	0.572	0.446	0.513	0.232	
		Modül-2 (K)	-0.167	1.075	-0.186	0.787	
		Modül-2 (Z)	0.843	0.133	0.852	0.306	
		Modül-3 (K)	-0.812	0.892	0.148	0.767	
		Modül-3 (O)	0.615	0.244	0.362	0.248	
		Modül-3 (Z)	0.883	0.513	0.639	0.752	
	250	Modül-1 (O)	0.649	0.740	0.495	0.670	
		Modül-2 (K)	-0.880	0.373	-0.335	0.898	
		Modül-2 (Z)	0.363	0.686	0.362	0.730	
		Modül-3 (K)	-0.556	0.586	-0.570	0.679	
		Modül-3 (O)	0.017	0.594	0.008	0.663	
		Modül-3 (Z)	0.036	0.879	0.795	1.285	
	12	2000	Modül-1 (O)	0.518	0.421	0.601	0.394
			Modül-2 (K)	-0.565	0.520	-0.129	0.730
			Modül-2 (Z)	0.480	0.883	0.038	1.18
			Modül-3 (K)	-0.207	1.082	-0.577	0.534
			Modül-3 (O)	0.661	0.408	0.418	0.389
			Modül-3 (Z)	0.802	0.883	0.826	0.509

EK-F: Etik Komisyonu Onay Bildirimi

Form: 40

Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

01 / 03 / 2018

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Anabilim Dalı Başkanlığı'na

Tez Başlığı / Konusu: BİREYSELLEŞTİRİLMİŞ ÇOK AŞAMALI TESTLERİN TEST BİRLEŞTİRME YÖNTEMLERİNE GÖRE İNCELENMESİ

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.

Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.


Ebru DOĞRUÖZ
(Öğrencinin Adı Soyadı, İmzası)

Öğrenci Bilgileri

Adı Soyadı	Ebru Doğruöz
Öğrenci No	N12244245
Anabilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.

Danışman Görüşü ve Onayı



Prof. Dr. Hülya KELEŞOĞLU
(İmza)
(Danışmanın Unvanı, Adı ve Soyadı)

EK-G: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.



07/03/2018

Ebru Doğruöz

EK-Ğ: Doktora Tez Çalışması Orijinallik Raporu

07/03/2018

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı: Bireyselleştirilmiş Çok Aşamalı Testlerin Test Birleştirme Yöntemlerine Göre İncelenmesi

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak Turnitin adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
07/03/2018	84	140,624	19/03/2018	%3	926611582

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Ebru DOĞRUÖZ

Öğrenci No.: N12244245

Ana Bilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: Y.Lisans Doktora Bütünleşik Dr.



07/03/2018

DANIŞMAN ONAYI

UYGUNDUR.
Prof. Dr. Hülya KELEÇİOĞLU



EK-H: Dissertation Originality Report

07/03/2018

HACETTEPE UNIVERSITY
Graduate School Of Educational Sciences
To The Department Of Educational Sciences

Thesis Title : Investigation Of Adaptive Multistage Test Based On Test Assembly Methods

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
07/03/2018	84	140,624	19/03/2018	%3	926611582

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: Ebru DOĞRUÖZ
Student No.: N12244245
Department: Educational Sciences
Program: Division of Measurement and Evaluation in Education
Status: Masters Ph.D. Integrated Ph.D.


07/03/2018

ADVISOR APPROVAL

APPROVED
Prof. Dr. Hülya KELECIOĞLU



EK-I: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversite'ye verilen kullanım hakları dışındaki bütün fikrî mülkiyet haklarım bende kalacak, tezimin tamamının veya bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinleri yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversite'ye teslim etmeyi taahhüt ederim.

Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.

(Bu seçenikle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etmeniz ve kütüphane bu talebinizi yerine getirse bile, teziniz arama motorlarının ön belleklerinde kalmaya devam edebilecektir)

Tezimin/Raporumun 07/03/2021 tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir).

Tezimin/Raporumun tarihine kadar erişime açılmasını istemiyorum ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.

Serbest Seçenek/Yazarın Seçimi:

.....
.....
.....

07/03/2018



Ebru DOĞRUÖZ

