ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
ÖLÇME VE DEĞERLENDİRME PROGRAMI

BİLGİSAYAR TEMELLİ BİREYSELLEŞTİRİLMİŞ TEST YAKLAŞIMLARININ TÜRKİYE'DEKİ MERKEZİ DİL SINAVLARINDA UYGULANABİLİRLİĞİNİN ARAŞTIRILMASI

DOKTORA TEZİ

ERCAN ÇOBAN

ANKARA
TEMMUZ, 2020


ANKARA ÜNİVERSİTESİ
1946

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
ÖLÇME VE DEĞERLENDİRME PROGRAMI

BİLGİSAYAR TEMELLİ BİREYSELLEŞTİRİLMİŞ TEST YAKLAŞIMLARININ TÜRKİYE'DEKİ MERKEZİ DİL SINAVLARINDA UYGULANABİLİRLİĞİNİN ARAŞTIRILMASI

DOKTORA TEZİ

ERCAN ÇOBAN

DANIŞMAN: DOÇ. DR. CELAL DEHA DOĞAN

ANKARA
TEMMUZ, 2020

ETİK İLKELERE UYGUNLUK BİLDİRİMİ

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

Ercan ÇOBAN

ÖZET

BİLGİSAYAR TEMELLİ BİREYSELLEŞTİRİLMİŞ TEST YAKLAŞIMLARININ TÜRKİYE'DEKİ MERKEZİ DİL SINAVLARINDA UYGULANABİLİRLİĞİNİN ARAŞTIRILMASI

ÇOBAN, Ercan
Doktora Tezi, Ölçme ve Değerlendirme Anabilim Dalı
Tez Danışmanı: Doç. Dr. Celal Deha DOĞAN
Temmuz, 2020, xv+84 sayfa

Bu çalışmada bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT) ve bilgisayarlı sınıflama testi (BST) bireyselleştirilmiş test yaklaşımlarından hangisinin Türkiye'de kullanılan dil sınavlarına daha uygun olduğunu belirlemek amaçlanmıştır. Bireyselleştirmenin olmadığı bilgisayarlı lineer test (BLT) ise bireyselleştirilmiş testlerin geleneksel testlere göre ne derece avantajlı olduğunu belirlemek için referans olarak kullanılmıştır. Bu test yaklaşımlarını, madde havuzu büyüklüğü ve test uzunluğunun değiştirildiği koşullarda sınıflama doğruluğu, test güvenliği ve madde havuzu kullanımı açısından karşılaştırmak için bilgisayar ortamında simülasyonlar yapılmıştır. BBT, ÇABT ve BLT ölçme kesinliği açısından da karşılaştırılmıştır. Madde seçim yönteminin BST uygulamalarının sınıflama doğruluğu üzerindeki etkisini araştırabilmek için, bu test yaklaşımı iki farklı madde seçim yöntemiyle uygulanmıştır. Türkiye'de kullanılan dil sınavları, içerik ve madde formatı açısından birbirine benzerdir. Bu çalışmada, bu sınavlardan biri olan Yüksek Öğretim Kurumları Dil Sınavı (YÖKDİL) ele alınmış ve BBT, ÇABT, BST ve BLT test yaklaşımları YÖKDİL bağlamında karşılaştırılmıştır. Araştırmada kullanılan madde havuzları YÖKDİL'in içerik ve madde sayısına dayalı olarak oluşturulmuştur. Araştırma sonuçlarına göre, BBT ölçme kesinliği açısından ÇABT ve BLT'ye göre daha iyidir. Sınıflama doğruluğu en yüksek olan test yaklaşımı, madde seçimini kesme puanına dayalı olarak yapan BST'dir. ÇABT, sınıflama doğruluğu açısından diğer test yaklaşımlarına göre daha kötü sonuçlar vermiştir. Fakat madde kullanım sıklığı oranları,

test çakışma oranları ve madde havuzu kullanımı açısından en iyi sonuçlar ÇABT simülasyonlarında elde edilmiştir.

Anahtar Sözcükler: Bilgisayarlı test, bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT), bilgisayarlı sınıflama testi (BST), dil sınavı

ABSTRACT

INVESTIGATION OF APPLICABILITY OF COMPUTER-BASED ADAPTIVE TESTING APPROACHES TO CENTRAL LANGUAGE EXAMS IN TURKEY

ÇOBAN, Ercan

Dissertation, Department of Measurement and Evaluation

Supervisor: Assoc. Prof. Dr. Celal Deha DOĞAN

July 2020, xv+84 pages

This study aimed to determine which of computerized adaptive test (CAT), multi-stage adaptive test (MST), and computerized classification test (CCT) is the most appropriate adaptive testing approach for the language exams applied in Turkey. Computerized linear test (CLT), which does not have any adaptive feature, was used as a reference test approach to determine how advantageous adaptive tests are compared to conventional tests. Computer simulations were conducted to compare these testing approaches in terms of classification accuracy, test security and item pool usage under different test length and pool size conditions. CAT, MST and CLT were also compared in terms of measurement precision. CCT simulations were conducted with two different item selection methods to investigate the effect of item selection method on classification accuracy of this testing approach. The language exams used in Turkey are similar in terms of content and item format. In this study, Foreign Language Exam for Higher Education Institutions (YOKDIL), which is one of these exams, was addressed and CAT, MST, CCT, and CLT were compared in the context of this exam. Item pools used in the study were generated based on content and item number of YOKDIL. Results of the study revealed that CAT was superior to MST and CLT in terms of measurement precision. CCT selecting items based on cut score was the test approach having the highest classification accuracy rates. MST provided worse results in terms of classification accuracy than did other test approaches. However, MST yielded the best results in terms of item exposure rates, test overlap rates, and item pool usage.

Keywords: Computerized test, computerized adaptive test (CAT), multistage adaptive test (MST), computerized classification test (CCT), language exam

ÖNSÖZ

Yurtdışında uygulanan sınavlarda madde tepki kuramına dayalı bireyselleştirilmiş test yaklaşımları yaygın bir şekilde kullanılırken Türkiye'de sınavlar genellikle kağıt-kalem şeklinde uygulanmaktadır. Bireyselleştirilmiş testlerin avantajlarından faydalanabilmek için Türkiye'de de bireyselleştirilmiş test yaklaşımlarının kullanımına başlanmalıdır. Bireyselleştirilmiş testlerin kullanımına başlamadan önce kağıt-kalem şekkinde uygulanmakta olan sınavların kullanım amacına en uygun bireyselleştirilmiş test yaklaşımını belirlemeye yönelik araştırmalar yapılması gerekmektedir. Bu çalışmada Türkiye'de kullanılan merkezi dil sınavlarına en uygun olan bireyselleştirilmiş test yaklaşımının hangisi olduğu araştırılmış ve araştırma sonucunda önerilerde bulunulmuştur.

Bu çalışmanın yapılması sürecinde görüş ve önerileriyle bana yol gösteren ve her türlü desteği veren değerli danışmanım Doç. Dr. Celal Deha DOĞAN'a tüm emekleri için çok teşekkür ederim. Zor zamanlarımda desteğini esirgemeyen saygıdeğer hocam Prof. Dr. R. Nükhet ÇIKRIKCI'ya; görüş ve önerileriyle tezime katkıda bulunan saygıdeğer jüri üyeleri Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU, Prof. Dr. Hülya KELECİOĞLU ve Doç.Dr. Derya GÖKMEN'e çok teşekkür ederim. Ayrıca doktora eğitimim süresince bilgi ve deneyimlerinden faydalandığım Prof. Dr. Nizamettin KOÇ, Prof. Dr. Ezel TAVŞANCIL, Dr. Öğr. Üyesi Ömer KUTLU, Doç. Dr. Ergül DEMİR, Doç. Dr. Deniz GÜLLEROĞLU'na ve lisansüstü eğitim hayatım süresince üzerimde emeği olan ölçme ve değerlendirme alanındaki tüm akademisyenlere teşekkür ederim.

Çalışmam sırasındaki tüm desteklerinden dolayı arkadaşlarım Ömer KAMIŞ, Muharrem ŞENGÜL, Fatih İMROL ve Merve ŞAHİN KÜRŞAD'a teşekkür ederim.

Hayatım boyunca her koşulda beni destekleyen ve zor zamanlarımda her zaman yanımda olan anne, baba ve kardeşlerime çok teşekkür ederim.

Ercan ÇOBAN

İÇİNDEKİLER

Sayfa

TABLOLAR DİZİNİ

Tablo **Sayfa**

ŞEKİLLER DİZİNİ

Şekil **Sayfa**

KISALTMALAR

AOK: Ağırlıklandırılmış Olabilirlik Kestirimi

AOOK: Aşamalı Olasılık Oran Testi

BBT: Bireyselleştirilmiş Bilgisayarlı Test

BKT: Bilgi Koşullu Tesadüfi Yöntemi

BLT: Bilgisayarlı Lineer Test

BSD: Beklenen Sonsal Dağılım

BST: Bilgisayarlı Sınıflama Testi

BST_KP: Kesme Puanına Dayalı Madde Seçimi Yapan Bilgisayarlı Sınıflama Testi

BST_GY: Geçici Yeteneğe Dayalı Madde Seçimi Yapan Bilgisayarlı Sınıflama Testi

ÇABT: Çok Aşamalı Bireyselleştirilmiş Test

GKPM: Genelleştirilmiş Kısmi Puan Modeli

KTK: Klasik Test Kuramı

MEB: Milli Eğitim Bakanlığı

MFB: Maksimum Fisher Bilgisi

MO: Maksimum Olabilirlik

MSD: Maksimum Sonsal Dağılım

MTK: Madde Tepki Kuramı

OTO: Otomatik Test Oluşturma

ÖSYM: Öğrenci Seçme ve Yerleştirme Merkezi

RMSE: Root Mean Square Error

YDS: Yabancı Dil Sınavı

YÖKDİL: Yüksek Öğretim Kurumu Dil Sınavı

BÖLÜM 1

GİRİŞ

Bu bölümde araştırmanın problemi, amacı ve önemi açıklanmıştır. Ayrıca araştırmanının sınırlılıklarına yer verilmiştir.

Problem

Teknoloji ve psikometrideki gelişmeler sonucunda bilgisayarların eğitim ve psikolojideki ölçme işlemlerinde kullanımı yaygınlaşmıştır. Test uygulayıcılar eğitim ve psikolojide kullanılan testleri bilgisayar ortamında uygulamaya başlamıştır. Hızlı puanlama yapılması, yenilikçi madde türlerinin uygulanmasına olanak sağlaması ve test güvenliğini artırması gibi avantajlarının olması (Chalhoub-Deville ve Deville, 1999) bilgisayarlı testlere ilgiyi arttırmıştır. Madde tepki kuramı (MTK) modellerinin geliştirilmesiyle birlikte psikometrik açıdan birçok avantaj sağlayan bireyselleştirilmiş bilgisayarlı test (computerized adaptive test, BBT), çok aşamalı bireyselleştirilmiş test (multistage adaptive test, ÇABT) ve bilgisayarlı sınıflama testi (computerized classification test, BST) gibi bilgisayarda uygulanan bireyselleştirilmiş test yaklaşımları yaygınlaşmaya başlamıştır.

BBT uygulamalarında her bireye kendi yetenek düzeyine uygun maddeler uygulanmaktadır. Bireye uygulanacak madde bireyin daha önce uygulanan maddelere verdiği cevaplara dayalı hesaplanan yetenek kestirimine göre seçilmektedir. Bireye geçici yetenek kestirim değerinde en fazla bilgi veren madde uygulanmaktadır. BBT'nin bu özelliği sayesinde, geleneksel kağıt-kalem testlerine göre daha kısa testlerle hedeflenen ölçme kesinliğine ulaşılmaktadır (Weiss, 1982).

BBT beş temel bileşenden oluşmaktadır. Bunlar madde havuzu, test başlatma yöntemi, madde seçim yöntemi, yetenek kestirim yöntemi ve sonlandırma kuralıdır (Thompson ve Weiss, 2011). BBT uygulaması için tüm bireylerin yetenek düzeylerine uygun yeterince madde içeren madde havuzuna ihtiyaç vardır (Green, Bock, Humphreys, Linn ve Reckase, 1984). Madde havuzunda parametre kestirimi önceden yapılmış maddeler yer almaktadır. BBT uygulamasında bireylere yetenek düzeylerine

uygun maddeler uygulanmaktadır. Testin başlangıcında bireyin yetenek düzeyi bilinmediği için test başlatma yönteminin belirlenmesi gerekmektedir. Testi başlatmak için farklı yöntemler bulunmaktadır. Tüm bireylerin başlangıç yetenek değerini "0" olarak belirleyerek ilk maddeyi bu yetenek düzeyine göre seçmek en basit yöntemdir. Bireylerin yetenekleri hakkında önsel bilgi (önceki test sonuçları vb.) olduğu durumlarda ise ilk madde seçimini bu bilgiye dayalı yapmak diğer bir yöntemdir (Thompson ve Weiss, 2011).

BBT uygulamalarının en önemli bileşenlerinden birisi madde seçim yöntemidir. BBT uygulamaların bireyselleştirilmiş olması madde seçim yöntemleriyle ilişkilidir. Kullanılan madde seçim yöntemleri sayesinde bireylere kendi yetenek düzeylerine uygun maddeler uygulanabilmesi, bu uygulamalara bireyselleştirilmiş olma özelliği kazandırmaktadır. BBT uygulamalarında kullanılabilecek birçok madde seçim yöntemi vardır. Bu yöntemler içerisinde maksimum Fisher bilgisi (maximum Fisher information, MFB) yöntemi ve Owen'ın Bayes yöntemi en yaygın kullanılan klasik madde seçim yöntemleridir. Maksimum Kullback-Leibler bilgisi ve olabilirlikle ağırlıklandırılmış bilgi yöntemleri ise modern madde seçim yöntemler arasındadır (van der Linden ve Pashley, 2010).

BBT uygulamalarının diğer bir bileşeni yetenek kestirim yöntemidir. Maksimum olabilirlik (maximum likelihood, MO), ağırlıklandırılmış olabilirlik kestirimi (weighted likelihood estimation, AOK), beklenen sonsal dağılım (expected a posteriori, BSD) ve maksimum sonsal dağılım (maximum a posteriori, MSD) BBT uygulamalarında kullanılan temel yetenek kestirim yöntemleridir (van der Linden ve Pashley, 2010).

Testin ne zaman sonlandırılacağı BBT uygulamalarındaki önemli noktalardan birisidir. Testi sonlandırmak için sabit ve değişen uzunluk sonlandırma kuralları bulunmaktadır. Sabit uzunluk sonlandırma kuralında tüm bireylere belirli sayıda madde uygulandığında test sonlandırılmaktadır. Değişen uzunluk sonlandırma kuralı farklı şekillerde uygulanabilmektedir. Standart hata sonlandırma kuralı, θ değişimi sonlandırma kuralı ve minimum bilgi sonlandırma kuralı gibi değişen uzunluk sonlandırma kuralları bulunmaktadır (Thompson ve Weiss, 2011).

BBT uygulamalarında bireye kendi yetenek düzeyinde en çok bilgi veren maddeler uygulanarak, bireyin yeteneğini en doğru şekilde kestirmek amaçlanmaktadır. Bu nedenle madde seçimi sırasında bireyin geçici yetenek düzeyinde en fazla bilgi veren madde seçilmektedir. Bu durum ayırt edicilik parametresi yüksek olan kaliteli maddelerin sık kullanımına yol açmaktadır (Chang ve Ying, 1999). Kaliteli maddelerin

sık kullanılması bu maddelerin açığa çıkmasına neden olarak test güvenliğine zarar vermektedir. Test güvenliğine zarar veren bu duruma karşı madde kullanım sıklığı kontrol yöntemleri (Georgiadou, Triantafillou ve Economides, 2007) geliştirilmiştir.

BBT uygulamalarında her bireye kendi yetenek düzeyine uygun maddeler uygulandığı için bireylere uygulanan maddeler farklıdır. Dolayısıyla her bireye farklı bir test formu uygulanmaktadır. Bu test formları uygulama sırasında oluşmaktadır. Uygulama sırasında oluşan bu test formlarının test belirtke tablosundaki tüm konuları dengeli bir şekilde temsil etmesi zordur. Bu durum test formlarının kapsam geçerliğini düşürmektedir. Test formlarının kapsam geçerliğini sağlanması için kapsam dengesinin kontrol edilmesi gerekmektedir (Segall, 2005). Kapsam dengesinin sağlanması için kısıtlanmış BBT (Kingsbury ve Zara, 1989), ağırlıklandırılmış sapmalar modeli (Stocking ve Swanson, 1993), gölge test yaklaşımı (van der Linden ve Reese, 1998), ağırlıklandırılmış ceza modeli (Shin, Chien, Way ve Swanson, 2009) ve maksimum öncelik indeksi (Cheng ve Chang, 2009) gibi yöntemler geliştirilmiştir.

Geleneksel kağıt-kalem testleriyle karşılaştırıldığında, BBT uygulamalarında daha kısa testlerle daha doğru ölçme işlemi yapılabilmektedir (Luecht ve Sireci, 2011; Segall, 2005). Böyle önemli bir avantajı olmasına rağmen BBT uygulamalarının bazı sınırlılıkları da vardır. Yüksek bilgi veren maddelerin aşırı kullanılması, test güvenliğine ilişkin bir tehdit oluşturabilmektedir. BBT uygulamalarında testler uygulama anında oluştuğu için kapsam dengesini sağlamak zordur. BBT uygulamalarında bir maddeyi yanıtlamadan diğerine geçmeye izin verilmediği için bireylerin zorlandıkları maddeleri atlama (skipping) olanakları yoktur. Ayrıca madde seçim yönteminin doğası gereği, geleneksel BBT uygulamalarında bireylerin daha önce uygulanan maddeye verdikleri yanıtı değiştirmesine izin verilmez. Bireylerin yanıtladıkları maddeleri tekrar gözden geçirme (item review) olanağı yoktur (Han, 2013; Luecht ve Sireci, 2011; Stone ve Davey, 2011). BBT uygulamalarının bu sınırlılıklarını çok aşamalı bireyselleştirilmiş test (ÇABT) uygulamaları kısmen ortadan kaldırmaktadır.

ÇABT, BBT ve kağıt-kalem testleri arasında uyumu sağlayan ve her iki test yaklaşımının avantajlarını birleştiren test yaklaşımıdır (Jodoin, Zenisky ve Hambleton, 2006; Zheng, Nozawa, Gao ve Chang, 2012). ÇABT'ler bireyselleştirilmiş testlerdir. Fakat bireyselleştirme madde düzeyinde değil modül olarak adlandırılan madde kümeleri düzeyindedir. Ayrıca kağıt-kalem testlerinde olduğu gibi ÇABT

uygulamalarında da bireyler yanıtlarını gözden geçirip değiştirebilmektedirler (Zheng vd., 2012).

Eğitim ve psikolojide testler seçme, yerleştirme, sertifikasyon, gelişimi izleme ve tanı gibi pratik amaçlarla kullanılmaktadır. Ölçme açısından bakıldığında ise testlerin iki temel amacı vardır. Bunlar bireylerin yeteneklerini kestirme ve bireyleri iki ya da daha fazla kategoriye sınıflamaktır (Eggen, 2010). BBT ve ÇABT hem bireylerin yeteneklerini kestirmek hem de bireyleri sınıflamak amacıyla kullanılabilmektedir. Diğer bir bilgisayarlı test yaklaşımı olan bilgisayarlı sınıflama testi (BST) ise sadece sınıflama amacıyla kullanılmaktadır. BST, BBT uygulamalarının özel bir biçimidir. BST uygulamalarının amacı, bireyleri iki ya da daha fazla kategoriye doğru bir şekilde sınıflamaktır (Eggen, 2011).

Bilgisayar teknolojisi ve psikometrideki gelişmelerle birlikte geleneksel kağıt-kalem testlerinin yerini yukarıda kısaca özetlenen bireyselleştirilmiş test yaklaşımları almıştır. Bu test yaklaşımları yurtdışında uzun süredir kullanılmaktadır. GMAT (Graduate Management Admission Test), ASVAB (Armed Services Vocational Aptitude Battery) ve TOEFL (Test of English as a Foreign Language) gibi sınavlar BBT olarak uygulanmaktır (He, 2010). GRE (Graduate Record Examination), 1993 tarihinden 2011'e kadar BBT olarak uygulanmıştır. Bu tarihten itibaren ÇABT olarak uygulanmaktadır (Briel ve Michel, 2014). Uniform CPA Exam ÇABT olarak uygulanmaktadır (Melican, Breithaupt ve Zhang, 2010). Hollanda'da kullanılan MATHCAT sınavı hem BST hem de BBT olarak uygulanmaktadır (Verschoor ve Straetmans, 2010). Japonya'da ise İngilizce yeterlik testi CASEC (Computerized Adaptive System for English Communication) BBT olarak uygulanmaktadır (Nogami ve Hayashi, 2010). Yurtdışındaki farklı ülkelerde kullanılan bu test yaklaşımları Türkiye'de henüz kullanılmamaktadır.

Türkiye'de Öğrenci Seçme ve Yerleştirme Merkezi (ÖSYM), Milli Eğitim Bakanlığı (MEB) ve Ankara Üniversitesi Sınav Yönetim Merkezi (ASYM) gibi kurumlar her yıl birçok sınav uygulamaktadırlar. Dünyada kağıt-kalem testlerinin yerini BBT, ÇABT ve BST gibi bireyselleştirilmiş test yaklaşımları alırken Türkiye'de ise sınavlar çoğunlukla kağıt-kalem şeklinde uygulanmaktadır. Bireyselleştirilmiş test yaklaşımlarının getirmiş olduğu avantajlardan faydalanabilmek için Türkiye'de de bu test yaklaşımlarının kullanımına başlanmalıdır. Öncelikle kağıt-kalem olarak uygulanan sınavlara en uygun bireyselleştirilmiş test yaklaşımını belirlemeye yönelik araştırmalar yapılması gerekmektedir. Özellikle yabancı dil sınavları, madde formatı açısından

bilgisayar ortamında uygulanmak için göreceli olarak daha uygundur. Türkiye'de kullanılan dil sınavlarında genellikle metinlere dayalı maddeler kullanıldığı için bu maddeleri bilgisayar ortamına aktarmak görece daha kolaydır. Ayrıca dil alanında bireyleri dil becerilerine göre sınıflamak için ileri, orta ve başlangıç gibi yeterlilik düzeyleri kullanılmaktadır. Dil sınavları, bu yönüyle sınıflama amacıyla kullanılan BST'yle benzer özelliklere sahiptir. Tüm bu nedenlerden dolayı, bireyselleştirilmiş testlere geçiş sürecine ilk olarak dil sınavlarından başlanabilir.

Türkiye'de bireylerin yabancı dil becerilerini ölçmek amacıyla kullanılan birkaç tane merkezi sınav bulunmaktadır. Bu sınavlardan Yabancı Dil Sınavı (YDS), hem kağıt kalem hem de bilgisayarlı lineer test olarak uygulanmaktadır. ASYM tarafından hazırlanan ve Anadolu Üniversitesi tarafından uygulanan Yüksek Öğretim Kurumları Dil Sınavı (YÖKDİL), kağıt-kalem şeklinde uygulanmaktadır. Bu sınavların maddeleri sınava katılan adaylarla paylaşılmaktadır. Bu nedenle her uygulama için tekrar madde yazılmaktadır. Bu sınavlar, bilgisayar ortamında bireyselleştirilmiş test olarak uygulanırsa maddeler adaylarla paylaşılmayacağı için maddeler tekrar kullanılabilecektir. Maddeler tekrar kullanılabileceği için madde yazım maliyeti göreceli olarak azalacaktır. Ayrıca daha az sayıda maddeyle katılımcıların dil yeterliği hakkında doğru kararlar verilebilecektir. Bireyselleştirilmiş testlerin bu avantajlarından faydalanabilmek için Türkiye'de kullanılan bu dil sınavlarının BBT, ÇABT ya da BST olarak uygulanabilirliğinin araştırılması gerekmektedir. Bu nedenle mevcut çalışmada bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT) ve bilgisayarlı sınıflama testi (BST) bireyselleştirilmiş test yaklaşımlarından hangisinin Türkiye'de kullanılan merkezi yabancı dil sınavlarına daha uygun olduğu araştırılmıştır.

Amaç

Bu araştırmanın amacı, üç farklı bireyselleştirilmiş test yaklaşımının Türkiye'de kullanılan merkezi yabancı dil sınavlarına uygunluk düzeyini belirlemektir. Araştırma kapsamında incelenen test yaklaşımları bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT) ve bilgisayarlı sınıflama testidir (BST). Bu üç bireyselleştirilmiş test yaklaşımı, hem birbirleriyle hem de bireyselleştirmenin olmadığı bilgisayarlı lineer test (BLT) ile karşılaştırılmıştır. BLT, bireyselleştirilmiş testlerin geleneksel testlere göre ne derece avantajlı olduğunu belirlemek için referans noktası olarak kullanılmıştır. Ayrıca madde seçim yönteminin BST uygulamalarının sınıflama

doğruluğu üzerindeki etkisini araştırabilmek için, bu test yaklaşımı iki farklı madde seçim yöntemiyle uygulanmıştır. Bu dört test yaklaşımı, madde havuzu büyüklüğü ve test uzunluğunun değiştirildiği koşullarda sınıflama doğruluğu (classification accuracy), test güvenliği (test security) ve madde havuzu kullanımı (item pool usage) açısından karşılaştırılmıştır. BBT, ÇABT ve BLT yaklaşımlarında yapılan yetenek kestirim değerlerinin doğruluk düzeyini değerlendirebilmek amacıyla bu üç test yaklaşımı ölçme kesinliği (measurement precision) açısından da karşılaştırılmıştır. Araştırmada aşağıdaki sorulara yanıt aranmıştır:

1. BBT, ÇABT ve BLT yaklaşımlarının her biri için ölçme kesinliği;
   a) Madde havuzu büyüklüğüne,
   b) Test uzunluğuna göre nasıl değişmektedir?
2. BBT, ÇABT, BST ve BLT yaklaşımlarının her biri için sınıflama doğruluğu;
   a) Madde havuzu büyüklüğüne,
   b) Test uzunluğuna göre nasıl değişmektedir?
3. BBT, ÇABT, BST ve BLT yaklaşımları kullanıldığında test güvenliğine ilişkin göstergeler;
   a) Madde havuzu büyüklüğüne,
   b) Test uzunluğuna göre nasıl değişmektedir?
4. BBT, ÇABT, BST ve BLT yaklaşımlarının her biri için madde havuzu kullanımı;
   a) Madde havuzu büyüklüğüne,
   b) Test uzunluğuna göre nasıl değişmektedir?

Önem

Türkiye'de her yıl onlarca merkezi sınav yapılmaktadır. Yapılan bu sınavlar çoğunlukla Klasik Test Kuramına (KTK) dayalı olarak hazırlanmakta ve kağıt-kalem şeklinde uygulanmaktadır. Dünyada ise GRE, TOEFL ve GMAT gibi sınavlarda MTK'ya dayalı bireyselleştirilmiş test yaklaşımları yaygın bir şekilde kullanılmaktadır. Bilgisayar teknolojisi ve MTK'nın avantajlarından faydalanabilmek için Türkiye'de de BBT, ÇABT ve BST gibi bireyselleştirilmiş test yaklaşımlarının kullanımına başlanmalıdır. ÖSYM e-sınav uygulamasını başlatarak bilgisayarlı lineer test uygulamasına geçmiştir. Dünyadaki gelişmelere uyum sağlanarak yakın gelecekte bireyselleştirilmiş testlere de geçiş yapılması kaçınılmazdır. Bu geçiş sürecinde

uygulanmakta olan ALES, YDS, KPSS, YGS ve YÖKDİL gibi sınavların kullanım amacına (yetenek kestirimi veya sınıflama) en uygun bireyselleştirilmiş test yaklaşımını belirlemeye yönelik araştırmalar yapılması gerekmektedir. Türkiye'de uygulanan merkezi dil sınavlarına en uygun bireyselleştirilmiş test yaklaşımının araştırıldığı bu araştırma, ÖSYM, ASYM ve MEB gibi test geliştiricilere ve uygulayıcılara yol gösterici olacaktır.

Türkiye'de YDS ve YÖKDİL gibi dil sınavları yıl içerisinde birden fazla kez uygulanmaktadır. Uygulama sonrası sınavlarda yer alan maddeler adaylarla paylaşıldığı için her uygulama için yeni maddeler yazılmaktadır. Madde yazım maliyeti dikkate alındığında her uygulama için yeniden madde yazılması ekonomik olarak büyük bir israfa yol açmaktadır. Bireyselleştirilmiş test uygulamalarında ise maddeler adaylarla paylaşılmadığı için havuzundaki maddeler tekrar kullanılabilmektedir. Dolayısıyla yazılan maddeler daha verimli bir şekilde kullanılmaktadır. YÖKDİL ve YDS gibi dil sınavları bireyselleştirilmiş test olarak uygulandığında maddeler tekrar kullanılabileceği için madde yazım maliyeti göreceli olarak daha az olacaktır. Türkiye'deki merkezi dil sınavlarına en uygun bireyselleştirilmiş testi belirlemeye yönelik yapılan bu çalışmanın sonuçları, test geliştiricilere ekonomik olarak katkıda bulunabilecek öneriler ortaya koyacaktır.

Türkiye'de bireyselleştirmiş testlere yönelik yapılan araştırmalarda genellikle BBT üzerine odaklanılmıştır (Aybek ve Çıkrıkçı, 2018; Boztunç-Öztürk ve Doğan, 2015; Bulut ve Kan, 2012; Eroğlu ve Kelecioğlu, 2015; Kalender, 2012; Kezer ve Koç, 2014). BST ve ÇABT hakkında ise sınırlı sayıda çalışma yapılmıştır. Fakat yapılan araştırmalarda genellikle Türkiye'deki dil sınavları üzerinde durulmamıştır. Mevcut araştırma, bu eksikliği gidererek yabancı dil üzerinde çalışan araştırmacılara önemli katkılarda bulunacaktır.

Araştırmada BBT, ÇABT, BST ve BLT yaklaşımları farklı madde havuzu büyüklüğü ve test uzunluğu koşulları altında karşılaştırılmıştır. Bu dört test yaklaşımının sınıflama doğruluğunun madde havuzu büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiği araştırılmıştır. Ayrıca BBT ve ÇABT yaklaşımlarının ölçme kesinliğinin test uzunluğu ve madde havuzu büyüklüğünden nasıl etkilendiği incelenmiştir. Bu nedenle araştırma sonuçları alan yazına katkıda bulunacaktır.

Sınırlılıklar

1. Ortak köklü maddelerin (madde takımı) çoklu puanlanan madde olarak ele alınması ve MTK modeli olarak genelleştirilmiş kısmi puan modelinin (GKPM) kullanılması araştırmanın sınırlılığıdır.

2. Araştırmada kesme puanı yeteneğin normal dağıldığı varsayımına dayalı olarak belirlenmiştir. Standart belirleme çalışmasının yapılmamış olması, araştırmanın sınırlılığıdır.

Tanımlar

Merkezi dil sınavı: Yetkili kurumlar tarafından bireylerin yabancı dil becerisini değerlendirmek için ülke genelinde uygulanan sınavdır.

BÖLÜM 2

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde araştırmada ele alınan test yaklaşımları hakkında kuramsal bilgilere ve bu araştırmayla ilgili araştırmalara yer verilmiştir. Araştırmada yer alan test yaklaşımlarının hepsi madde tep dayalı olduğu için ilk olarak bu kuram hakkında bilgi verilmiştir. Daha sonra sırasıyla bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT) ve bilgisayarlı sınıflama testi (BST) kısaca tanıtılmıştır. Kuramsal bilgilerden sonra ilgili araştırmalar verilmiştir.

Madde Tepki Kuramı

Madde tepki kuramı (MTK), bireyin test maddelerine verdiği yanıtlar ve maddelerin psikometrik özelliklerine dayalı olarak bireyin gizil özelliğini (yetenek, tutum, vb.) kestiren bir test kuramıdır. Bu test kuramı, bireyin bir maddeye verdiği yanıt ile bireyin yeteneği arasındaki ilişkiyi açıklayan matematiksel modellerden oluşmaktadır. Bireyin yanıtı ve yetenek arasındaki ilişki, madde tepki fonksiyonu (item response function) ya da madde karakteristik eğrisi (item charactersitic curve) olarak adlandırılan matematiksel fonksiyonlar aracılığıyla açıklanmaktadır. Madde karakteristik eğrisi, bireyin bir maddeye belirli bir yanıtı verme olasılığını bireyin yeteneği ve madde parametrelerinin fonksiyonu olarak tanımlamaktadır (Hambleton ve Swaminathan, 1985).

Eğitim ve psikolojik özelliklerin ölçülmesinde klasik test kuramı (KTK) sıklıkla kullanılmaktadır. Bu test kuramında birey özellikleri ve test özellikleri birbirine bağlıdır. Bireyin yetenek kestirimi, testin güçlük düzeyine göre değişebilmektedir. Benzer şekilde, test ve madde özellikleri uygulama yapılan grubun yetenek düzeyine göre değişebilmektedir. MTK, parametre değişmezliği özelliği sayesinde KTK'nın bu dezavantajını ortadan kaldırmaktadır. MTK'da madde parametreleri testi alan bireylerin yetenek dağılımına göre değişmemektedir. Kestirilen yetenek parametreleri de uygulanan testteki maddelerin özelliklerine göre değişmemektedir (Hambleton, Swaminathan ve Rogers, 1991).

Madde Tepki Kuramı Varsayımları

Madde tepki kuramının üç temel varsayımı vardır (DeMars, 2010). MTK 'nın ilk varsayımı tek boyutluluktur. Tek boyutluluk varsayımına göre testteki maddelerin tek bir gizil yapıyı ölçmesi gerekmektedir. Test sonuçlarını yetenek dışında çeşitli bilişsel ve psikolojik özellikler de etkileyebileceği için pratikte bu varsayımı sağlamak zordur. Bu nedenle test sonuçlarını belirleyen başat bir yapının varlığı kanıtlandığında tek boyutluk varsayımının sağlandığı kabul edilir. Bu başat yapı, test ile ölçülen gizil yapıdır (Hambleton ve Swaminathan, 1985). Ölçülen özellik dışında başka özellikler de test sonuçları üzerinde önemli düzeyde etkiye sahip olduğunda tek boyutluluk varsayımı ihlal edilmektedir. Örneğin, matematik becerisini ölçmek için geliştirilen bir testin maddeleri okuduğunu anlama becerisi gibi farklı özellikleri de ölçerse bu varsayım ihlal edilmiş olur.

MTK'nın ikinci varsayımı yerel bağımsızlıktır. Bu varsayıma göre yetenek sabit tutulduğunda testi alan bireylerin herhangi iki test maddesine verdiği yanıtlar istatistiksel olarak birbirinden bağımsız olmalıdır. Diğer bir ifadeyle, sabit bir yetenek düzeyindeki bireylerin farklı maddelere verdikleri yanıtlar arasında ilişki olmamalıdır. Bu varsayım tek boyutluluk varsayımıyla ilişkilidir. Bireylerin test maddelerine verdikleri yanıtlar üzerinde sadece ölçülen özellik etkili olduğunda bu varsayım sağlanmaktadır. Bireylerin maddelere verdikleri yanıtları etkileyen başka faktörler olduğunda ise bu varsayım ihlal edilmektedir. Ayrıca bir madde, diğer maddelerin doğru yanıtlanmasına katkıda bulunacak bilgi ya da ipucu verdiğinde yerel bağımsızlık ihlal edilmektedir (Hambleton vd., 1991). Ortak bir köke dayalı olarak yazılan maddeler kullanıldığında da yerel bağımsızlık varsayımı ihlal edilmektedir (Sireci, Thissen ve Wainer, 1991). Yerel bağımsızlık varsayımı sağlandığında, bir birey için herhangi bir yanıt örüntüsünün gerçekleşme olasılığı örüntüdeki madde yanıtlarının gerçekleşme olasılıklarının çarpımına eşittir (Hambleton ve Swaminathan, 1985).

MTK'nın üçüncü varsayımı seçilen model ve veri arasındaki uyumla ilgilidir. Bu varsayıma göre seçilen MTK modeli, bireyin yeteneği ve maddelere verdiği yanıtlar arasındaki gerçek ilişkiyi yansıtmalıdır (Hambleton vd., 1991). Diğer bir ifadeyle, model-veri uyumunun iyi olması için en doğru model seçilmelidir (DeMars, 2010). Örneğin, tahminle yanıtlama olasılığının yüksek olduğu bir test için kullanılan model, bu durumu dikkate alan bir model olmalıdır.

Madde Tepki Kuramı Modelleri

Maddelerin puanlanma şekli (ikili /çoklu) ve testin ölçtüğü özelik sayısı (tek boyutlu/çok boyutlu) gibi faktörlere göre farklı MTK modelleri bulunmaktadır. Bu çalışmada tek boyutlu MTK modellerinden biri kullanıldığı için çok boyutlu MTK modelleri üzerinde durulmamıştır.

MTK modelleri maddelerin puanlanma şekline göre iki gruba ayrılmaktadır. Çoktan seçmeli ve doğru-yanlış maddeleri gibi ikili puanlan maddeler için geliştirilen modeller iki kategorili (dichotomous) MTK modelleri olarak adlandırılmaktadır. Kısmi puanlamanın yapıldığı maddeler gibi ikiden fazla puanlama olasılığının olduğu maddeler için geliştirilen modeller ise çok kategorili (polytomous) MTK modelleri olarak adlandırılmaktadır.

İkili puanlanan maddeler için geliştirilen MTK modelleri, modelde yer alan parametre sayısına göre adlandırılmaktadır. Bir parametreli lojistik model (1 PLM), iki parametreli lojistik model (2 PLM) ve üç parametreli lojistik model, bu modeller arasında en sık kullanılan üç modeldir. Bir parametreli lojistik modelde sadece güçlük parametresi (b) yer almaktadır. Bu modelde madde ayırt edicilik parametresi (a) tüm maddeler için sabittir. Şans parametresi (c) ise tüm maddeler için sıfıra sabitlenmektedir. Bu modelde, yetenek düzeyi θ_j olan bir *j* bireyinin güçlük parametresi b_i olan bir i maddesine doğru yanıt verme olasılığı,

$$P_i(\theta_j) = \frac{e^{(\theta_j - b_i)}}{1+e^{(\theta_j - b_i)}} \tag{2.1}$$

eşitliğine göre hesaplanmaktadır. İki parametreli lojistik modelde güçlük parametresine ek olarak ayırt edicilik parametresi de yer almaktadır. Bu modelde, yetenek düzeyi θ_j olan bir *j* bireyinin *i* maddesine doğru yanıt verme olasılığı,

$$P_i(\theta_j) = \frac{e^{Da_i(\theta_j - b_i)}}{1+e^{Da_i(\theta_j - b_i)}} \tag{2.2}$$

eşitliğine göre hesaplanmaktadır. Bu eşitlikte b_i *i* maddesinin güçlük parametresini, a_i *i* maddesinin ayırt edicilik parametresini ve *D* ise ölçekleme katsayısını (1.7) ifade etmektedir. Üç parametreli lojistik modelde ise güçlük ve ayırt edicilik parametresine ek olarak şans parametresi yer almaktadır. Bu modelde, yetenek düzeyi θ_j olan bir *j* bireyinin *i* maddesine doğru yanıt verme olasılığı,

$$P_i(\theta_j) = c_i + (1-c_i)\frac{e^{Da_i(\theta_j - b_i)}}{1+e^{Da_i(\theta_j - b_i)}} \quad (2.3)$$

eşitliğiyle hesaplanmaktadır. Bu eşitlikte b_i *i* maddesinin güçlük parametresini, a_i *i* maddesinin ayırt edicilik parametresini, c_i *i* maddesinin şans parametresini ve *D* ise ölçekleme katsayısını (1.7) ifade etmektedir (Hambleton vd., 1991).

Çoklu Puanlanan Maddeler İçin Madde Tepki Kuramı Modelleri

Eğitim ve psikolojide kullanılan bazı maddelerde doğru-yanlış gibi ikili puanlama yapılmaz. Kısmi puanlamanın yapıldığı açık uçlu bir matematik maddesinde bireylerin alabileceği olası puan sayısı ikiden fazladır. Benzer şekilde, bireylerin tutumunu ölçen Likert tipi maddelerde ikiden fazla puan olasılığı bulunmaktadır. Çoklu puanlamanın yapıldığı bu tür maddeler için geliştirilmiş MTK modelleri de bulunmaktadır. Aşamalı tepki modeli, modifiye edilmiş aşamalı tepki modeli, kısmi puan modeli, genelleştirilmiş kısmi puan modeli, derecelendirme ölçeği modeli ve sınıflamalı tepki modeli bu modeller arasındadır. Bu çalışmada genelleştirilmiş kısmi puan modeli kullanıldığı için bu model hakkında ayrıntılı bilgi verilmiştir.

Genelleştirilmiş kısmi puan modelinde (GKPM) tüm maddelerin ayırıcılık gücü farklıdır. Bu modelde her bir madde için ayrı bir eğim parametresi (α_i) vardır. Modelde her bir yanıt kategorisi için kategori yanıt eğrileri (category response curve) yer almaktadır. Kategori yanıt eğrileri, her bir kategorinin yanıtlanma olasılığının yetenek düzeyine göre değişimini göstermektedir. Bu modelde iki ardışık kategori arasındaki geçiş için gerekli olan yetenek düzeyiyle ilişkili olan kategori kesişim parametresi (δ_{ij}) bulunmaktadır. Adım güçlüğü (step difficulty) olarak da adlandırılan bu parametreler kategori yanıt eğrilerin kesiştiği noktadaki yetenek düzeyine eşittir. Bu modelde, yetenek düzeyi θ olan bir bireyin $x=0,1,\ldots, m_i$ şeklinde puanlanan m_i+1 yanıt kategorisine sahip bir *i* maddesini $x=j$ yanıt kategorisinde yanıtlama olasılığı,

$$P_{ix}(\theta) = \frac{e^{\sum_{j=0}^{x}\alpha_i(\theta-\delta_{ij})}}{\sum_{r=0}^{m_i}\left[e^{\sum_{j=0}^{r}\alpha_i(\theta-\delta_{ij})}\right]} \quad (2.4)$$

eşitliğiyle hesaplanmaktadır. Bu eşitlikte $\sum_{j=0}^{0}\alpha_i(\theta-\delta_{ij})$= 0'dır (Embretson ve Reise, 2000).

Madde ve Test Bilgi Fonksiyonu

MTK'da ölçmenin standart hatası ve güvenirlik bilgi fonksiyonlarına dayalı olarak hesaplanmaktadır (Demars, 2010). Bilgi fonksiyonları madde ve testlerin tanımlanması, madde seçimi ve testlerin karşılaştırılması gibi süreçlerde kullanılmaktadır. Hem madde hem de testin tamamı için bilgi fonksiyonu tanımlanmaktadır. Bir maddenin θ yetenek düzeyinde sağladığı bilgi, madde bilgi fonksiyonu aracılığıyla belirlenmektedir (Hambleton vd., 1991). İkili puanlanan bir i maddesi için madde bilgisi,

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)(1-P_i(\theta))} \tag{2.5}$$

eşitliğiyle hesaplanmaktadır. Bu eşitlikte $I_i(\theta)$ i maddesinin bilgi fonksiyonunu, $P_i(\theta)$ madde tepki fonksiyonunu, $P'_i(\theta)$ ise madde tepki fonksiyonunun birinci türevini simgelemektedir. Madde bilgi fonksiyonları, toplanabilme özelliğine sahiptir. Bu özellik sayesinde test bilgi fonksiyonu hesaplanabilmektedir. Madde sayısının n olduğu bir test için test bilgi fonksiyonu $TI(\theta)$,

$$TI(\theta) = \sum_{i=1}^{n} I_i(\theta) \tag{2.6}$$

eşitliğiyle hesaplanmaktadır. Bu eşitlikte $I_i(\theta)$ madde bilgi fonksiyonunu simgelemektedir. Test bilgisi ve kestirimlerin standart hatası ters orantılıdır. Maksimum olabilirlik kestirim yönteminde kestirimlerin standart hatası $SE(\theta)$,

$$SE(\theta) = \frac{1}{\sqrt{TI(\theta)}} \tag{2.7}$$

eşitliğiyle hesaplanmaktadır (Embretson ve Reise, 2000).

Çoklu puanlanan maddelerde hem her bir kategori için hem de madde için bilgi fonksiyonu tanımlanmaktadır. Samejima (1969) çoklu puanlanan maddeler için geliştirilen tüm modellere uygulanabilen bilgi fonksiyonları tanımlamıştır. Herhangi bir i maddesinin $x = 0,1....,m_i$ kategorileri için kategori bilgi fonksiyonunu $I_{ix}(\theta)$ olarak tanımlamıştır. Madde bilgi fonksiyonu $I_i(\theta)$'yı hesaplamak için,

$$I_i(\theta) = \sum_{x=0}^{m_i} I_{ix}(\theta)\ P_{ix}(\theta) \tag{2.8}$$

eşitliğini tanımlamıştır. Bu eşitlikte $P_{ix}(\theta)$ her bir kategori için kategori yanıt eğrisini göstermektedir. Bu eşitlik matematiksel olarak sadeleştirilmiş ve madde bilgi fonksiyonunu hesaplamak için,

$$I_i(\theta) = \sum_{x=0}^{m_i} \frac{[P'_{ix}(\theta)]^2}{P_{ix}(\theta)} \quad (2.9)$$

eşitliği elde edilmiştir. Bu eşitlikte $P'_{ix}(\theta)$, kategori yanıt eğrisinin birinci dereceden türevidir. Test bilgi fonksiyonu ise madde bilgi fonksiyonlarının toplamı olarak tanımlanmıştır.

Bireyselleştirilmiş Bilgisayarlı Test

Bireyselleştirilmiş bilgisayarlı test (BBT), bireylere kendi yetenek düzeylerine uygun maddelerin uygulandığı bir bilgisayarlı test yaklaşımıdır. Tüm bireylere aynı maddelerin uygulandığı geleneksel kağıt-kalem testlerinin aksine, BBT uygulamalarında her bireye ayrı bir test uygulanmaktadır. BBT uygulamalarında, bireye uygulanan maddeler bireyin önceki maddelere verdiği yanıtlara göre belirlenen geçici yetenek kestirimine uygun olacak şekilde seçilmektedir. BBT'nin bu avantajı sayesinde, geleneksel kağıt-kalem testlerine göre daha kısa testlerle yüksek ölçme kesinliğine ulaşılabilmektedir (Weiss, 1982). BBT beş temel bileşenden oluşmaktadır. Bunlar madde havuzu, test başlatma yöntemi, madde seçim yöntemi, yetenek kestirim yöntemi ve sonlandırma kuralıdır (Thompson ve Weiss, 2011). Bu temel bileşenlerin dışında, BBT uygulamalarında iki önemli bileşen daha vardır. Bunlar madde kullanım sıklığı kontrol yöntemi (item exposure control method) ve kapsam dengeleme yöntemidir (content balancing method).

Madde Havuzu

Madde havuzu, parametre kestirimi uygun bir MTK modeline göre yapılmış maddelerden oluşmaktadır. Madde havuzunun özellikleri, BBT uygulamalarında önemli bir yere sahiptir. Madde havuzunun büyüklüğü (Stocking, 1994) ve kalitesi (Flaugher, 2000), BBT uygulamalarında yapılan yetenek kestirimlerinin doğruluğu üzerinde etkili olan faktörlerdir.

BBT uygulamalarında havuzda yer alan maddeler belirli bir süre boyunca tekrar tekrar kullanılmaktadır. Bazı maddelerin bu süreçte sık kullanılması test güvenliği için tehdit oluşturmaktadır. Test güvenliğini sağlamak için sık kullanılan ve açığa çıkma olasılığı yüksek olan maddeler bir süre sonra havuzdan çıkarılmaktadır. Havuzda

yeterince madde olmadığında BBT uygulamalarının devamlılığını sağlamak zorlaşacaktır. Bu nedenle BBT uygulamalarında havuz büyüklüğü önemli bir faktördür. Stocking (1994)'e göre sabit test uzunluğuna göre uygulanan BBT uygulamalarında havuz büyüklüğü, test uzunluğunun en az 12 katı olmalıdır. Maddelerin puanlanma şekli ideal havuz büyüklüğü üzerinde etkili olan diğer bir faktördür. Dodd, Ayala ve Koch (1995) e göre, ikili puanlanan maddeleri kullanan BBT uygulamaları için daha büyük havuz gerekmektedir. Madde bilgileri daha fazla olduğu için çoklu puanlanan maddeler kullanıldığında daha küçük havuzlar yeterli olabilmektedir. Ayrıca, kapsam geçerliği ve test güvenliğini sağlamak için kapsam dengeleme ve madde kullanım sıklığı kontrol yöntemleri kullanıldığında daha büyük havuzlara ihtiyaç vardır.

Madde havuzunun kalitesi havuzdaki maddelerin güçlük düzeyi, verdikleri bilgi düzeyi ve kapsamı temsil etme derecesi gibi faktörlerle ilişkilidir. Kapsam geçerliğinin sağlanabilmesi için havuzdaki maddeler ölçülen özelliği yeterince temsil etmelidir. Havuzda ölçülen tüm konu alanları için yeterli sayıda madde yer almalıdır. BBT uygulamalarında her bireye bireyselleştirilmiş ayrı bir test uygulandığı için havuzdaki maddeler geniş bir yetenek aralığını kapsamalıdır (Parshall, Spray, Kalohn ve Davey, 2002). Havuzdaki maddelerin güçlük parametreleri, farklı yetenek düzeyindeki bireylerin yeteneğine uygun olmalıdır. Havuzdaki maddelerin bilgi düzeyi ise testin amacına uygun olmalıdır. Norm dayanaklı testler için havuz bilgi fonksiyonu dikdörtgensel bir şekle sahip olmalı ve havuz tüm yetenek düzeylerinde yüksek bilgi vermelidir (Reckase, 1981). Ölçüt dayanaklı testlerde ise havuz bilgi fonksiyonu maksimum değerini kesme puanı civarında almalıdır (Parshall vd., 2002).

Başlatma Kuralı

BBT uygulamalarında madde seçimi yapılırken bireyin önceden uygulanan maddelere verdiği yanıtlara göre kestirilen geçici yetenek değeri kullanılmaktadır. Testin başlangıcında bireye ait herhangi bir geçici yetenek değeri olmadığı için madde seçim sürecini başlatacak bir başlatma kuralının belirlenmesi gerekmektedir. BBT uygulamalarında testi başlatmak için üç farklı yaklaşım bulunmaktadır. Bu yaklaşımlar orta güçlükte maddelerle başlama, kolay maddelerle başlama ve bireyin yeteneği hakkındaki ön bilgiye göre başlamadır (Parshall vd., 2002).

Testi orta güçlükteki maddelerle başlatmanın en kolay yolu bireylerin başlangıç yetenek düzeyini sıfıra eşitlemektir. Bu yöntemde tüm bireylerin başlangıç yetenek

düzeyi sıfıra sabitlenmektedir. Başlangıçta tüm bireylerin yeteneğinin sıfıra eşitlenmesi orta güçlükteki bazı maddelerin aşırı kullanılmasına yol açabilmektedir. Test güvenliğini arttırmak için bu yönteme alternatif başka bir yöntem geliştirilmiştir. Bu yöntemde bireylerin başlangıç yetenek düzeyi -0.5 ile 0.5 arasından rastgele seçilen bir değere eşitlenmektedir (Thompson ve Weiss, 2011). Kolay maddelerle başlama yaklaşımında, testin başlangıcında bireylere kolay maddeler uygulanmaktadır. Başlangıç yetenek düzeyini -1 ile 0 arasından rastgele seçilen bir değere eşitlemek, bu yaklaşımı uygulama yöntemlerinden birisidir (Hambleton ve Xing, 2006). Bireyin yeteneği hakkındaki bilgiye göre başlama yaklaşımında, başlangıç madde seçimi önceki test puanları ve derslerden aldığı notlar gibi bireyin yetenek düzeyi hakkında önceden bilinen bilgilere dayalı olarak yapılmaktadır. Bireyin başlangıç yetenek düzeyi bu bilgilere göre belirlenmektedir (Parshall vd., 2002).

Madde Seçim Yöntemi

BBT uygulamalarında bireylere kendi yeteneklerine uygun maddeler uygulanmaktadır. Uygulama sırasında bireyin yeteneğine uygun maddeleri uygulayabilmek için madde seçim yöntemleri kullanılmaktadır. Madde seçim yöntemleri, bazı algoritmalara dayalı olarak en uygun madde seçimini yaparak BBT'ye bireyselleştirilmiş olma özelliği kazandırmaktadırlar. BBT uygulamalarında madde seçimi için iki temel yaklaşım bulunmaktadır. Bunlar madde bilgisi yaklaşımı ve Bayesçi yaklaşımdır. Madde bilgisi yaklaşımı, Fisher bilgisini ya da Kullback-Leibler bilgisini kullanmaktadır. Bayesçi yaklaşım ise yetenek düzeyine ait önsel ve sonsal dağılımları kullanmaktadır (van Rijn, Eggen, Hemker ve Sanders, 2002).

Maksimum Fisher bilgisi (MFB) madde seçim yöntemi, en yaygın kullanılan madde bilgisi yaklaşımına dayalı yöntemlerden birisidir. Bu yöntemde, bireye geçici yetenek düzeyinde en çok bilgi veren madde uygulanmaktadır. Bu yöntemin başarılı bir şekilde uygulanabilmesi, geçici yetenek kestirimi ve gerçek yetenek düzeyinin birbirine yakın olmasına bağlıdır. BBT uygulamalarının başlangıcında gerçek ve geçici yetenek değerleri arasındaki fark fazla olduğunda, bu yöntem birey için uygun olmayan maddeleri seçebilmektedir (Ho ve Dodd, 2012). Ayrıca bu yöntem ayırıcılık düzeyi yüksek olan maddelerin çok sık kullanılmasına yol açabilmektedir (Chang ve Ying, 1999).

Bayesçi madde seçim yöntemlerinde bir sonraki madde seçilirken bireyin yeteneğine ait sonsal dağılım kullanılmaktadır. Owen (1975) tarafından geliştirilen yöntemde beklenen sonsal dağılımın varyansını en küçük yapan madde seçilmektedir. Bu yöntem, yeni madde seçimi sırasında önsel dağılım olarak son sonsal dağılımı doğrudan kullanmak yerine sonsal dağılımın ortalama ve varyansına göre oluşturulan normal dağılımı kullanmaktadır. Bilgisayar teknolojisinin gelişmesiyle birlikte sonsal dağılımı doğrudan kullanabilen yöntemler geliştirilmiştir. Van der Linden (1998), sonsal dağılımı bir sonraki madde seçiminde doğrudan kullanan dört farklı yöntem önermiştir. Bu yöntemler maksimum sonsal ağırlıklandırılmış bilgi (maximum posterior-weighted information), maksimum beklenen bilgi (maximum expected information), minimum beklenen sonsal varyans (minimum expected posterior variance) ve maksimum beklenen sonsal ağırlıklandırılmış bilgidir (maximum expected posterior-weighted information).

Yetenek Kestirim Yöntemi

BBT uygulamalarının en önemli bileşenlerinden birisi de yetenek kestirim yöntemidir. Yetenek kestirim yöntemleri hem geçici yetenek kestirim değerlerini hem son yetenek kestirim değerini belirlemek için kullanılmaktadır. Yetenek kestiriminde iki temel yaklaşım bulunmaktadır. Bunlar olabilirlik fonksiyonu yaklaşımı ve Bayesçi yaklaşımdır. Olabilirlik fonksiyonuna dayalı olarak kestirim yapan yöntemler, maksimum olabilirlik (maximum likelihood) ve ağırlıklandırılmış olabilirlik (weighted likelihood) kestirim yöntemleridir. Owen'ın yöntemi, beklenen sonsal dağılım (expected a posteriori) ve maksimum sonsal dağılım (maximum a posteriori) Bayesçi yaklaşımlardır.

Olabilirlik fonksiyonuna dayalı kestirim yapan yöntemler arasında en yaygın kullanılan yöntem maksimum olabilirlik (MO) yöntemidir. MO yöntemi, olabilirlik fonksiyonunun en büyük değerini aldığını θ yetenek düzeyini belirler. Bireyin kestirilen yeteneği, olabilirlik fonksiyonunu maksimum yapan yetenek değeridir. MO yönteminde kestirim yapmak için Newton-Raphson yöntemi gibi matematiksel yöntemler kullanılmaktadır. MO yöntemi, asimptotik özellikleri sayesinde uzun testlerde yansız kestirim yapma olanağı sağlamaktadır. Fakat bu yöntem, tüm yanıtların doğru ya da yanlış olduğu durumlarda yetenek kestirimi yapamamaktadır. Bu yöntemin kestirim yapabilmesi için bireyin yanıt örüntüsünde farklı yanıt kategorilerinin olması

gerekmektedir (Hambleton vd., 1991). Olabilirlik fonksiyonuna dayalı olarak kestirim yapan diğer bir yöntem ise ağırlıklandırılmış olabilirlik yöntemidir. Bu yöntemde, ağırlıklandırılmış olabilirlik fonksiyonunu maksimum yapan θ yetenek düzeyi yetenek kestirimi olarak belirlenmektedir (Warm, 1989).

Bayesçi yaklaşımlar, kestirim yapmak için yeteneğin sonsal dağılımını kullanmaktadırlar. Yeteneğin sonsal dağılımı, yanıt örüntüsünün olabilirlik fonksiyonu ile yeteneğin önsel dağılımının çarpımına dayalı olarak hesaplanmaktadır. Owen'ın yöntemi (Owen, 1975), yetenek kestirimi olarak yetenek sonsal dağılımının ortalamasını kullanmaktadır. Bu yöntem, her bir aşamada önsel dağılım olarak ortalaması ve standart sapması bir önceki sonsal dağılımın ortalama ve standart sapmasına eşit olan ayrı bir normal dağılım kullanmaktadır. Owen'ın yöntemine benzer şekilde, beklenen sonsal dağılım (BSD) yöntemi de yetenek kestirimi olarak yetenek sonsal dağılımının ortalamasını kullanmaktadır. Fakat BSD yönteminde önsel dağılım olarak her aşamada yeni bir normal dağılım belirlenmez. BSD, önsel dağılım olarak bir önceki sonsal dağılımı kullanmaktadır. Diğer bir Bayesçi yöntem olan maksimum sonsal dağılım (MSD) ise yetenek kestirimi olarak yetenek sonsal dağılımının modunu kullanmaktadır. MSD yöntemi de MO yöntemi gibi kestirim yaparken tekrara dayalı (iterative) matematiksel işlemleri kullanmaktadır. Bayesçi yaklaşımlar önsel dağılımın ortalamasına doğru yanlılık gösterebilmektedirler. Fakat Bayesçi yaklaşımlar tüm yanıt örüntüleri için yetenek kestirimi yapabilmektedir (Embreston ve Reise, 2000).

Sonlandırma Kuralı

BBT uygulamalarında testi sonlandırmak için iki temel sonlandırma kuralı bulunmaktadır. Bu kurallar, sabit ve değişen uzunluk sonlandırma kurallarıdır. Sabit uzunluk kuralında, bireylere önceden belirlenen belirli sayıda madde uygulandıktan sonra test sonlandırılmaktadır. Bu kurala göre sonlandırma yapıldığında tüm bireylere uygulanan madde sayısı aynı olmaktadır. Bu sonlandırma kuralının uygulanması kolaydır. Fakat bu sonlandırma kuralı uygulandığında tüm bireyler için aynı düzeyde ölçme kesinliği sağlamak zordur (Parshall vd., 2002).

Değişen uzunluk sonlandırma kuralında, bireylere uygulanan madde sayıları farklılık göstermektedir. Testi sonlandırmak için farklı değişen uzunluk sonlandırma kuralları kullanılmaktadır. Standart hata sonlandırma kuralı, θ değişimi sonlandırma kuralı ve minimum bilgi sonlandırma kuralı, BBT uygulamalarında kullanılan başlıca

değişen uzunluk sonlandırma kurallarıdır. Standart hata sonlandırma kuralında, yetenek kestiriminin standart hatası önceden belirlenen sabit bir değere ulaştığında test sonlandırılmaktadır. Kestirilen yetenek değerlerindeki değişimi dikkate alan θ değişimi sonlandırma kuralında, ardışık iki yetenek değeri arasındaki fark önceden belirlenen bir değerin altına düştüğünde test sonlandırılmaktadır. Minimum bilgi kuralında ise havuzdaki maddelerin bilgi düzeyleri kullanılmaktadır. Bu kurala göre, bireyin son kestirilen yetenek düzeyinde önceden belirlenen minimum bilgiyi verebilecek madde kalmadığında test sonlandırılmaktadır (Thompson ve Weiss, 2011).

Madde Kullanım Sıklığı Kontrol Yöntemleri

BBT uygulamalarında bireye geçici yetenek düzeyine en uygun maddeler uygulanmakta ve bireyin yeteneğini en doğru şekilde kestirmek amaçlanmaktadır. Yapılan yetenek kestirimlerinin doğruluğunu arttırmak için madde seçimi sırasında bireyin geçici yetenek düzeyinde en fazla bilgi veren madde seçilmektedir. Madde seçimi sırasında en fazla bilgi veren maddelerin seçilmesi, ayırt edicilik parametresi yüksek olan maddelerin sık kullanılmasına neden olmaktadır (Chang ve Ying, 1999). Ayırt edicilik parametresi yüksek maddelerin sık kullanılması, bu maddelerin zamanla açığa çıkmasına yol açmakta ve test güvenliğine zarar vermektedir. Bu nedenle bazı maddelerin sık kullanılmasını engellemek ve test güvenliğini sağlamak için madde kullanım sıklığı kontrol yöntemleri geliştirilmiştir.

Madde kullanım sıklığını kontrol yöntemleri, tesadüfi seçme (randomization) yöntemleri, koşullu seçme (conditional selection) yöntemleri, tabakalı (stratified) yöntemler, birleştirilmiş (combined) yöntemler ve çok aşamalı bireyselleştirilmiş test desenleri (multiple stage adaptive test designs) olarak beş kategoriye ayrılmaktadır. Tesadüfi seçme yöntemlerinde, geçici yetenek düzeyinde maksimum bilgi veren maddeler arasından rastgele seçilen bir madde bireylere uygulanmaktadır. Bilgi koşullu tesadüfi (BKT) yöntemi (randomesque) ve 5-4-3-2-1 yöntemi en yaygın kullanılan tesadüfi seçme yöntemleri arasındadır. Kingsbury ve Zara (1989) tarafından önerilen BKT yönteminde, öncelikle bireyin geçici yetenek düzeyinde en çok bilgi veren belirli sayıdaki madde seçilmektedir. Daha sonra bu maddeler arasından rastgele seçilen bir madde bireye uygulanmaktadır. Diğer bir tesadüfi seçme yöntemi olan 5-4-3-2-1 yönteminde (McBride ve Martin, 1983) ise ilk olarak bireyin başlangıç yetenek düzeyinde en çok bilgi veren beş maddeden biri rastgele seçilerek uygulanmaktadır.

İkinci aşamada bireyin kestirilen geçici yetenek düzeyimde en çok bilgi veren dört maddeden biri rastgele seçilerek uygulanmaktadır. Bu işlem üçüncü ve dördüncü maddelerin seçiminde de benzer şekilde devam etmektedir. Beşinci maddeden itibaren tesadüfi seçme işlemine son verilmekte ve bireyin geçici yetenek düzeyinde en çok bilgi veren madde uygulanmaktadır. Koşullu yöntemlerde her bir madde için madde kullanım sıklığı parametresi belirlenmektedir. Simülasyonlar sonucu belirlenen bu parametreler sayesinde, bilgi düzeyinden dolayı seçilme olasılığı yüksek olan maddelerin uygulanma olasılığını azaltmak amaçlanmaktadır. Tabakalı yöntemlerde, havuzdaki maddeler ayırt edicilik parametrelerine göre tabakalara ayrılmaktadır. Uygulamanın başında bireylere ayırt ediciliği en düşük tabakadaki maddeler uygulanmaktadır. Daha sonra kalan diğer tüm tabakalardan belli sayıda madde aşamalı olarak uygulanmaktadır. Tabakalı yöntemlerde, bilgi düzeyi yüksek maddelerin testin başlangıç aşamalarında gereksiz bir şekilde kullanılmasını engellemek amaçlanmaktadır. Birleştirilmiş yöntemler, tesadüfi seçme yöntemleri, koşullu seçme yöntemleri ve tabakalı yöntemler arasından seçilen farklı madde kullanım sıklığı kontrol yöntemlerinin birleştirilmesiyle oluşturulmaktadır. Çok aşamalı bireyselleştirilmiş test ise farklı bir bireyselleştirilmiş test yaklaşımıdır. Bu test yaklaşımında bireyler önceden oluşturulan panellere rastgele atandığı için bireyler farklı maddelerle karşılaşmaktadırlar. Böylelikle maddelerin sık kullanılması engellenmekte ve madde kullanım sıklığı kontrol altında tutulmaktadır (Georgiadou vd., 2007).

Kapsam Dengeleme Yöntemleri

BBT uygulamalarında madde seçimi bireyin yetenek düzeyine uygun yapıldığı için her bireye uygulama sırasında oluşan farklı test formları uygulanmaktadır. Test formları uygulama sırasında oluştuğu için test belirtke tablosundaki konu dağılımına uygun test formu oluşturmak zordur. Konu dağılımı açısından dengesiz test formlarının uygulanması, kapsam geçerliğine zarar vermektedir. Kapsam geçerliğinin sağlanması için tüm konuların dengeli bir şekilde uygulanması gerekmektedir. Konuların dengeli bir şekilde uygulanması için bazı kapsam dengeleme yöntemleri geliştirilmiştir. Kingsbury ve Zara (1989) tarafından geliştirilen kısıtlanmış BBT (constrained CAT) bu yöntemlerden biridir. Uygulaması kolay olduğu (Shin vd., 2009) için sıklıkla kullanılan bu yöntemde, belirtke tablosuna dayalı olarak her bir konu alanı için hedef yüzde değerleri belirlenmektedir. Test sırasında herhangi bir madde uygulandıktan sonra, daha

önce uygulanmış tüm maddelerin konu alanı dağılımına göre gerçek yüzde değerleri hesaplanmaktadır. Uygulamanın her aşamasında konu alanlarının gerçek ve hedef yüzde değerleri karşılaştırılmaktadır. Gerçek ve hedef yüzde değerleri arasındaki farkın en yüksek olduğu konu alanı belirlenmekte ve madde seçimi bu konu alanından yapılmaktadır.

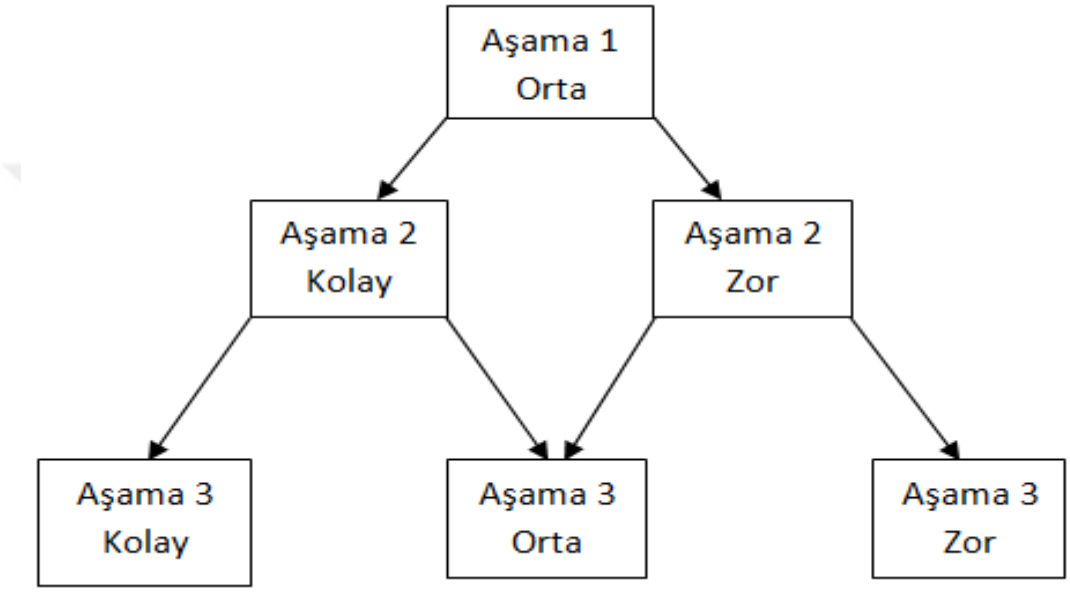
Çok Aşamalı Bireyselleştirilmiş Test

Çok aşamalı bireyselleştirilmiş test (ÇABT), geleneksel kağıt-kalem testleri ve BBT'nin bazı özelliklerini birleştirerek bu iki test yaklaşımı arasında dengeyi sağlayan test yaklaşımdır. BBT uygulamalarında bireyler zorlandıkları maddeleri atlayıp bir sonraki maddeye geçemezler. Ayrıca bireyler kendilerine uygulanan bir maddeyi yanıtladıktan sonra yanıtlarını değiştiremezler. Kağıt-kalem testlerine benzer şekilde, ÇABT uygulamalarında bireyler modül içinde zorlandıkları maddeyi atlayabilmekte ve daha önce verdikleri yanıtları değiştirebilmektedirler (Luecht ve Sireci, 2011). ÇABT uygulamalarında, BBT'ye benzer şekilde bireylerin yeteneklerine göre adaptasyon yapılmaktadır. Fakat adaptasyon BBT'de olduğu gibi madde düzeyinde değildir. ÇABT uygulamaları, aşamalardan oluşmakta ve aşamalar arası geçiş yapılırken bireylerin yeteneklerine uygun madde kümeleri uygulanmaktadır. Modül olarak adlandırılan bu madde kümeleri, uygulamadan önce oluşturulmaktadır. (Hendrickson, 2007; Jodoin vd., 2006)

Çok Aşamalı Bireyselleştirilmiş Testin Temel Bileşenleri

ÇABT uygulamaları dört temel bileşenden oluşmaktadır. Bunlar modül, panel, aşama ve yoldur (Luecht ve Nungester, 1998). Modüller belirli sayıda madde içeren madde kümeleridir. Modüller, önceden belirlenen hedef test bilgi fonksiyonlarına ve test belirtke tablosuna dayalı olarak oluşturulmaktadır. Modüller güçlük düzeyine göre kolay, orta ya da zor modül olarak adlandırılmaktadır. Modüllerin bir araya gelmesiyle paneller oluşmaktadır. Her panel birkaç tane aşamadan oluşmaktadır. Birinci aşamada tüm bireylere orta güçlük düzeyindeki yönlendirme modülü (routing module) uygulanmaktadır. Bireyler yönlendirme modülündeki performanslarına göre ikinci aşamadaki kolay, orta ya da zor modülden birisine yönlendirilmektedir. Bir bireyin yönlendirme modülünden son aşamadaki modüle kadar aldığı tüm modüller onun

geçtiği yolu (pathway) oluşturmaktadır (Zheng vd., 2012). ÇABT uygulamalarında birbirine paralel paneller bulunmaktadır. Bireyler bu panellere rastgele atanmaktadır. Böylelikle modül ve madde kullanım sıklığı kontrol altına alınmaktadır (Luecht, 2003). Aşama sayısı, her bir aşamadaki modül sayısı, modüllerdeki madde sayısı ve modüllerdeki maddelerin güçlük dağılımı değiştirilerek farklı panel desenleri oluşturulabilmektedir. Şekil 1'de 1-2-3 panel deseni örnek olarak verilmiştir. Bu desene göre oluşturulan bir panelde üç aşama, altı modül ve dört yol bulunmaktadır.



Şekil 1. 1-2-3 Panel Deseni

Test Oluşturma

ÇABT uygulamalarında test desenine ilişkin kararlar (panel sayısı, modül sayısı, her bir modüldeki madde sayısı vb.) verildikten sonra modül ve panellerin oluşturulması gerekmektedir. ÇABT uygulamalarında modül ve panellerin oluşturulması manuel olarak (Davis ve Dodd, 2003; Keng, 2008) ya da otomatik test oluşturma (automated test assembly; OTO) programlarıyla (Xing ve Hambleton, 2004; Zheng vd., 2012) yapılmaktadır. Panellerin oluşturulmasında aşağıdan-yukarıya yöntemi (bottom-up), yukarıdan-aşağıya yöntemi (top-down) veya bu iki yöntemin birleştirilmesiyle oluşturulan karma (mixture) yöntem kullanılmaktadır. Aşağıdan-yukarıya yönteminde tüm modüller test bilgi fonksiyonu ve kapsama yönelik belirlenen hedefleri sağlamaktadır. Bu nedenle aynı güçlük düzeyindeki modüller birbiri yerine

kullanılabilmektedir. Yukarıdan-aşağıya yönteminde yollar (pathways) test bilgi fonksiyonu ve kapsama yönelik hedefleri sağlamaktadır. Bu yöntemde modüller panel içinde ya da paneller arasında yer değiştiremez. Karma yöntemde hem modül düzeyinde hem de test düzeyinde hedefler belirlenmektedir (Luecht ve Nungester, 2000). Örneğin, her bir modül test bilgi fonksiyonuna yönelik belirlenen hedefleri sağlarken, yönlendirme modülünden başlayarak son aşamadaki modüllere giden tüm yollar kapsama yönelik hedefleri sağlamaktadır. Böylelikle tüm bireylere kapsam açısından dengeli testler uygulanabilmektedir.

Teknolojideki gelişmelerle birlikte test oluşturmada (test assembly) otomatik test oluşturma programları daha sık kullanılmaktadır. Otomatik test oluşturma (OTO) programları, lineer programlama gibi matematiksel optimizasyon işlemlerini kullanarak belirlenen hedefleri sağlayan en uygun modül ve panelleri oluşturmayı amaçlamaktadır. Bu programlar sayesinde çok sayıda panel ve modül eş zamanlı olarak oluşturulabilmektedir.

Yönlendirme Yöntemleri

ÇABT uygulamalarında aşamalar arası geçiş yapılırken bireylere uygulanacak modül yönlendirme yöntemlerine dayalı olarak belirlenmektedir. ÇABT uygulamalarında bireyleri bir aşamadan diğerine yönlendirme için farklı yöntemler vardır. Bunlar dinamik ve statik yöntemler olarak adlandırılmaktadır. Dinamik yöntemler, Fisher bilgisi ya da yeteneğin sonsal dağılımına göre yönlendirme yapmaktadır. Bu yöntemlerden birisi olan Maksimum Fisher bilgisi yönteminde, birey kestirilen yetenek düzeyinde en fazla bilgi veren modüle yönlendirilmektedir. Statik yönlendirme yöntemlerinde yönlendirme önceden belirlenen kesme puanlarına göre yapılmaktadır (Weissman, 2014). Statik yöntemlerden birisi olan tanımlanmış popülasyon aralığı (defined population interval) yönteminde belirli yollarda gidecek bireylerin oranı belirlenmektedir. Belirlenen bu orana göre kesme puanları hesaplanmaktadır. Yaklaşık maksimum bilgi (approximate maximum information) yönteminde ise kesme puanlarını belirlemek için kümülatif test bilgi fonksiyonları kullanılmaktadır (Luecht, Brumfield ve Breithaupt, 2006).

Bilgisayarlı Sınıflama Testi

Bilgisayarlı sınıflama testi (BST), bireyleri iki ya da daha fazla kategoriye doğru bir şekilde sınıflamayı amaçlayan bir bilgisayarlı test yaklaşımdır (Eggen, 2011). BBT ve ÇABT uygulamaları hem yetenek kestirmek hem de bireyleri sınıflamak amacıyla kullanılabilmektedir. BST ise sadece sınıflama amacıyla kullanılmaktadır.

BST uygulamalarında genel olarak iki tür madde seçim yöntemi bulunmaktadır. Bunlar yeteneğe ve kesme puanına dayalı madde seçim yöntemleridir. Yeteneğe dayalı madde seçim yönteminde bireye geçici yetenek düzeyinde en fazla bilgi veren madde uygulanmaktadır. Kesme puanına dayalı madde seçim yönteminde bireylere kesme puanında en fazla bilgi veren maddeler uygulanmaktadır (Thompson, 2009). BST uygulamalarında test sonlandırma üç yönteme göre yapılmaktadır. Bunlar yetenek güven aralığı (Eggen ve Straetmans, 2000; Kingsbury ve Weiss, 1983), aşamalı olasılık oran testi (Eggen ve Straetmans, 2000; Reckase, 1983) ve Bayesçi karar teorisi (Vos, 2000) yöntemleridir. Yetenek güven aralığı (ability confidence interval) yönteminde bireyin yetenek güven aralığı kesme puanının tamamen altında ya da üstünde kaldığında test sonlandırılmaktadır. Aşamalı olasılık oran testinde (sequential probability ratio test) bireyin yeteneğinin kesme puanının altında ve üstünde belirlenen yetenek değerlerine eşit olup olmadığı test edilmektedir. Bayesçi karar teorisi (Bayesian decision theory) yönteminde birey hakkında verilen sınıflama kararının kayıp ve kazançları dikkate alınmaktadır (Thompson, 2009). Bu çalışmada aşamalı olasılık oran testi kullanıldığı için bu yöntem aşağıda ayrıntılı olarak açıklanmıştır.

Aşamalı Olasılık Oran Testi

Aşamalı olasılık oran testi (AOOT), sınıflama yapmak için olabilirlik fonksiyonunu kullanmaktadır. AOOT uygulanırken her bir kesme puanı θ_k için H_0 ve H_1 hipotezleri kurulmaktadır. Bu hipotezler, kesme puanı etrafındaki değişmezlik bölgesinin sınırlarına göre oluşturulmaktadır. Yetenek düzeyi θ olan bir bireyin bulunduğu kategoriyi belirlemek için

$$H_0 : \theta \leq \theta_k - \delta = \theta_1 \quad (2.10)$$

$$H_1 : \theta \geq \theta_k + \delta = \theta_2 \quad (2.11)$$

hipotezleri test edilmektedir. Bu hipotezlerde yer alan θ_1 ve θ_2, değişmezlik bölgesinin alt ve üst sınırlardır. Değişmezlik bölgesinin alt ve üst sınırları, δ genişlik değerine göre belirlenmektedir. Yukarıdaki hipotezlerin doğruluğu hakkında karar verilirken kabul edilebilir hata oranları,

$$P(H_0 \text{ kabul } | H_0 \text{ doğru}) \geq 1-\alpha \tag{2.12}$$

$$P(H_0 \text{ kabul } | H_1 \text{ doğru}) \leq \beta \tag{2.13}$$

eşitsizlikleriyle tanımlanmaktadır. Bu eşitsizliklerde α 1.tip hata oranını ve β 2.tip hata oranını simgelemektedir. Hipotezleri test etmek için kullanılan test istatistiği, olabilirlik fonksiyonunun θ_1 ve θ_2 yetenek düzeylerindeki değerlerinin birbirine oranlanmasıyla hesaplanmaktadır. Bir bireyin k sayıda maddeye verdiği yanıtlar $\mathbf{x} = x_1, x_2, \ldots, x_k$ olmak üzere olabilirlik oranı,

$$LR_k(\theta_2, \theta_1 ; \mathbf{x}) = \frac{LR_k(\theta_2; \mathbf{x})}{LR_k(\theta_1; \mathbf{x})} \tag{2.14}$$

eşitliğiyle belirlenmektedir. Bu oranın yüksek olması bireyin θ yetenek düzeyinin kesme puanının üstünde olduğunu; düşük olması ise bireyin θ yetenek düzeyinin kesme puanının altında olduğunu göstermektedir. Hipotez testleri hakkındaki kararlar,

Eğer $\beta/(1-\alpha) < LR_k(\theta_2, \theta_1 ; \mathbf{x}) < (1-\beta)/\alpha$ ise teste devam et; (2.15)

Eğer $LR_k(\theta_2, \theta_1 ; \mathbf{x}) \leq \beta/(1-\alpha)$ ise H_0'ı kabul et; (2.16)

Eğer $LR_k(\theta_2, \theta_1 ; \mathbf{x}) \geq (1-\beta)/\alpha$ ise H_0'ı reddet, (2.17)

ölçütlere dayalı olarak verilmektedir (Eggen,1999). Bu ölçütlere göre bireyin kategorisi hakkında karar verebilmek için $LR_k(\theta_2, \theta_1 ; \mathbf{x}) \geq (1-\beta)/\alpha = A$ ya da $LR_k(\theta_2, \theta_1 ; \mathbf{x}) \leq \beta/(1-\alpha) = B$ koşullarının sağlanması gerekmektedir. Değişen uzunluktaki testlerde $B < LR_k(\theta_2, \theta_1 ; \mathbf{x}) < A$ olduğu durumlarda test devam etmektedir. Fakat AOOT, sabit uzunluktaki testlerde de kullanılabilmektedir (Parshall vd., 2002). Sabit uzunluktaki testlerde tüm maddeler uygulandıktan sonra $LR_k(\theta_2, \theta_1 ; \mathbf{x}) \geq A$ ise birey başarılı, $LR_k(\theta_2, \theta_1 ; \mathbf{x}) \leq B$ ise birey başarısız kategorisinde yer almaktadır. Eğer $B < LR_k(\theta_2, \theta_1 ; \mathbf{x}) < A$ koşuluyla karşılaşılırsa, olabilirlik oranının A ve B değerlerine olan uzaklıklarına göre karar verilmektedir. Olabilirlik oranı A değerine daha yakınsa birey başarılı, B değerine yakınsa başarısız kategorisinde yer almaktadır (Kim, 2010).

İlgili Araştırmalar

Patsula (1999), parametre kestirimi üç parametreli lojistik modele göre yapılmış 418 maddeden oluşan bir havuzu kullanarak yaptığı çalışmada BBT, ÇABT ve BLT test yaklaşımlarını ölçme kesinliği açısından karşılaştırmıştır. ÇABT uygulamalarında aşama sayısı, her bir aşamadaki modül sayısı ve modüllerde yer alan madde sayısını manipüle ettiği çalışmada 14 farklı koşul için simülasyon yapmıştır. Simülasyonlar sonucunda BBT'nin ÇABT ve BLT'ye göre daha doğru yetenek kestirimi yaptığını bulmuştur. ÇABT uygulamalarında, BLT'ye göre daha doğru yetenek kestirimleri yapıldığı sonucuna ulaşmıştır.

Davis ve Dodd (2003), yaptıkları çalışmada BBT ve ÇABT test yaklaşımlarını ölçme kesinliği, test güvenliği ve madde havuzu kullanımı açısından karşılaştırmışlardır. BBT simülasyonlarında maksimum Fisher bilgisi ve 0.10 logit içi (within .10 logit) madde seçim yöntemlerini kullanmışlardır. Ayrıca madde kullanım sıklığı oranları açısından en iyi sonuçları veren rastgele madde seçim yöntemini referans olarak kullanmışlardır. Çoklu puanlanan okuma metinlerinden oluşan madde havuzundaki maddelerin parametrelerini, kısmi puanlama modeline dayalı olarak kestirmişlerdir. Araştırma sonucunda BBT'nin ölçme kesinliği açısından ÇABT'a göre daha iyi sonuçlar verdiğini görmüşlerdir. BBT uygulamalarında elde edilen standart hata, ortalama mutlak fark ve RMSE değerleri daha düşük çıkmıştır. Fakat test güvenliği ve madde havuzu kullanımı açısından ÇABT daha iyi sonuçlar vermiştir. ÇABT simülasyonunda elde edilen maksimum madde kullanım sıklığı oranı, test çakışma oranı ve hiç kullanılmayan madde oranı BBT'ye göre daha düşük çıkmıştır.

Jodoin (2003), test uzunluğu, madde kullanım sıklığı kontrol düzeyi ve bazı madde havuzu özelliklerini manipüle ettiği çalışmasında BBT, ÇABT ve BLT test yaklaşımlarını ölçme kesinliği ve sınıflama doğruluğu açısından karşılaştırmıştır. Parametre kestirimi üç parametreli lojistik modele göre yapılmış ikili puanlanan maddelerden oluşan bir havuz kullanmıştır. Araştırma sonuçlarına göre, madde havuzunun ayırt ediciliği, test ve madde havuzunun kapsam açısından benzerliği ve test uzunluğu arttığında tüm test yaklaşımlarının ölçme kesinliği ve sınıflama doğruluğu artmaktadır. Madde kullanım sıklığı daha sınırlayıcı bir şekilde kontrol edildiğinde ölçme kesinliği ve sınıflama doğruluğu azalmaktadır. BBT, ölçme kesinliği ve sınıflama doğruluğu açısından ÇABT ve BLT'ye gore daha iyi sonuçlar vermiştir. ÇABT ise genel olarak BLT'ye gore daha iyi sonuçlar vermiştir.

Xing ve Hambleton (2004), havuz büyüklüğü ve havuzdaki maddelerin kalitesini manipüle ettikleri çalışmada BBT, ÇABT ve BLT test yaklaşımlarını sınıflama doğruluğu ve sınıflama tutarlılığı açısından karşılaştırmışlardır. Parametre kestirimi üç parametreli lojistik modele göre yapılmış altı farklı madde havuzu kullanmışlardır. Havuz büyüklüklerini 240 ve 480 olarak belirlemişlerdir. Madde havuzundaki maddelerin ortalama ayırt edicilik değeri için üç farklı düzey (0.60, 1.00, 1.40) belirlemişlerdir. Araştırma sonucunda, test yaklaşımı fark etmeksizin havuz büyüklüğü ve havuzdaki maddelerin kalitesi arttırıldığında sınıflama doğruluğu ve tutarlılığının arttığını bulmuşlardır. BBT, ÇABT ve BLT yaklaşımları, sınıflama doğruluğu ve tutarlılığı açısından benzer sonuçlar vermiştir. BBT, biraz daha iyi sonuçlar vermesine rağmen test yaklaşımları arasındaki fark çok az çıkmıştır.

Hambleton ve Xing (2006), parametre kestirimi üç parametreli lojistik modele göre yapılmış 600 maddeden oluşan bir havuzu kullanarak yaptıkları çalışmada BBT, ÇABT ve BLT test yaklaşımlarını sınıflama doğruluğu ve sınıflama tutarlılığı açısından karşılaştırmışlardır. ÇABT ve BLT simülasyonlarında hedef bilgi fonksiyonlarını manipüle etmişlerdir. Hedef bilgi fonksiyonlarını, hem bireylerin ortalama yetenek düzeyine hem de kesme puanına göre belirlemişlerdir. Araştırma soncunda, sınıflama doğruluğu ve tutarlılığı açısından en iyi sonuçları veren test yaklaşımının BBT olduğunu bulmuşlardır. ÇABT, hedef bilgi fonksiyonu kesme puanıyla eşlendiği koşullarda BLT'ye göre daha iyi sonuçlar vermiştir.

Keng (2008), test uzunluğu, havuz büyüklüğü ve bireylerin yetenek dağılımını manipüle ettiği çalışmasında BBT ve ÇABT test yaklaşımlarını ölçme kesinliği, test güvenliği ve havuz kullanımı açısından karşılaştırmıştır. Madde takımı (testlet) şeklindeki okuma parçalarından oluşan iki farklı havuz kullanmış ve madde takımlarının parametre kestirimini madde takımları tepki modeline (testlet response model) dayalı olarak yapmıştır. Bu havuzlarda 46 (1008 madde) ve 31 (741 madde) madde takımı yer almıştır. BBT simülasyonlarında, hem madde düzeyinde hem de madde takımı düzeyinde adaptasyon kullanmıştır. Araştırma sonucunda, BBT'nin ölçme kesinliği ve test güvenliği açısından ÇABT'a göre daha iyi sonuçlar verdiğini bulmuştur. ÇABT ise havuz kullanımı açısından daha iyi sonuçlar vermiştir.

Kim (2010), test uzunluğu ve kesme puanını manipüle ettiği çalışmasında BBT, ÇABT ve BST test yaklaşımlarını sınıflama doğruluğu, test güvenliği ve havuz kullanımı açısından karşılaştırmıştır. Hem ikili hem çoklu puanlanan maddeler oluşan 424 maddelik bir havuz kullanmış ve parametre kestirimini genelleştirilmiş kısmi

puanlama modeline dayalı olarak yapmıştır. BBT ve BST uygulamalarında madde kullanım sıklığını kontrol etmek için kademeli kısıtlı yöntemi (progressive restricted method) kullanmıştır. BST simülasyonlarında sınıflama kararı için aşamalı olasılık oran testini (AOOT) kullanmış ve madde seçimini kesme puanına dayalı olarak yapmıştır. Araştırma sonucunda, BBT ve BST'nin sınıflama doğruluğu açısından ÇABT'a göre daha iyi sonuçlar verdiğini bulmuştur. BBT, BST'ye göre biraz daha iyi sonuçlar vermesine rağmen aralarındaki fark az çıkmıştır. Test güvenliği açısından en iyi sonuçlara da BBT simülasyonlarında ulaşmıştır. Kademeli kısıtlı yöntemi, BBT uygulamalarında madde kullanım sıklığını etkili bir şekilde kontrol etmiştir. Havuz kullanımı açısından en iyi sonuçlara ÇABT simülasyonlarında ulaşmıştır.

Wang (2017), yaptığı çalışmasında BBT ve ÇABT'ı bu test yaklaşımlarının özelliklerine uygun olarak oluşturulan havuzlara dayalı olarak karşılaştırmıştır. Bu test yaklaşımlarını aynı havuz bağlamında karşılaştırmanın adil olmadığını iddia etmiş ve araştırmasında BBT ve ÇABT için büyük bir havuzdan iki farklı işlevsel havuz oluşturmuştur. Araştırma sonucunda, ÇABT 'ın ölçme kesinliği açısından BBT'ye göre biraz daha iyi sonuçlar verdiğini bulmuştur.

BÖLÜM 3

YÖNTEM

Bu bölümde araştırma modeli, simülasyon deseni, verilerin üretilmesi, verilerin analizi ve araştırmada ele alınan test yaklaşımlarının simülasyon koşullarıyla ilgili bilgilere yer verilmiştir.

Araştırma Modeli

Bu araştırmada bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT) ve bilgisayarlı sınıflama testi (BST) bireyselleştirilmiş test yaklaşımları Türkiye'deki dil sınavlarına uygunluk açısından karşılaştırılmıştır. Ayrıca bireyselleştirilmiş bir test olmayan bilgisayar lineer test (BLT) referans olarak kullanılmıştır. Bu test yaklaşımlarını sınıflama doğruluğu, test güvenliği ve madde havuzu kullanımı gibi ölçütler açısından karşılaştırabilmek için simülasyonlar yapılmıştır. Yapay veri üretilerek farklı koşullar altında BBT, ÇABT, BST ve BLT simülasyonları yapıldığı için bu araştırma bir simülasyon çalışmasıdır.

Simülasyon Deseni

Araştırmada BBT, ÇABT, BST ve BLT madde havuzu büyüklüğü ve test uzunluğunun değiştirildiği koşullarda sınıflama doğruluğu, test güvenliği ve madde havuzu kullanımı açısından karşılaştırılmıştır. Ayrıca BBT, ÇABT ve BLT ölçme kesinliği açısından karşılaştırılmıştır. Madde seçim yönteminin BST uygulamalarının sınıflama doğruluğu üzerindeki etkisini araştırabilmek için, bu test yaklaşımı iki farklı madde seçim yöntemiyle uygulanmıştır. Kesme puanında maksimum bilgi veren maddenin seçildiği BST (BST_KP) ve geçici yetenek düzeyinde maksimum bilgi veren maddenin seçildiği BST (BST_GY) uygulamaları için ayrı simülasyonlar yapılmıştır. Araştırmanın bağımsız değişkenleri bilgisayarlı test yaklaşımı, madde havuzu büyüklüğü ve test uzunluğudur. Araştırmanın bağımlı değişkenleri ise sınıflama doğruluğu, test güvenliği, madde havuzu kullanımı ve ölçme kesinliğidir.

Madde havuzu bireyselleştirilmiş test uygulamalarında önemli bir yere sahiptir. BBT uygulamalarında geniş bir yetenek aralığına uygun maddelerden oluşan büyük madde havuzuna ihtiyaç vardır (Segall, 2005). ÇABT uygulamalarında farklı test uzunluğu ve kapsam dengesini koşullarını sağlayan modül ve panellerin oluşturulabilmesi, madde havuzunun nitelik ve nicelik yönünden yeterli olmasına bağlıdır (Hendrickson, 2007). Bu nedenle bu araştırmanın bağımsız değişkenlerinden birisi madde havuzu büyüklüğüdür.

Test uzunluğu arttıkça ölçme sonuçlarının güvenirliği artmaktadır (Crocker ve Algina, 1986). Böylelikle bireyler hakkında verilen kararların doğruluğu artmaktadır. Test uzunluğu arttırıldıkça BBT, ÇABT ve BST yaklaşımlarının sınıflama doğruluklarının nasıl bir değişim gösterdiğini incelemek için test uzunluğu manipüle edilmiştir.

Araştırmada beş bilgisayarlı test yaklaşımı koşulu (BBT, ÇABT, BST_KP, BST_GY, BLT), üç madde havuzu büyüklüğü koşulu (400, 600, 800) ve üç test uzunluğu koşulu (27, 33, 39) dikkate alındığı için toplam 45 (5x3x3) koşul bulunmaktadır. Araştırmanın simülasyon deseni Tablo 1'de özetlenmiştir.

Tablo 1

Simülasyon Deseni

Manipüle Edilen Değişken	Düzey	Düzey Sayısı
Test Yaklaşımı	BBT/ ÇABT/ BST_KP/ BST_GY /BLT	5
Madde Havuzu Büyüklüğü	400/ 600/ 800	3
Test Uzunluğu	27/ 33/ 39	3
Toplam Koşul Sayısı		5x3x3=45

Türkiye'de kullanılan dil sınavları, içerik ve madde formatı açısından birbirine benzerdir. Bu çalışmada, bu sınavlardan biri olan Yüksek Öğretim Kurumları Dil Sınavı (YÖKDİL) ele alınmış ve BBT, ÇABT, BST ve BLT test yaklaşımları YÖKDİL bağlamında karşılaştırılmıştır. Araştırmada kullanılan madde havuzları YÖKDİL'in içerik ve madde sayısına dayalı olarak oluşturulmuştur. YÖKDİL'in bireyselleştirilmiş test olarak uygulanmaya başlanması durumda başlangıç aşamasından büyük bir madde havuzu olmayacaktır. Araştırmada bu durum dikkate alınarak madde havuzu büyüklükleri 400, 600 ve 800 olarak belirlenmiştir. YÖKDİL'de yer alan ortak köklü maddeler çoklu puanlanan madde olarak ele alındığı için test uzunluğu 62'ye düşmektedir. Bireyselleştirilmiş testler kullanıldığında geleneksel kağıt-kalem

testlerinin yaklaşık yarı uzunluğundaki testlerle aynı ölçme kesinliğine ulaşılmaktadır (Weiss, 1982). Bu durum dikkate alınarak test uzunlukları 27, 33 ve 39 olarak belirlenmiştir. ÇABT simülasyonlarında eşit uzunlukta modüller içeren üç aşama olacağı için test uzunlukları 3'ün katları olacak şekilde seçilmiştir.

Araştırmada test yaklaşımı, havuz büyüklüğü ve test uzunluğu manipüle edilirken bazı değişkenler ise sabit tutulmuştur. Kesme puanı, MTK modeli, yetenek kestirim yöntemi, kapsam dengesi kontrol yöntemi ve madde kullanım sıklığı kontrol yöntemi sabitlenen değişkenlerdir.

Kesme Puanının Belirlenmesi

BBT, ÇABT ve BST test yaklaşımlarına göre bireylerin İngilizce yeterlik açısından yeterli ya da yetersiz olarak sınıflanabilmesi için bir kesme puanının belirlenmesi gerekmektedir. Yükseköğretim Kurumunun doçentlik başvurusu için belirlediği kesme puanı 55'tir. Bu araştırmada kesme puanı belirlenirken bu puan dikkate alınmıştır. Sınıflamalı test araştırmalarında kesme puanı belirleme yaklaşımlarından birisi de ölçülen yeteneğin normal dağıldığı varsayımına dayalıdır (Kim, 2010; van Groen, Eggen ve Veldkamp, 2014). Kesme puanı bu varsayıma dayalı olarak belirlenmiştir. YÖKDİL puan ölçeği 0 ile 100 arasındadır. Yeteneğin normal dağıldığı varsayımına dayalı olarak bu ölçekteki puanlar ortalaması 0 standart sapması 1 olan yetenek ölçeğine dönüştürüldüğünde 55 puana denk gelen θ değeri 0.13'tür. Bu nedenle araştırmada kesme puanı olarak $\theta_k = 0.13$ kullanılmıştır.

MTK Modeli

İkili ve çoklu puanlanan maddeler için farklı MTK modelleri kullanılmaktadır. YÖKDİL, 80 tane ikili puanlanan çoktan seçmeli madde içermektedir. Bu maddeler içerisinde ortak köke dayalı maddeler bulunmaktadır. Ortak bir metindeki 5 boşluğun doldurulduğu 2 boşluk doldurma testi ve ortak bir okuma parçasına göre 3 sorunun sorulduğu 5 okuma parçası yer almaktadır. Madde takımları (testlet) olarak adlandırılan (Wainer ve Kiely, 1987) ortak köklü maddelerde MTK'nın yerel bağımsızlık varsayımı ihlal edilmektedir (Sireci vd., 1991). Yerel bağımsızlık ihlal edildiğinde ayırt edicilik parametreleri ve bilgi fonksiyonları olduğundan yüksek kestirilirken yetenek kestirimlerinin standart hataları gerçek değerinden düşük kestirilmektedir. Bu nedenle

MTK'ya dayalı yapılan birey ve madde parametre kestirimleri hatalı olmaktadır (Sireci vd., 1991; Wainer ve Wang, 2000). Alan yazında yerel bağımsızlık ihlalinin yol açtığı problemleri ortadan kaldırmak için madde takımlarını çoklu puanlanan madde olarak değerlendirme (Thissen, Steinberg ve Mooney, 1989) ve madde takımları tepki modelini (testlet response theory) kullanma (Wainer, Bradlow ve Wang, 2007) gibi yaklaşımlar bulunmaktadır. Alan yazından daha yaygın olduğu için (Davis ve Dodd, 2003; Davis, Pastor, Dodd, Chiang ve Fitzpatrick, 2003; Pastor, Dodd ve Chang, 2002) bu araştırmada madde takımları çoklu puanlanan madde olarak ele alınmıştır. Madde takımları çoklu puanlanan madde olarak ele alındığında madde havuzunda hem ikili hem de çoklu puanlanan maddeler yer almıştır. Bu nedenle MTK modeli olarak genelleştirilmiş kısmi puan modeli (GKPM) kullanılmıştır.

Yetenek Kestirim Yöntemi

MTK'ya dayalı yetenek kestirimi için maksimum olabilirlik (maximum likelihood estimation-MLE), beklenen sonsal dağılım (expected a posteriori-EAP) ve maksimum sonsal dağılım (maximum a posteriori-MAP) gibi yöntemler bulunmaktadır. Bayesçi yaklaşımlar olarak adlandırılan beklenen sonsal dağılım (BSD) ve maksimum sonsal dağılım (MSD) yöntemleriyle kestirim yapılabilmesi için bireylerin yetenekleri hakkında önsel bilgiye ihtiyaç vardır. Maksimum olabilirlik (MO), önsel bilgi olmadan kestirim yapabilmesine rağmen tüm maddeleri doğru ya da yanlış yanıtlayan bireylerin yetenek kestirimini yapamamaktadır (Embretson ve Reise, 2000) . Bu nedenle bu araştırmada Bayesçi yaklaşımlardan BSD kullanılmıştır. Önsel yetenek dağılımının standart normal dağılım (N~(0,1)) olduğu varsayılmıştır.

Kapsam Dengesi ve Madde Kullanım Sıklığı Kontrol Yöntemleri

Kapsam dengesini sağlamak için kısıtlanmış BBT (Kingsbury ve Zara, 1989), ağırlıklandırılmış sapmalar modeli (Stocking ve Swanson, 1993) ve gölge test yaklaşımı (van der Linden ve Reese, 1998) gibi farklı yöntemler bulunmaktadır. Kısıtlanmış BBT yöntemi en sık kullanılan (Kim, 2010) ve uygulaması kolay olan (Shin vd., 2009) kapsam dengesi kontrol yöntemidir. Bu nedenle araştırmada BBT, ÇABT ve BST simülasyonlarında kapsam dengesini sağlamak için kısıtlanmış BBT yöntemi kullanılmıştır.

Madde kullanım sıklığını kontrol etmek için farklı yöntemler geliştirilmiştir. Bunlar tesadüfi seçme (randomization) yöntemleri, koşullu seçme (conditional selection) yöntemleri, tabakalı (stratified) yöntemler, birleştirilmiş (combined) yöntemler ve çok aşamalı bireyselleştirilmiş test desenleri (multiple stage adaptive test designs) olarak beş kategoriye ayrılmaktadır (Georgiadou vd., 2007). Tesadüfi seçme yöntemlerinin programlanması ve kullanımı basittir (Davis, 2004). Bu nedenle araştırmada BBT ve BST simülasyonlarında madde kullanım sıklığını kontrol etmek için Kingsbury ve Zara (1989) tarafından geliştirilen bilgi koşullu tesadüfi (BKT) yöntemi (Randomesque) kullanılmıştır. Bu yöntemde bireye geçici yetenek düzeyinde maksimum bilgi veren 5 madde arasından rastgele seçilen bir madde uygulanmıştır.

Verilerin Üretilmesi

BBT, ÇABT, BST ve BLT simülasyonlarında kullanılacak madde havuzu ve birey yetenek parametreleri R programlama dili kullanılarak üretilmiştir.

Madde Havuzunun Oluşturulması

Araştırmada kullanılan madde havuzları YÖKDİL'in içerik ve madde sayısına dayalı olarak oluşturulmuştur. YÖKDİL, lisansüstü eğitime veya doçentlik sınavına başvuracak adayların dil yeterliğini belirlemek amacıyla geliştirilmiştir. Tüm adaylara aynı maddelerin uygulandığı Yabancı Dil Sınavının (YDS) aksine Fen Bilimleri, Sağlık Bilimleri ve Sosyal Bilimler gibi farklı alanlarda uygulanmaktadır.

Araştırmada farklı büyüklükteki üç madde havuzu oluşturulmuştur. Madde havuzu büyüklükleri 400, 600 ve 800' dür. YÖKDİL'de yer alan madde takımları çoklu puanlanan madde olarak ele alındığı için havuzlarda hem ikili hem de çoklu puanlanan maddeler yer almıştır.

YÖKDİL'de 80 tane ikili puanlanan madde yer almaktadır. Bu maddelerin içeriğe göre dağılımı Tablo 2'de verilmiştir.

Tablo 2

YÖKDİL Maddelerinin İçeriğe Göre Dağılımı

İçerik	n	%
Kelime Bilgisi	6	7.50
Dilbilgisi	14	17.50
Cümle Tamamlama	11	13.75
İngilizce-Türkçe Çeviri	6	7.50
Türkçe-İngilizce Çeviri	6	7.50
Paragraf Tamamlama	6	7.50
Anlam Bütünlüğünü Bozan Cümleyi Bulma	6	7.50
Okuma Parçası	15	18.75
Boşluk Doldurma Testi (Cloze Test)	10	12.50
Toplam	80	100

YÖKDİL ‘de yer alan boşluk doldurma testi ve okuma parçası maddeleri ortak köke dayalı olarak hazırlanmaktadır. Her biri 5 ikili puanlanan madde içeren 2 boşluk doldurma testi ve her biri 3 ikili puanlanan madde içeren 5 okuma parçası vardır. Madde takımı şeklinde sorulan boşluk doldurma testi ve okuma parçaları çoklu puanlanan madde olarak ele alınmıştır. Boşluk doldurma testleri 6 kategorili (0-5), okuma parçaları ise 4 kategorili (0-3) çoklu puanlanan madde olarak ele alınmıştır. Bu değişim yapıldıktan sonra YÖKDİL hem ikili hem de çoklu puanlanan maddeler içeren 62 maddelik bir karma test haline dönüşmektedir. Değişim yapıldıktan sonra maddelerin içerik ve kategoriye göre dağılımı Tablo 3’teki gibi olmaktadır.

Tablo 3

Değişiklikten Sonra YÖKDİL Maddelerinin İçerik ve Kategori Sayısına Göre Dağılımı

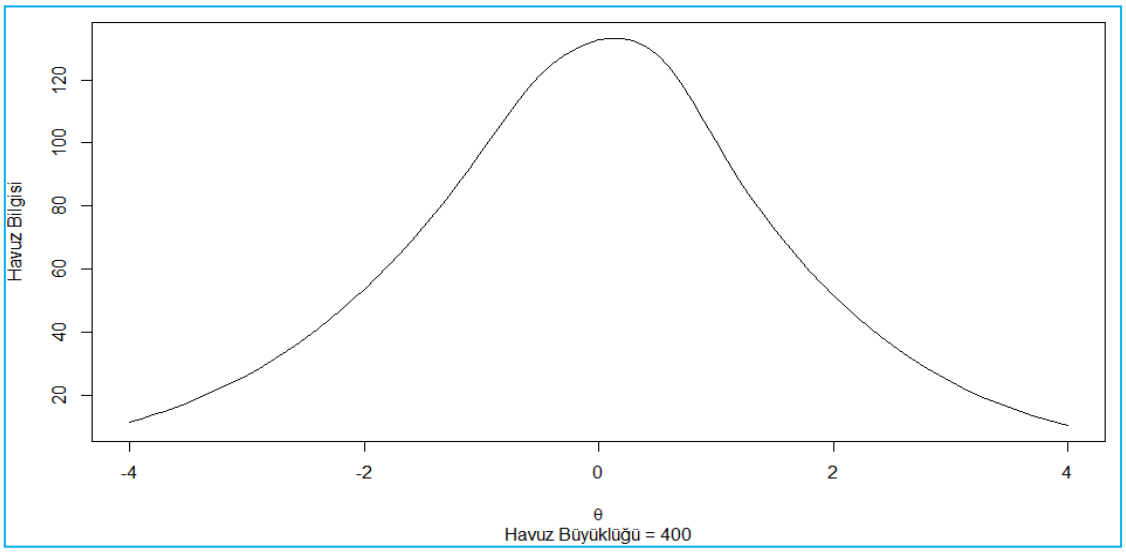
	Kategori Sayısı							
	2		4		6		Toplam	
İçerik	n	%	n	%	n	%	n	%
Kelime Bilgisi	6	9.68					6	9.68
Dilbilgisi	14	22.58					14	22.58
Cümle Tamamlama	11	17.74					11	17.74
İngilizce-Türkçe Çeviri	6	9.68					6	9.68
Türkçe-İngilizce Çeviri	6	9.68					6	9.68
Paragraf Tamamlama	6	9.68					6	9.68
Anlam Bütünlüğünü Bozan Cümleyi Bulma	6	9.68					6	9.68
Okuma Parçası			5	8.06			5	8.06
Boşluk Doldurma Testi					2	3.23	2	3.23
Toplam	55	88.72	5	8.06	2	3.23	62	100

Tablo 3'te görüldüğü üzere madde takımları çoklu puanlanan maddeye dönüştürüldüğünde test uzunluğu 62'ye düşmüştür. Madde takımları çoklu puanlanan madde olarak ele alındığında madde havuzunda hem ikili hem de çoklu puanlanan maddeler yer almıştır. Bu nedenle MTK modeli olarak genelleştirilmiş kısmi puan modeli (GKPM) kullanılmıştır. Madde havuzu oluşturulurken Tablo 3'te yer alan içerik ve madde formatı biçimlerinin oranları dikkate alınmıştır. Havuzlarda yer alan maddelerin içeriğe göre dağılımı Tablo 4'te verilmiştir. Havuzlara ait havuz bilgileri Şekil 2, Şekil 3 ve Şekil 4'te verilmiştir

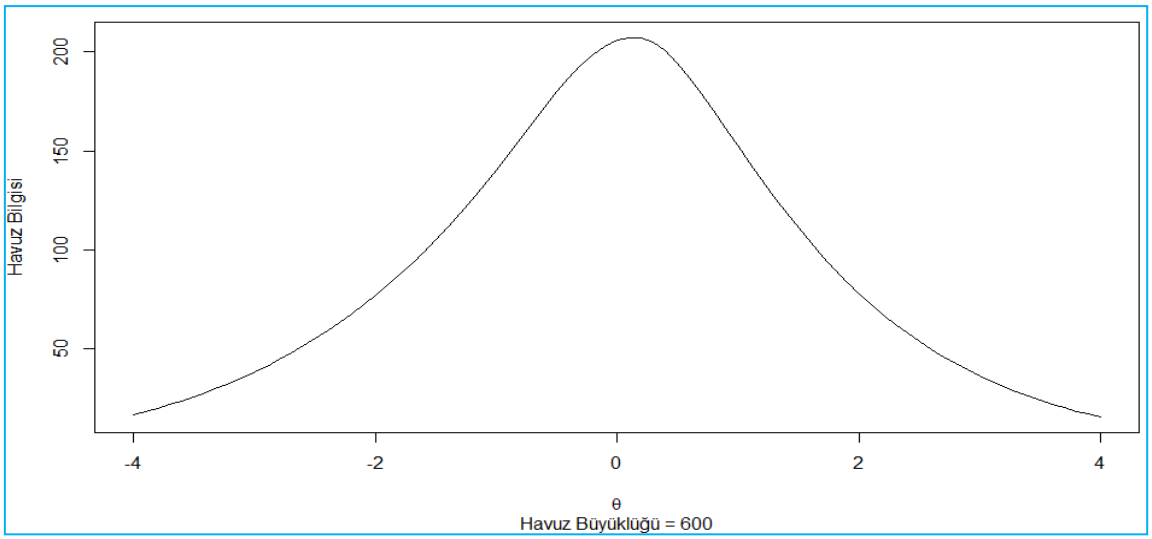
Tablo 4

Havuzlardaki Madde Sayılarının İçeriğe Göre Dağılımı

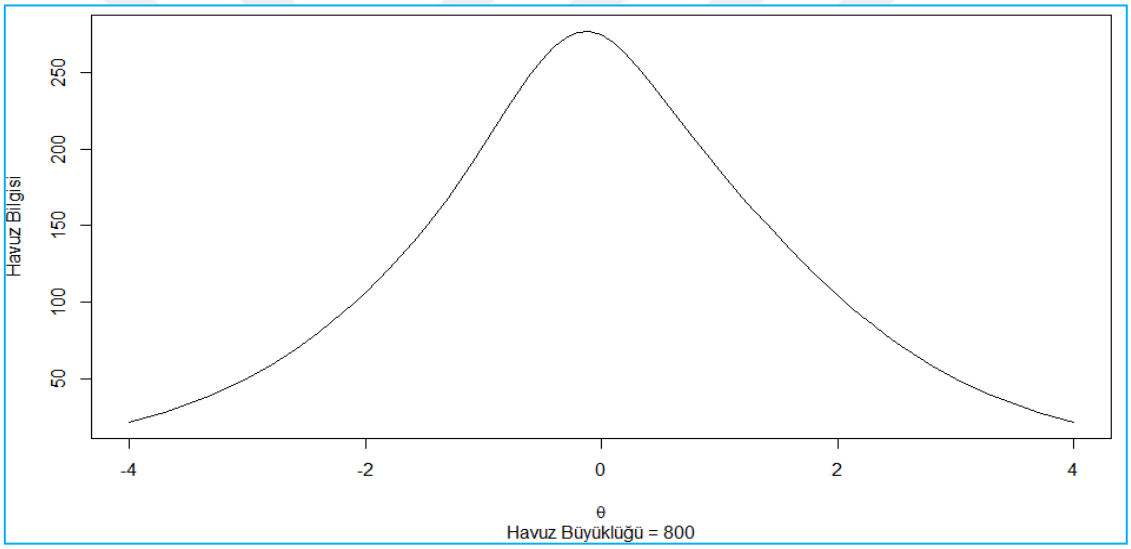
	Havuz Büyüklüğü		
İçerik	400	600	800
Kelime Bilgisi	39	58	77
Dilbilgisi	90	136	181
Cümle Tamamlama	70	107	143
İngilizce-Türkçe Çeviri	39	58	77
Türkçe-İngilizce Çeviri	39	58	77
Paragraf Tamamlama	39	58	77
Anlam Bütünlüğünü Bozan Cümleyi Bulma	39	58	77
Okuma Parçası	32	48	65
Boşluk Doldurma Testi	13	19	26



Şekil 2. Madde Sayısı 400 Olan Havuzun Bilgisi



Şekil 3. Madde Sayısı 600 Olan Havuzun Bilgisi



Şekil 4. Madde Sayısı 800 Olan Havuzun Bilgisi

Havuz bilgilerine ait şekiller (Şekil 2, Şekil 3, Şekil 4) incelendiğinde, tüm havuzların maksimum bilgiyi ortalama yetenek düzeyi etrafında verdiği görülmektedir. Havuz büyüklüğünün artmasına bağlı olarak bilgi artmaktadır. Havuz bilgi değerlerinin maksimum olduğu θ değerleri, 400, 600 ve 800 maddelik havuzlar için sırasıyla 0.10, 0.10 ve -0.10'dur.

Yetenek Parametrelerinin Elde Edilmesi

BBT, ÇABT, BST ve BLT simülasyonlarında kullanılan 1000 bireyin yetenek parametreleri, ortalaması 0 ve standart sapması 1 olan normal dağılımdan türetilmiştir.

Madde havuzundaki maddeler dikkate alınarak bu bireylerin yanıt örüntüleri türetilmiştir.

Simülasyon çalışmalarında örnekleme hatasını azaltmak için tekrarlar (replications) yapılması gerekmektedir. Bireyselleştirilmiş testlerle ilgili yapılan simülasyon çalışmalarında tekrar sayısının 10 ile 1000 arasında değiştiği görülmektedir (Cheng, Patton ve Shao, 2015; Kang ve Chang, 2016; Kim, 2010; Murphy, Dodd ve Vaughn, 2010; Spray ve Reckase, 1996; Yao, 2013; Yao, 2014; Widiatma, 2004; Zhang ve Li, 2016). Harwell, Stone, Hsu ve Kirisci (1996) ise örnekleme hatasını azaltmak için en az 25 tekrar yapılması gerektiğini belirtmişlerdir. Bu araştırmada örnekleme hatasını azaltmak için 30 tekrar yapılmıştır. Örneklemdeki 1000 bireyin her biri için 30 yanıt örüntüsü türetilmiş ve simülasyonlar her bir yanıt örüntüsü için tekrarlanmıştır. Verilerin analizi yapılırken 30 tekrar sonucu elde edilen değerlerin ortalaması kullanılmıştır.

BBT Simülasyonları

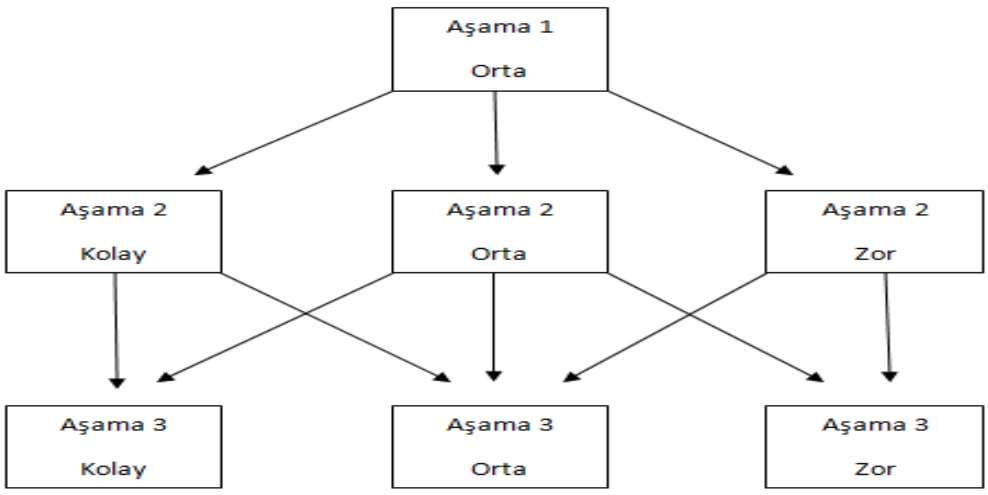
Araştırmada YÖKDİL'deki madde takımları çoklu puanlanan madde olarak alındığı için MTK modeli olarak genelleştirilmiş kısmi puanlama modeli kullanılmıştır. Bireylerin başlangıç yetenek kestirim değeri sıfır olarak belirlenmiştir. Madde seçiminde maksimum Fisher bilgisi (MFB) yöntemi kullanılmıştır. Madde kulanım sıklığını kontrol etmek için Kingsbury ve Zara (1989) tarafından geliştirilen bilgi koşullu tesadüfi (BKT) yöntemi (randomesque) kullanılmıştır. Bu yönteme göre, geçici yetenek kestirim değerinde en fazla bilgi veren 5 madde arasından rastgele seçilen madde bireye uygulanmıştır. Kapsam dengesinin sağlanması için Kingsbury ve Zara (1989) tarafından önerilen kısıtlanmış BBT (constrained CAT) yöntemi kullanılmıştır. Geçici ve son yetenek kestirimi beklenen sonsal dağılım (BSD) yöntemiyle yapılmıştır. BSD yönteminde önsel yetenek dağılımı standart normal dağılım (N~(0,1)) olarak kabul edilmiştir. Sonlandırma kuralı olarak sabit uzunluk yöntemi kullanılmıştır. Test uzunluğu araştırmanın bağımsız değişkenlerinden birisi olduğu için testi sonlandırmada üç farklı test uzunluğu (27, 33, 39) kullanılmıştır. BBT simülasyonları yapılırken R programla dilinde yer alan catR (Magis ve Raiche, 2012) paketindeki fonksiyonlardan yararlanılmıştır.

ÇABT Simülasyonları

Çok aşamalı bireyselleştirilmiş test uygulaması için öncelikli olarak modül ve paneller oluşturulmuştur. Modül ve paneller oluşturulduktan sonra her bir birey panellerden birine rastgele atanmış ve ÇABT uygulamaları gerçekleştirilmiştir.

Modül ve Panellerin Oluşturulması

Modül ve paneller otomatik test oluşturma (OTO) yöntemiyle yapılmıştır. ÇABT simülasyonları için 3 panel oluşturulmuştur. Her panel 3 aşamadan oluşmuştur. Araştırmada 1-3-3 panel deseni (Şekil 5) kullanılmıştır. Bu panel deseninde, birinci aşamada paneldeki tüm bireylere aynı yönlendirme testi uygulanmıştır. Yönlendirme testi orta güçlüktedir. İkinci ve üçüncü aşamada ise zorluk derecesine göre kolay, orta ve zor olan üçer modül yer almaktadır. İkinci aşama ve üçüncü aşama arasında bireyler bulundukları modülden sadece güçlük düzeyi açısından aynı ya da ardışık olan modüle geçebilmektedirler. Bu nedenle yönlendirme testinde başlayıp üçüncü aşamada biten toplam 7 yol vardır.



Şekil 5. 1-3-3 Panel Deseni

Panellerdeki modüllerin test uzunluğu eşittir. Test uzunluğu 27, 33 ve 39 olduğu durumlarda modüllerde sırasıyla 9, 11 ve 13 madde yer almaktadır. Teste az sayıda çoklu puanlanan madde yer aldığı için kapsam dengelemesini tüm modüllerde sağlamak zordur. Örneğin, 27 maddelik teste her bireye bir tane boşluk doldurma testi (6

kategorili madde**)** uygulanacağı için her modülde boşluk doldurma testi olması uygun değildir. Bu nedenle paneller oluşturulurken yukarıdan aşağı (top-down) ve aşağıdan yukarı (bottom-up) panel oluşturma yöntemlerinin birlikte kullanıldığı karma (mixture) panel oluşturma yöntemi kullanılmıştır. Kapsam dengelenmesi modül düzeyinde değil test düzeyinde sağlanmıştır. Panel içerisindeki yedi yolun her biri kapsam dengesini sağlamaktadır. Böylelikle testi alan tüm bireylere kapsam açısından dengeli testler uygulanmıştır. Test bilgi fonksiyonuna ilişkin hedefler ise modül düzeyinde sağlanmıştır.

Modüllerin güçlük düzeyleri hedef test bilgi fonksiyonlarına göre belirlenmiştir. Kolay, orta ve zor güçlük düzeyindeki modüllerin test bilgi fonksiyonları sırasıyla -1.00, 0.00 ve 1.00 yetenek düzeylerinde maksimum değeri vermektedirler. Modül ve panelleri oluşturmak için R programlama dilindeki IpSolveAPI (Konis, 2014) paketinde yer alan fonksiyonlardan faydalanılmıştır.

ÇABT Uygulamaları

ÇABT uygulamalarında birey 3 panelden birine rastgele atanmıştır. Bireyin ilk aşamadaki ve sonraki aşamalardaki yeteneği BSD yöntemine göre kestirilmiştir. BSD yönteminde önsel yetenek dağılım, standart normal dağılım (N~(0,1)) kabul edilmiştir. Bireyler sonraki aşamalara yönlendirilirken maksimum Fisher bilgisi yönlendirme yöntemi kullanılmıştır. Birey ilk aşamada kestirilen yetenek düzeyinde en fazla bilgi veren ikinci aşama modülüne yönlendirilmiştir. İkinci aşama tamamlandıktan sonra, birey kestirilen yetenek düzeyinde maksimum bilgi veren üçüncü aşama modülüne yönlendirilmiştir. Fakat ikinci aşamadan üçüncü aşamaya geçişte, bireyler sadece aynı ya da ardışık güçlük düzeyine sahip modüllere yönlendirilmişlerdir. Örneğin, kolay modülden sadece kolay ve orta güçlükteki modüle geçilebilir. ÇABT uygulamaları R programlama dilindeki mstR (Magis, Yan ve von Davier, 2018) paketinin fonksiyonlarına dayalı olarak araştırmacı tarafından yazılan fonksiyonlarla yapılmıştır.

BST Simülasyonları

Bilgisayarlı sınıflama testi simülasyonlarında aşamalı olasılık oran testi (AOOT) yöntemi kullanılmıştır. İngilizce yeterliğini belirlemek için kullanılacak kesme puanı 0.13'tür. AOOT yönteminde hipotez testleri için kullanılacak 1.tür ve 2.tür hata oranları

0.05 olarak belirlenmiştir. Alan yazındaki araştırmalara (Eggen, 2011; Lin, 2011; Thompson, 2009; van Groen, Eggen ve Veldkamp, 2016) benzer şekilde değişmezlik bölgesinin (indifference region) genişliği 0.40 (δ=0.20) olarak belirlenmiştir. Bu durumda, değişmezlik bölgesinin sınırları -0.07 ve 0.33 yetenek değerleridir.

AOOT yönteminde madde seçimi için iki yöntem bulunmaktadır. Bu yöntemler, geçici yetenek kestiriminde maksimum bilgi veren maddeyi seçme ve kesme puanında maksimum bilgi veren maddeyi seçmedir. Madde seçim yönteminin BST uygulamalarının sınıflama doğruluğu üzerindeki etkisini araştırabilmek için bu test yaklaşımı her iki madde seçim yöntemiyle de uygulanmıştır. Kesme puanında maksimum bilgi veren maddenin seçildiği BST (BST_KP) ve geçici yetenek düzeyinde maksimum bilgi veren maddenin seçildiği BST (BST_GY) uygulamaları için ayrı simülasyonlar yapılmıştır. Kapsam dengesini sağlamak için kısıtlanmış BBT (constrained CAT) yöntemi kullanılmıştır. Madde kullanım sıklığını kontrol etmek için bilgi koşullu tesadüfi (BKT) yöntemi (randomesque) kullanılmıştır. Testi sonlandırmada sabit test uzunluğu yöntemi kullanılmıştır. Simülasyon koşuluna göre test uzunluğu 27, 33 ve 39 olduğunda test sonlandırılmıştır. Test bittikten sonra değişmezlik bölgesinin sınırları olan -0.07 ve 0.33 yetenek değerlerinde hesaplanan olabilirlik fonksiyonu değerleri oranlanmıştır. Elde edilen olabilirlik oranına göre birey hakkında "yeterli/yetersiz" kararı verilmiştir. BST simülasyonları R programlama dilindeki catIrt (Nydick, 2014) paketinin fonksiyonlarına dayalı olarak araştırmacı tarafından yazılan fonksiyonlarla yapılmıştır.

BLT Simülasyonları

Bilgisayarlı lineer test simülasyonları için öncelikli olarak test formlarının oluşturulması gerekmektedir. Test formları, lineer programlamaya dayalı olarak otomatik test oluşturma (OTO) yöntemiyle oluşturulmuştur. Testler, geleneksel kağıt-kalem testlerine benzer şekilde orta güçlükte oluşturulmuştur. Testlerin hedef test bilgi fonksiyonu, $\theta = 0.00$ noktasında maksimum değeri vermektedir. Her bir test formu kapsam dengesini sağlayacak şekilde oluşturulmuştur. Madde kullanım sıklığının ÇABT simülasyonlarıyla karşılaştırılabilir olması için tüm test uzunluğu koşullarında üç farklı test formu oluşturulmuştur. Simülasyon sırasında bireyler üç test formundan birine rastgele atanmıştır. Böylelikle, madde kullanım sıklığı kontrol altına alınmıştır. Bireylerin yetenek kestirimi, beklenen sonsal dağılım (BSD) yöntemiyle yapılmıştır.

Verilerin Analizi

BBT, ÇABT, BST ve BLT yaklaşımlarının sınıflama doğruluğunu belirlemek için doğru sınıflama yüzdeleri karşılaştırılmıştır. BBT, ÇABT ve BLT uygulamalarında bireylerin İngilizce yeterlik düzeyi (yeterli/yetersiz) hakkındaki sınıflama kararı, kestirilen yetenek düzeylerine göre verilmiştir. Bireylerin yetenek kestirim değerleri kesme puanından küçük ise bireyler yetersiz olarak değerlendirilmiştir.

Dört test yaklaşımını test güvenliği açısından karşılaştırmak için madde kullanım sıklığı oranına ait betimsel istatistikler ve test çakışma oranı incelenmiştir. Madde kullanım sıklığı oranı, maddenin uygulanma sayısının testi alan toplam kişi sayısına bölünmesiyle hesaplanmaktadır (He, Diao ve Hauser, 2014). Dört test yaklaşımını madde havuzu kullanımı açısından karşılaştırmak için ise hiç kullanılmayan madde sayı ve yüzdesi incelenmiştir.

BBT, ÇABT ve BLT yaklaşımlarını ölçme kesinliği açısından karşılaştırmak için yanlılık, mutlak yanlılık, RMSE (root mean square error), ortalama standart hata ve uyum katsayısı (fidelity coefficient) incelenmiştir. N örneklem büyüklüğü, θ_i i bireyinin gerçek yetenek düzeyi ve $\hat{\theta}_i$ i bireyinin kestirilen yetenek düzeyi olmak üzere yanlılık, mutlak yanlılık ve RMSE;

$$Yanlılık = \frac{\sum_{i=1}^{N}(\hat{\theta}_i - \theta_i)}{N} \tag{3.1}$$

$$Mutlak\ Yanlılık = \frac{\sum_{i=1}^{N}|(\hat{\theta}_i - \theta_i)|}{N} \tag{3.2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{\theta}_i - \theta_i)^2}{N}} \tag{3.3}$$

formülleriyle hesaplanmaktadır. Uyum katsayısını($r_{\theta,\hat{\theta}}$) belirlemek için ise gerçek ve kestirilen yetenek düzeyleri arasındaki Pearson Momentler Çarpımı Korelasyon Katsayısı hesaplanmıştır.

BÖLÜM 4

BULGULAR VE YORUMLAR

Bu bölümde, belirlenen koşullara ilişkin yapılan simülasyonlar sonucu elde edilen bulgulara yer verilmiştir. Ayrıca elde edilen bulgulara dayalı olarak yapılan yorumlar raporlanmıştır.

Ölçme Kesinliğine İlişkin Bulgular ve Yorumlar

Bu bölümde bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT) ve bilgisayarlı lineer test (BLT) uygulamalarında ölçme kesinliğinin havuz büyüklüğü ve test uzunluğuna göre değişimine ilişkin bulgu ve yorumlara yer verilmiştir. Ölçme kesinliğinin göstergesi olarak yanlılık, mutlak yanlılık, ortalama standart hata, RMSE ve uyum katsayısı incelenmiştir.

İlk olarak, BBT uygulamalarında ölçme kesinliğinin havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiği incelenmiştir. BBT uygulamalarında yanlılık, mutlak yanlılık, ortalama standart hata, RMSE ve uyum katsayısına ilişkin elde edilen bulgular Tablo 5'te verilmiştir.

Tablo 5

BBT Uygulamalarında Ölçme Kesinliğinin Havuz Büyüklüğü ve Test Uzunluğuna Göre Değişimi

Havuz Büyüklüğü	Test Uzunluğu	Yanlılık	Mutlak Yanlılık	RMSE	SH	Uyum Katsayısı
400	27	0.000	0.233	0.295	0.292	0.956
	33	0.000	0.212	0.267	0.265	0.964
	39	-0.003	0.200	0.252	0.251	0.968
600	27	0.000	0.223	0.281	0.280	0.960
	33	0.001	0.199	0.252	0.253	0.968
	39	0.002	0.190	0.240	0.240	0.971
800	27	0.002	0.221	0.279	0.279	0.961
	33	0.000	0.203	0.256	0.253	0.967
	39	0.002	0.190	0.240	0.239	0.971

Tablo 5'te görüldüğü üzere, tüm BBT simülasyon koşullarında yanlılık sıfır ya da sıfıra çok yakın bir değerdir. Bu bulgu, simülasyon olarak uygulanan bireyselleştirilmiş bilgisayarlı testlerin yansız olduğunu göstermektedir. Havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça mutlak yanlılık, RMSE ve standart hata değerleri azalmaktadır. Alan yazınla (Cheng, Patton ve Shao, 2015; Deng, Ansley ve Chang, 2010; Doong, 2009) tutarlı olan bu bulgu, test uzunluğu arttıkça gerçek ve kestirilen θ değerlerinin birbirine yaklaştığını göstermektedir. Test uzunluğu arttıkça uyum katsayısı da artmaktadır. Fakat artış küçük olduğu için virgülden sonra binde birler basamağında görülmektedir. Diğer bir ifadeyle, test uzunluğunun arttırılması gerçek ve kestirilen yetenek değerleri arasındaki korelasyonda az da olsa bir artışa neden olmaktadır. Cheng ve diğerleri (2015), test uzunluğunu 15'ten 40'a çıkardıkları çalışmada uyum katsayısında daha dikkate değer artış bulmuşlardır. Bu çalışmada uyum katsayındaki artışın az olması, test uzunlukları arasındaki farkın yeterince büyük olmamasıyla ilişkili olabilir. Sabit test uzunluğunda, havuz büyüklüğünün 600 ve 800 olduğu koşullarda elde edilen mutlak yanlılık, RMSE ve standart hata değerleri havuz büyüklüğünün 400 olduğu koşullara göre daha düşüktür. Havuz büyüklüğü 600 ve 800 iken birbirine yakın mutlak yanlılık, RMSE ve standart hata değerlerine ulaşılmıştır. Bu bulguya göre BBT uygulamalarında havuz büyüklüğünü arttırmak ölçme sonuçlarının doğruluğunu her zaman arttırmayabilir. Uyum katsayısının ise havuz büyüklüğünden daha az etkilendiği görülmektedir. Havuz büyüklüğü 400 iken elde edilen uyum katsayıları, havuz büyüklüğü 600 ve 800 iken elde edilen uyum katsayılarından küçük olmasına rağmen aradaki fark azdır. Havuz büyüklüğü 600 ve 800 iken elde edilen uyum katsayıları hemen hemen birbirinin aynısıdır. Havuz büyüklüğünün uyum katsayısı üzerindeki etkisinin zayıf olduğunu gösteren bu bulgu, Leroux, Lopez, Hembry ve Dodd (2013) tarafından yapılan çalışmanın bulgularıyla benzerdir. Bu araştırmacılar, iki farklı havuz büyüklüğü (300 ve 540) kullandıkları çalışmada birbirine benzer uyum katsayıları elde etmişlerdir.

ÇABT uygulamalarında ölçme kesinliğinin havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiği incelenmiştir. ÇABT uygulamalarında ölçme kesinliğinin göstergesi olan yanlılık, mutlak yanlılık, ortalama standart hata, RMSE ve uyum katsayısına ilişkin bulgular Tablo 6'da verilmiştir.

Tablo 6

ÇABT Uygulamalarında Ölçme Kesinliğinin Havuz Büyüklüğü ve Test Uzunluğuna Göre Değişimi

Havuz Büyüklüğü	Test Uzunluğu	Yanlılık	Mutlak Yanlılık	RMSE	SH	Uyum Katsayısı
400	27	0.001	0.246	0.312	0.309	0.951
	33	0.000	0.223	0.284	0.281	0.960
	39	0.001	0.215	0.272	0.270	0.963
600	27	0.002	0.245	0.310	0.305	0.952
	33	0.001	0.221	0.281	0.278	0.961
	39	0.000	0.209	0.265	0.261	0.965
800	27	0.002	0.238	0.301	0.297	0.954
	33	-0.001	0.216	0.274	0.270	0.962
	39	-0.001	0.209	0.264	0.262	0.965

Tablo 6'da görüldüğü üzere, tüm ÇABT simülasyon koşullarında yanlılık sıfır ya da sıfıra çok yakın bir değerdir. Havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça mutlak yanlılık, RMSE ve standart hata değerleri azalmaktadır. Alan yazında daha önce yapılan çalışmalarla (Luo ve Kim, 2018; Sari ve Huggins-Manley, 2017) uyumlu olan bu bulgu, ÇABT uygulamalarında test uzunluğu arttıkça kestirilen yetenek değerlerinin gerçek değerlere yaklaştığını göstermektedir. Havuz büyüklüğü sabitken test uzunluğu arttıkça uyum katsayısı da artmaktadır. Test uzunluğundaki artışa bağlı olarak gerçek ve kestirilen θ değerleri arasındaki korelasyonun arttığını gösteren bu bulgu, alan yazındaki çalışmalarda (Keng, 2008; Sari ve Huggins-Manley, 2017) elde edilen bulgularla benzerdir. Sabit test uzunluğunda, havuz büyüklüğü arttırıldığında mutlak yanlılık, RMSE ve standart hata değerleri azalmaktadır. Özellikle 400 ve 800 maddelik havuzlar kullanıldığında elde edilen değerler arasında fark fazladır. Fakat 600 maddelik havuz kullanıldığında elde edilen değerler, test uzunluğu 27 ve 33 iken 400 maddelik havuz kullanıldığında elde edilen değerlere yakındır. Test uzunluğu 39 iken ise 600 ve 800 maddelik havuzlar kullanıldığında elde edilen değerler birbirine yakındır. Sabit test uzunluğunda, havuz büyüklüğü arttırıldığında elde edilen uyum katsayılarının ise birbirine yakın olduğu görülmektedir. Keng (2008), yaptığı çalışmada benzer bir sonuca ulaşmıştır. Bu bulguya göre ÇABT uygulamalarında, havuz büyüklüğünün gerçek ve kestirilen yetenek değerleri arasındaki korelasyon üzerinde önemli bir etkisi bulunmamaktadır.

Bireyselleştirme olmadığı için bireyselleştirilmiş testler için referans olarak kullanılan BLT uygulamalarında ölçme kesinliğinin havuz büyüklüğü ve test

uzunluğuna göre nasıl bir değişim gösterdiği incelenmiştir. BLT uygulamalarında ölçme kesinliğinin göstergesi olan yanlılık, mutlak yanlılık, ortalama standart hata, RMSE ve uyum katsayısına ilişkin bulgular Tablo 7'de verilmiştir.

Tablo 7

BLT Uygulamalarında Ölçme Kesinliğinin Havuz Büyüklüğü ve Test Uzunluğuna Göre Değişimi

Havuz Büyüklüğü	Test Uzunluğu	Yanlılık	Mutlak Yanlılık	RMSE	SH	Uyum Katsayısı
400	27	0.003	0.260	0.334	0.325	0.944
	33	-0.001	0.237	0.304	0.297	0.954
	39	0.002	0.226	0.289	0.283	0.958
600	27	0.001	0.254	0.326	0.319	0.946
	33	0.004	0.235	0.302	0.294	0.954
	39	0.003	0.223	0.286	0.279	0.959
800	27	0.002	0.254	0.327	0.319	0.946
	33	0.003	0.233	0.300	0.292	0.955
	39	0.000	0.221	0.285	0.276	0.959

Tablo 7'de görüldüğü üzere, BLT uygulamalarında havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça mutlak yanlılık, RMSE ve standart hata değerleri azalırken uyum katsayısı artmaktadır. Bu bulguya göre, BLT uygulamalarında test uzunluğu arttıkça kestirilen yetenek değerleri gerçek yetenek değerlerine yaklaşmaktadır. Diğer bir ifadeyle, daha doğru kestirim yapılmaktadır. Sabit test uzunluğunda, havuz büyüklüğü arttırıldığında mutlak yanlılık, RMSE ve standart hata azalmaktadır. Fakat değişim küçük olduğu için virgülden sonra binde birler basamağında görülmektedir. Özellikle 600 ve 800 maddelik havuzlar kullanıldığında elde edilen değerler arasındaki fark azdır. Havuz büyüklüğünün arttırılması sonucunda uyum katsayısında da önemli değişiklikler görülmemektedir. Sabit test uzunluğunda, havuz büyüklüğü değiştirildiğinde edilen uyum katsayıları birbirine çok yakındır.

BBT, ÇABT ve BLT yaklaşımlarını ölçme kesinliği açısından karşılaştırmak için farklı simülasyon koşullarında hesaplanan yanlılık, mutlak yanlılık, RMSE, standart hata ve uyum katsayısı değerlerinin ortalaması alınmıştır. Bu üç test yaklaşımının ölçme kesinliği açısından karşılaştırılmasına ilişkin bulgular Tablo 8'de verilmiştir.

Tablo 8

Ölçme Kesinliğinin Test Yaklaşımına Göre Değişimi

Test Yaklaşımı	Yanlılık	Mutlak Yanlılık	RMSE	SH	Uyum Katsayısı
BBT	0.000	0.208	0.262	0.261	0.965
ÇABT	0.001	0.225	0.285	0.281	0.959
BLT	0.002	0.238	0.306	0.298	0.953

Tablo 8'de görüldüğü üzere, BBT uygulamalarında elde edilen mutlak yanlılık, RMSE ve standart hata değerleri diğer test yaklaşımlarında elde edilen değerlere göre daha düşüktür. Bu bulguyla uyumlu olarak BBT uygulamalarında uyum katsayısı daha yüksek çıkmıştır. En yüksek mutlak yanlılık, RMSE ve standart hata değerleri BLT uygulamalarında elde edilmiştir. Uyum katsayısının en düşük olduğu test yaklaşımı da BLT' dir. Alan yazındaki çalışmalarla (Davis ve Dodd, 2003; Keng, 2008; Zheng ve Chang, 2015) paralellik gösteren bu bulgulara göre, gerçek değer en yakın yetenek kestirimleri BBT uygulamalarında elde edilmektedir. Daha sonra sırasıyla ÇABT ve BLT gelmektedir. Test yaklaşımları arasındaki bu farklılık, adaptasyon düzeyiyle açıklanabilir. BBT uygulamalarında madde düzeyinde adaptasyon vardır. Bireye uygulanacak her bir madde, bireyin geçici yetenek kestirim değerine göre seçilmektedir. ÇABT uygulamalarında ise modül düzeyinde adaptasyon vardır. Bireylere test başlangıcında orta güçlükte bir yönlendirme modülü uygulanmaktadır. Daha sonraki aşamalarda modül seçimi, bireylerin geçici yetenek kestirim değerlerine göre yapılmaktadır. BLT uygulamalarında ise bireylerin yeteneğine göre adaptasyon yoktur. Bireylere önceden hazırlanan test formları uygulanmaktadır.

Sınıflama Doğruluğuna İlişkin Bulgular ve Yorumlar

Bu bölümde bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT), bilgisayarlı sınıflama testi (BST) ve bilgisayarlı lineer test (BLT) uygulamalarında sınıflama doğruluğunun havuz büyüklüğü ve test uzunluğuna göre değişimine ilişkin bulgu ve yorumlara yer verilmiştir. Sınıflama doğruluğunun göstergesi olarak doğru sınıflama yüzdesi incelenmiştir.

BBT koşulları için yapılan simülasyonlar sonucunda sınıflama doğruluğunun havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiği incelenmiştir. Farklı

BBT simülasyon koşullarında elde edilen doğru sınıflama yüzdelerine ilişkin bulgular Tablo 9'da verilmiştir.

Tablo 9

BBT Uygulamalarında Sınıflama Doğruluğunun Havuz Büyüklüğü ve Test Uzunluğuna Göre Değişimi

Havuz Büyüklüğü	Test Uzunluğu	Doğru Sınıflama Yüzdesi
400	27	91.54
	33	92.63
	39	92.84
600	27	92.43
	33	93.09
	39	93.44
800	27	92.41
	33	92.90
	39	93.33

Tablo 9'da görüldüğü üzere, BBT uygulamalarında havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça doğru sınıflama yüzdesi artmaktadır. Bu bulgu, test uzunluğu arttıkça daha doğru yetenek kestirimleri yapıldığının diğer bir göstergesi olarak değerlendirilebilir. Ölçme kesinliğine ilişkin bulgularla uyumlu olan bu bulgu, test uzunluğu arttıkça gerçek ve kestirilen θ değerlerinin birbirine yaklaşmasıyla açıklanabilir. Çünkü kestirilen yetenek değeri gerçek değere yaklaştıkça bireylerin doğru sınıflanma olasılığı artacaktır. Sabit test uzunluğunda, havuz büyüklüğü 600 ve 800 olduğu koşullarda elde edilen doğru sınıflama yüzdesinin havuz büyüklüğü 400 iken elde edilen sınıflama yüzdesinden yüksek olduğu görülmektedir. Fakat havuz büyüklüğü 600 olduğunda doğru sınıflama yüzdesi havuz büyüklüğünün 800 olduğu duruma göre yüksek çıkmıştır. Bu durum, havuz bilgisinin maksimum değeri aldığı yetenek düzeyiyle ilişkili olabilir. Hambleton ve diğerleri (1991), bireyleri sınıflamayı amaçlayan testlerin kesme puanı etrafında daha çok bilgi vermesi gerektiği belirtmişlerdir. Embretson ve Reise (2000) ise sınıflama testlerinde güçlük düzeyi kesme puanı civarında olan maddelerden oluşan bir havuzun kullanılmasını önermişlerdir. Bu çalışmada büyüklüğü 600 olan havuz, en yüksek bilgiyi $\theta = 0.10$ değerinde verirken, 800 maddelik havuz en yüksek bilgiyi $\theta = -0.10$ değerinde vermektedir. Büyüklüğü 600 olan havuzun kesme puanı olan $\theta = 0.13$ yetenek düzeyi etrafında daha çok bilgi vermesi, doğru sınıflama yüzdesini arttırmış olabilir.

ÇABT koşulları için yapılan simülasyonlar sonucunda sınıflama doğruluğunun havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiği incelenmiştir. Farklı ÇABT simülasyon koşullarında elde edilen doğru sınıflama yüzdelerine ilişkin bulgular Tablo 10'da verilmiştir.

Tablo 10

ÇABT Uygulamalarında Sınıflama Doğruluğunun Havuz Büyüklüğü ve Test Uzunluğuna Göre Değişimi

Havuz Büyüklüğü	Test Uzunluğu	Doğru Sınıflama Yüzdesi
400	27	91.00
	33	91.91
	39	91.99
600	27	91.73
	33	92.21
	39	92.36
800	27	91.74
	33	92.15
	39	92.66

Tablo 10'da görüldüğü üzere, ÇABT uygulamalarında havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça doğru sınıflama yüzdesi artmaktadır. Kim (2010)'un bulgularıyla benzerlik gösteren bu bulgu, ÇABT uygulamalarında test uzunluğu arttıkça daha doğru ölçme sonuçlarına ulaşıldığını göstermektedir. Sabit test uzunluğunda, havuz büyüklüğü arttıkça genel olarak doğru sınıflama yüzdesinin arttığı görülmektedir. ÇABT uygulamalarında oluşturulan modül ve panellerin kalitesi, havuz büyüklüğü ve kalitesine bağlıdır (Hendrickson, 2007). ÇABT simülasyon koşullarında havuz büyüklüğü arttıkça sınıflama doğruluğunun artması bu durumla ilişkili olabilir.

BST koşulları için yapılan simülasyonlar sonucunda sınıflama doğruluğunun havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiği incelenmiştir. Hem **kesme puanında** maksimum bilgi veren maddenin seçildiği BST (BST_KP) hem de **geçici yetenek düzeyinde** maksimum bilgi veren maddenin seçildiği BST (BST_GY) uygulamaları için doğru sınıflama yüzdeleri hesaplanmıştır. BST_KP simülasyon koşullarında elde edilen doğru sınıflama yüzdelerine ilişkin bulgular Tablo 11'de, BST_GY simülasyon koşullarında elde edilen doğru sınıflama yüzdelerine ilişkin bulgular ise Tablo 12'de, verilmiştir.

Tablo 11

BST_KP Uygulamalarında Sınıflama Doğruluğunun Havuz Büyüklüğü ve Test Uzunluğuna Göre Değişimi

Havuz Büyüklüğü	Test Uzunluğu	Doğru Sınıflama Yüzdesi
400	27	91.76
	33	92.46
	39	92.83
600	27	92.48
	33	93.21
	39	93.49
800	27	92.65
	33	93.12
	39	93.31

Tablo 11'e göre, BST_KP uygulamalarında havuz büyüklüğünden bağımsız olarak test uzunluğu arttıkça doğru sınıflama yüzdesi artmaktadır. Bu bulgu, BST_KP uygulamalarında test uzunluğu arttıkça sınıflama doğruluğunun arttığını göstermektedir. Sabit test uzunluğunda, havuz büyüklüğü 600 ve 800 olduğu koşullarda elde edilen doğru sınıflama yüzdesi havuz büyüklüğü 400 iken elde edilen sınıflama yüzdesinden yüksektir. Diğer taraftan, test uzunluğu 33 ve 39 iken 600 maddelik havuz kullanıldığında elde edilen doğru sınıflama yüzdesi, 800 maddelik havuz kullanıldığında elde edilen doğru sınıflama yüzdesinden yüksek bulunmuştur. Bu bulgu, havuzların maksimum bilgi verdiği yetenek düzeylerinin kesme puanına olan uzaklığıyla açıklanabilir. Thompson (2009) hem kesme puanı etrafında maksimum bilgi veren hem de görece daha geniş bir θ aralığında yüksek bilgi veren iki havuz kullandığı çalışmasında, kesme puanı etrafında maksimum bilgi veren havuz kullanıldığında sınıflama doğruluğunun daha yüksek olduğunu bulmuştur. Bu çalışmada 600 maddelik havuz, en yüksek bilgiyi $\theta = 0.10$ değerinde verirken 800 maddelik havuz en yüksek bilgiyi $\theta = -0.10$ değerinde vermektedir. Büyüklüğü 600 olan havuzun bilgisinin kesme puanı olan $\theta = 0.13$'e yakın bir yetenek düzeyinde maksimum değere sahip olması, doğru sınıflama yüzdesini arttırmış olabilir. Böylelikle, 600 ve 800 maddelik havuzlar arasındaki havuz büyüklüğüne bağlı fark ortadan kalkmış olabilir.

Tablo 12

BST_GY Uygulamalarında Sınıflama Doğruluğunun Havuz Büyüklüğü ve Test Uzunluğuna Göre Değişimi

Havuz Büyüklüğü	Test Uzunluğu	Doğru Sınıflama Yüzdesi
400	27	91.63
	33	92.41
	39	92.57
600	27	92.52
	33	93.06
	39	93.51
800	27	92.12
	33	93.04
	39	93.22

Tablo 12'de görüldüğü üzere, BST_GY uygulamalarında havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça sınıflama doğruluğu artmaktadır. Sabit test uzunluğunda, havuz büyüklüğü 600 ve 800 olduğu koşullarda elde edilen doğru sınıflama yüzdeleri, havuz büyüklüğü 400 iken elde edilen sınıflama yüzdesinden yüksektir. Fakat havuz büyüklüğü 600 olduğunda elde edilen doğru sınıflama yüzdesi, havuz büyüklüğünün 800 olduğu duruma göre yüksek çıkmıştır. Kesme puanı civarında daha fazla bilgi veren 600 maddelik havuzun kullanılması, sınıflama doğruluğunu arttırmıştır.

BLT koşulları için yapılan simülasyonlar sonucunda sınıflama doğruluğunun havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiği incelenmiştir. Farklı BLT simülasyon koşullarında elde edilen doğru sınıflama yüzdelerine ilişkin bulgular Tablo 13'te verilmiştir.

Tablo 13

BLT Uygulamalarında Sınıflama Doğruluğunun Havuz Büyüklüğü ve Test Uzunluğuna Göre Değişimi

Havuz Büyüklüğü	Test Uzunluğu	Doğru Sınıflama Yüzdesi
400	27	91.58
	33	92.49
	39	92.84
600	27	92.38
	33	92.85
	39	93.13
800	27	92.42
	33	93.15
	39	93.22

Tablo 13'te görüldüğü üzere, havuz büyüklüğünden bağımsız olarak test uzunluğu arttıkça BLT uygulamalarında doğru sınıflama yüzdesi artmaktadır. Bu bulgu, test uzunluğundaki artışa bağlı olarak daha doğru ölçme sonuçları elde edilmesinin doğal bir sonucu olarak değerlendirilebilir. Sabit test uzunluğunda, havuz büyüklüğü arttıkça doğru sınıflama yüzdesinin arttığı görülmektedir. Bu bulgu, havuz büyüklüğüne bağlı olarak oluşturulan formların kalitesindeki değişimle ilişkilendirilebilir. Bilgisayarlı lineer test uygulamalarında test formları önceden oluşturulmaktadır. Havuz büyüklüğü arttıkça test formları oluşturulurken kullanılabilecek kaliteli madde alternatifi artacaktır. Böylelikle, farklı yetenek düzeylerine uygun ve ayırt ediciliği yüksek maddelerden oluşan test formları daha kolay bir şekilde oluşturulabilir.

BBT, ÇABT, BST_KP, BST_GY ve BLT yaklaşımlarını, sınıflama doğruluğu açısından karşılaştırabilmek için farklı simülasyon koşullarında elde edilen doğru sınıflama yüzdelerinin ortalaması alınmıştır. Simülasyon koşullarından bağımsız olarak bu test yaklaşımlarının doğru sınıflama yüzdelerine ilişkin bulgular Tablo 14'te verilmiştir.

Tablo 14

Sınıflama Doğruluğunun Test Yaklaşımına Göre Değişimi

Test Yaklaşımı	Doğru Sınıflama Yüzdesi
BBT	92.73
ÇABT	91.97
BST_KP	92.81
BST_GY	92.68
BLT	92.67

Tablo 14'te görüldüğü üzere, doğru sınıflama yüzdesi en yüksek olan test yaklaşımı BST_KP'dir (% 92.81), en düşük olan test yaklaşımı ise ÇABT'dır (% 91.97). Kesme puanı civarında maksimum bilgi veren madde seçim yöntemi kullanıldığında, bilgisayarlı sınıflama testinin sınıflama doğruluğunun diğer test yaklaşımlarına göre daha yüksek olduğu görülmektedir. Bireyselleştirilmiş bir test yaklaşımı olan ÇABT'ın sınıflama doğruluğu, herhangi bir bireyselleştirilmenin olmadığı BLT'den düşük çıkmıştır. Xing ve Hambleton (2004)'e göre bireyselleştirilmiş testlerin ölçme kesinliğinin fazla olması, sınıflama doğruluğu açısından her zaman avantaj sağlamayabilir. Sınıflama doğruluğu için önemli olan, kesme puanı etrafında ölçme kesinliğinin yüksek olmasıdır. Çok aşamalı bireyselleştirilmiş testin doğru sınıflama

yüzdesinin, bilgisayarlı lineer teste göre düşük olması bu durumla açıklanabilir. Bu çalışmada kesme puanı 0.13'tür. Bilgisayarlı lineer test simülasyon koşullarında oluşturulan tüm test formlarının bilgi fonksiyonu, maksimum bilgiyi $\theta = 0.00$ yetenek düzeyinde vermektedir. Formların maksimum bilgi verdiği yetenek düzeyi, kesme puanına ($\theta = 0.13$) yakın bir değerdir. Diğer taraftan, çok aşamalı bireyselleştirilmiş test koşullarında kolay ve zor modüller sırasıyla $\theta = -1.00$ ve $\theta = 1.00$ yetenek düzeyinde maksimum bilgi verecek şekilde oluşturulmuştur. Yönlendirme testinden sonra gelen ikinci ve üçüncü aşamasında kolay ve zor modüllerin uygulanması, sınıflama doğruluğunu olumsuz etkilemiş olabilir. Özellikle gerçek yetenek değeri kesme puanına yakın olan bireyler, yanlış sınıfta yer almış olabilir. BBT uygulamalarında ise ölçme kesinliğinin yüksek olması, sınıflama doğruluğuna yansımıştır. Madde düzeyinde adaptasyon olduğu için bireylerin gerçek yetenek değerlerine yakın kestirim değerlerine ulaşılması, sınıflama doğruluğunu arttırmıştır.

Test Güvenliğine İlişkin Bulgular ve Yorumlar

Bu bölümde bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT), bilgisayarlı sınıflama testi (BST) ve bilgisayarlı lineer test (BLT) uygulamalarında test güvenliğinin havuz büyüklüğü ve test uzunluğuna göre değişimine ilişkin bulgu ve yorumlara yer verilmiştir. Test güvenliğinin göstergesi olarak madde kullanım sıklığı oranına ait betimsel istatistikler ve test çakışma oranı incelenmiştir.

BBT için yapılan simülasyonlar sonucunda test güvenliğinin havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiği incelenmiştir. Test güvenliğinin göstergesi olan madde kullanım sıklığı oranına ait betimsel istatistikler ve test çakışma oranı Tablo 15'te verilmiştir.

Tablo 15

BBT Uygulamaları İçin Madde Kullanım Sıklığı Oranına Ait Betimsel İstatistikler

Havuz Büyüklüğü	Test Uzunluğu	Madde Kullanım Sıklığı				Test Çakışma Oranı
		Ortalama	SS	Minimum	Maksimum	
400	27	0.068	0.104	0.000	0.504	0.229
	33	0.083	0.121	0.000	0.584	0.258
	39	0.098	0.135	0.000	0.582	0.285
600	27	0.045	0.088	0.000	0.492	0.215
	33	0.055	0.102	0.000	0.554	0.244
	39	0.065	0.114	0.000	0.593	0.266
800	27	0.034	0.073	0.000	0.446	0.192
	33	0.041	0.086	0.000	0.526	0.219
	39	0.049	0.097	0.000	0.522	0.241

Tablo 15'e göre, BBT uygulamalarında havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça madde kullanım sıklığı oranı ortalaması ve test çakışma oranı artmaktadır. Test uzunluğu artırıldığında bireylere uygulanan madde sayısı arttığı için madde kullanım sıklığı oranları da artmıştır. Benzer şekilde, test uzunluğunun artırılması sonucunda bireylere uygulanan ortak madde sayısının artması test çakışma oranlarında artışa yol açmıştır. Deng ve diğerleri (2010) de yaptıkları çalışmada, test uzunluğu arttıkça test çakışma oranının arttığını bulmuşlardır. Test uzunluğu sabitken, havuz büyüklüğü arttıkça madde kullanım sıklığı oranı ortalaması, maksimum madde kullanım sıklığı ve test çakışma oranı azalmaktadır.

ÇABT uygulamalarında panel ve modüller oluşturulurken havuzdaki maddelerin tamamı kullanılmamaktadır. Dolayısıyla, simülasyonlar sırasında kullanılmayan maddelerin uygulanma olasılığı bulunmamaktadır. Bu nedenle alan yazına (Davis ve Dodd, 2003; Keng, 2008; Kim, Chung, Dodd ve Park, 2012) benzer şekilde ÇABT simülasyonlarının madde kullanım sıklığı oranları havuzun kullanılan bölümüne göre belirlenmiştir. ÇABT uygulamalarında madde kullanım sıklığı oranına ait betimsel istatistikler ve test çakışma oranı Tablo 16'da verilmiştir

Tablo 16

ÇABT Uygulamaları İçin Madde Kullanım Sıklığı Oranına Ait Betimsel İstatistikler

Havuz Büyüklüğü	Test Uzunluğu	Kullanılan Madde Sayısı*	Madde Kullanım Sıklığı				Test Çakışma Oranı
			Ortalama	SS	Minimum	Maksimum	
400	27	189	0.143	0.084	0.061	0.349	0.192
	33	231	0.143	0.081	0.068	0.349	0.189
	39	273	0.143	0.081	0.069	0.350	0.189
600	27	189	0.143	0.082	0.067	0.346	0.190
	33	231	0.143	0.082	0.070	0.348	0.190
	39	273	0.143	0.082	0.067	0.347	0.189
800	27	189	0.143	0.082	0.036	0.347	0.190
	33	231	0.143	0.081	0.076	0.346	0.188
	39	273	0.143	0.085	0.039	0.349	0.193

*ÇABT uygulamalarında modül ve paneller oluşturulurken kullanılan madde sayısı
SS: Standart sapma

Tablo 16'da görüldüğü üzere, ÇABT uygulamalarında madde kullanım sıklığı oranları test uzunluğu ve havuz büyüklüğünden bağımsızdır. Bu durum ÇABT uygulamalarında, modül ve panellerin uygulamadan önce oluşturulmasının bir sonucu olarak değerlendirilebilir. Test uzunluğunun artması, madde kullanım sıklığı oranının ortalamasını değiştirmemektedir. ÇABT uygulamalarında modül ve paneller oluşturulurken test uzunluğundaki artışla orantılı olarak modüllerde kullanılan madde sayısı da artmaktadır. Örneğin, test uzunluğunun 27 olduğu tüm koşullarda 189 madde kullanılırken test uzunluğu 39 olduğunda ise 273 madde kullanılmaktadır. Test uzunluğu ve kullanılan maddenin birlikte artması sonucunda ortalama madde kullanım sıklığı değişmemektedir. Havuz büyüklüğü ve test uzunluğu fark etmeksizin, maksimum kullanım sıklığının 0.35'i geçmediği görülmektedir. Bireylerin üç panelden birine rastgele atanması sonucu maksimum madde kullanım sıklığı sınırlandırılmıştır. Bu bulgu, ÇABT uygulamalarında madde kullanım sıklığının test uzunluğu ya da havuz büyüklüğüyle değil, panel sayısıyla ilişkili olduğunu göstermektedir. Farklı havuz büyüklüğü ve test uzunluğu koşullarında elde edilen test çakışma oranlarının birbirine çok yakın olduğu görülmektedir. Uygulama öncesinde oluşturulan ve ortak madde içermeyen üç panele bireylerin rastgele atanması sayesinde, test çakışma oranı kontrol altına alınmıştır.

BST için yapılan simülasyonlar sonucunda test güvenliğinin havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiği incelenmiştir. Hem **kesme puanında**

maksimum bilgi veren maddenin seçildiği BST (BST_KP) hem de **geçici yetenek düzeyinde** maksimum bilgi veren maddenin seçildiği BST (BST_GY) uygulamaları için madde kullanım sıklığı oranına ait betimsel istatistikler ve test çakışma oranı hesaplanmıştır. BST_KP simülasyon koşulları için elde edilen bulgular Tablo 17'de, BST_GY simülasyon koşullarında elde edilen bulgular ise Tablo 18'de verilmiştir.

Tablo 17

BST_KP Uygulamaları İçin Madde Kullanım Sıklığı Oranına Ait Betimsel İstatistikler

Havuz Büyüklüğü	Test Uzunluğu	Madde Kullanım Sıklığı				Test Çakışma Oranı
		Ortalama	SS	Minimum	Maksimum	
400	27	0.068	0.168	0.000	0.755	0.487
	33	0.083	0.197	0.000	0.848	0.550
	39	0.098	0.220	0.000	0.846	0.591
600	27	0.045	0.141	0.000	0.755	0.486
	33	0.055	0.165	0.000	0.844	0.550
	39	0.065	0.185	0.000	0.845	0.591
800	27	0.034	0.124	0.000	0.753	0.486
	33	0.041	0.145	0.000	0.847	0.550
	39	0.049	0.163	0.000	0.845	0.591

SS: Standart sapma

Tablo 17'de görüldüğü üzere, BST_KP uygulamalarında havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça madde kullanım sıklığı oranı ortalaması ve test çakışma oranı artmaktadır. Test uzunluğunun sabit olduğu durumda havuz büyüklüğü arttıkça madde kullanım sıklığı oranı ortalaması azalmaktadır. Fakat test çakışma oranı, havuz büyüklüğüne göre değişiklik göstermemiştir. Madde kullanım sıklığının maksimum değerlerinin yüksek olması, bazı maddelerin çok sık kullanıldığını göstermektedir. BST_KP uygulamalarında, kesme puanında en yüksek bilgiyi veren maddenin seçilmesi yaklaşımı kullanılmıştır. Thompson (2009)' a göre BST uygulamalarında kesme puanında maksimum bilgi veren maddenin seçilmesi, güçlük düzeyi kesme puanına yakın olan maddelerin sık kullanılmasına yol açmaktadır. Bu madde seçim yöntemi, bazı maddelerin aşırı kullanımına yol açmış olabilir. Madde kullanım sıklığını kontrol etmek için kullanılan yöntem, bazı maddelerin aşırı kullanılmasına engel olamamıştır. Test çakışma oranının havuz büyüklüğüne göre değişmemesi de bu madde seçim yöntemiyle ilişkilendirilebilir. Havuz büyüklüğü fark etmeksizin, bireylere kesme puanında en yüksek bilgiyi veren belirli maddeler uygulanmıştır. Bu durum, bireylere uygulanan ortak madde sayısını arttırmıştır.

Bireylere kesme puanı civarında yüksek bilgi veren sınırlı sayıdaki madde uygulandığı için test çakışma oranları yüksek çıkmıştır.

Tablo 18

BST_GY Uygulamaları İçin Madde Kullanım Sıklığı Oranına Ait Betimsel İstatistikler

Havuz Büyüklüğü	Test Uzunluğu	Madde Kullanım Sıklığı				Test Çakışma Oranı
		Ortalama	SS	Minimum	Maksimum	
400	27	0.068	0.104	0.000	0.506	0.228
	33	0.083	0.121	0.000	0.583	0.259
	39	0.098	0.135	0.000	0.581	0.284
600	27	0.045	0.088	0.000	0.490	0.215
	33	0.055	0.102	0.000	0.558	0.244
	39	0.065	0.115	0.000	0.590	0.266
800	27	0.034	0.073	0.000	0.450	0.192
	33	0.041	0.086	0.000	0.523	0.218
	39	0.049	0.097	0.000	0.531	0.241

SS:Standart sapma

Tablo 18'de görüldüğü üzere, BST_GY uygulamalarında havuz büyüklüğünden bağımsız olarak test uzunluğu arttıkça madde kullanım sıklığı oranı ortalaması ve test çakışma oranı artmaktadır. Test uzunluğu artırıldığında bireylere daha fazla sayıda madde uygulandığı için madde kullanım sıklığı oranları artmıştır. Benzer şekilde, test uzunluğunun artırılması sonucunda bireylere uygulanan ortak madde sayısının artması test çakışma oranlarını da arttırmıştır. Test uzunluğu sabitken, havuz büyüklüğü arttıkça madde kullanım sıklığı oranı ortalaması ve maksimum madde kullanım sıklığı azalmaktadır. Ayrıca BST_KP uygulamalarında farklı olarak, havuz büyüklüğü artıkça test çakışma oranı da azalmaktadır. BST_GY uygulamalarında, her bireye geçici yetenek düzeyine uygun maddeler uygulandığı için daha geniş bir yetenek aralığındaki maddeler kullanılmaktadır. Bu nedenle bireylere uygulanan ortak madde sayısı görece daha azdır. Havuz büyüklüğünün artırılması sonucunda ise bireylere uygulanan ortak madde sayısı biraz daha azalmıştır.

ÇABT uygulamalarına benzer şekilde, bilgisayarlı lineer test (BLT) uygulamalarında da sadece önceden hazırlanmış test formlarında yer alan maddeler bireylere uygulanmaktadır. Bu nedenle madde kullanım sıklığı sadece formlarda yer alan maddelere dayalı olarak hesaplanmıştır. BLT uygulamalarında madde kullanım sıklığı oranına ait betimsel istatistikler Tablo 19'da verilmiştir.

Tablo 19

BLT Uygulamaları İçin Madde Kullanım Sıklığı Oranına Ait Betimsel İstatistikler

Havuz Büyüklüğü	Test Uzunluğu	Kullanılan Madde Sayısı*	Madde Kullanım Sıklığı				Test Çakışma Oranı
			Ortalama	SS	Minimum	Maksimum	
400	27	81	0.333	0.013	0.318	0.347	0.334
	33	99	0.333	0.012	0.320	0.348	0.334
	39	117	0.333	0.016	0.316	0.351	0.334
600	27	81	0.333	0.013	0.318	0.347	0.334
	33	99	0.333	0.014	0.317	0.350	0.334
	39	117	0.333	0.014	0.317	0.349	0.334
800	27	81	0.333	0.013	0.319	0.350	0.334
	33	99	0.333	0.013	0.319	0.349	0.334
	39	117	0.333	0.013	0.319	0.350	0.334

*BLT uygulamalarında sadece formlardaki maddeler kullanılmaktadır.
SS: Standart sapma

Tablo 19'de görüldüğü üzere, BLT uygulamalarında madde kullanım sıklığı ve test çakışma oranı havuz büyüklüğü ve test uzunluğundan bağımsızdır. Madde kullanım sıklığı ve test çakışma oranı, form sayısına bağlı olarak değişmektedir. Araştırmada bireyler üç formdan birine rastgele atanmışlardır. Böylelikle, madde kullanım sıklığı kontrol altına alınmıştır.

BBT, ÇABT, BST_KP, BST_GY ve BLT yaklaşımlarını, test güvenliği açısından karşılaştırabilmek için farklı simülasyon koşullarında elde edilen madde kullanım sıklığı oranına ait betimsel istatistiklerin ve test çakışma oranlarının ortalaması alınmıştır. Simülasyon koşullarından bağımsız olarak bu test yaklaşımlarının test güvenliğine ilişkin bulgular Tablo 20'de özetlenmiştir.

Tablo 20

Madde Kullanım Sıklığının Test Yaklaşımına Göre Değişimi

Test Yaklaşımı	Madde Kullanım Sıklığı				Test Çakışma Oranı
	Ortalama	SS	Minimum	Maksimum	
BBT	0.060	0.102	0.000	0.534	0.239
ÇABT	0.143	0.082	0.062	0.348	0.190
BST_KP	0.060	0.167	0.000	0.815	0.542
BST_GY	0.060	0.102	0.000	0.534	0.239
BLT	0.333	0.013	0.318	0.349	0.334

SS: Standart sapma

Tablo 20'de görüldüğü üzere, maksimum madde kullanım sıklığı en yüksek olan test yaklaşımı BST_KP'dir (0.815). BST_KP uygulamalarında, bazı maddeler bireylerin % 80'inden fazlasına uygulanmıştır. Maksimum madde kullanım sıklığı açısından BST_KP'den sonra, BBT ve BST_GY gelmektedir (0.534). ÇABT ve BLT uygulamalarında ise maksimum madde kullanım sıklığının 0.35'i geçmediği görülmektedir. Test çakışma oranı en yüksek olan test yaklaşımı da BST_KP'dir. Test çakışma oranı en küçük olan test yaklaşımı ise ÇABT'dır. ÇABT uygulamalarında maksimum madde kullanım sıklığının 0.35'i geçmemesi, paneller oluşturulurken maddelerin sadece bir kez kullanılması ve bireylerin üç panele rastgele atanmasıyla açıklanabilir. BLT uygulamalarında ise formlarda ortak madde olmaması ve bireylerin üç formdan birine rastgele atanması, maddelerin maksimum kullanım sıklığını sınırlamıştır. ÇABT uygulamalarında kullanılan üç panelde ortak madde yer almamaktadır. ÇABT'ın test çakışma oranının düşük olması, bu durumla ilişkilendirilebilir. BST_KP uygulamalarında testi alan bireylerin % 81.5'ine (1000 bireyin 815'i) uygulanan madde(ler) olması ve test çakışma oranının yüksek olması ise madde seçim yöntemiyle açıklanabilir. Madde seçim sürecinde kesme puanında maksimum bilgi veren maddelerin seçilmesi, özellikle güçlük düzeyi kesme puanına yakın maddelerin sık kullanılmasına yol açmış olabilir. Bilgi koşullu tesadüfi madde sıklığı kontrol yöntemi, her seferinde kesme puanında maksimum bilgi veren beş madde arasından birini rastgele seçmesine rağmen bazı maddelerin aşırı kullanılmasına engel olamamıştır. Kim (2010), BBT ve BST_KP simülasyonlarında madde kullanım sıklığını kontrol etmek için kademeli kısıtlı (progressive-restricted) yöntemini (Revuelta ve Ponsoda, 1998) kullandığı çalışmada, bu çalışmadan farklı bir sonuca ulaşmıştır. Kademeli kısıtlı yöntemi sayesinde, hem BBT hem de BST_KP koşullarında maksimum madde kullanım sıklığını 0.35'e sabitlemiştir. Böylelikle, madde kullanım sıklığı 0.35'e ulaşan maddenin tekrar seçilmesinin önüne geçilmiştir. Bu çalışmada kullanılan madde kullanım sıklığı kontrol yöntemi, maksimum madde kullanım sıklığını sınırlama olanağı sağlamamaktadır.

Madde Havuzu Kullanımına İlişkin Bulgular ve Yorumlar

Bu bölümde bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT), bilgisayarlı sınıflama testi (BST) ve bilgisayarlı lineer test (BLT) uygulamalarında madde havuzu kullanımının havuz büyüklüğü ve test

uzunluğuna göre değişimine ilişkin bulgu ve yorumlara yer verilmiştir. Havuz kullanımının göstergesi olarak hiç kullanılmayan madde sayı ve yüzdesi incelenmiştir.

BBT uygulamalarında madde havuzu kullanımın havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiğini belirlemek için hiç kullanılmayan madde sayısı ve yüzdesi hesaplanmıştır. Farklı BBT koşulları için hiç kullanılmayan madde sayısı ve yüzdesi Tablo 21'de verilmiştir.

Tablo 21

BBT Uygulamalarında Havuz Kullanımı

		Hiç Kullanılmayan Madde	
Havuz Büyüklüğü	Test Uzunluğu	N	%
400	27	161	40.25
	33	148	37.00
	39	133	33.25
600	27	336	56.00
	33	319	53.17
	39	303	50.50
800	27	520	65.00
	33	502	62.75
	39	485	60.63

Tablo 21'de görüldüğü üzere, BBT uygulamalarında havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça hiç kullanılmayan madde sayısı ve yüzdesi azalmaktadır. Bu durum, uzun testlerde bireylere daha fazla madde uygulanmasının doğal bir sonucu olarak düşünülebilir. Bu bulgu, alan yazındaki çalışmalarla uyumludur. Deng ve diğerleri (2010) madde kullanım sıklığı kontrol yöntemi ve test uzunluğunu manipüle ettikleri çalışmada benzer sonuca ulaşmışlardır. Doong (2009) ise kullanılmayan madde oranı yerine kullanılan madde oranını dikkate aldığı çalışmasında, test uzunluğu arttıkça kullanılan madde oranının arttığını bulmuştur. Diğer taraftan havuz büyüklüğünün arttırılması sonucunda, hiç kullanılmayan madde yüzdelerinin önemli derecede arttığı görülmektedir. Pastor ve diğerleri de (2002) havuz büyüklüğü arttıkça kullanılmayan madde oranının arttığını bulmuşlardır.

Bu çalışmada test yaklaşımlarının farklı simülasyon koşullarındaki madde havuzu kullanımı, hiç kullanılmayan (uygulanmayan) madde sayı ve yüzdesine dayalı olarak değerlendirilmiştir. ÇABT uygulamalarında **panel ve modüller oluşturulurken** havuzdaki maddelerin bir kısmı kullanılmamıştır. Dolayısıyla, modül ve panellerde yer

almayan maddelerin simülasyonlar sırasında **uygulanma olasılığı** bulunmamaktadır. Bu nedenle alan yazına (Davis ve Dodd, 2003; Keng, 2008; Kim, Chung, Dodd ve Park, 2012) benzer şekilde ÇABT simülasyonlarının madde havuzu kullanımı, panel ve modüllerin oluşturulması sırasında kullanılan madde sayılarına dayalı olarak belirlenmiştir. Madde havuzu kullanımını değerlendirmek için **panel ve modüller oluşturulmasında** kullanılan maddeler arasında hiç **uygulanmayan** maddelerin sayı ve yüzdesi belirlenmiştir. Farklı ÇABT koşullarındaki madde havuzu kullanımına ilişkin bulgular Tablo 22'de verilmiştir.

Tablo 22

ÇABT Uygulamalarında Havuz Kullanımı

		Hiç Kullanılmayan Madde	
Havuz Büyüklüğü*	Test Uzunluğu	N	%
400	27	0	0.00
	33	0	0.00
	39	0	0.00
600	27	0	0.00
	33	0	0.00
	39	0	0.00
800	27	0	0.00
	33	0	0.00
	39	0	0.00

*ÇABT simülasyonlarında hiç kullanılmayan madde sayı ve yüzdesi, modül ve paneller oluşturulurken kullanılan maddelere dayalı olarak hesaplanmıştır.

ÇABT simülasyonlarında **modül ve paneller oluşturulurken** 27, 33 ve 39 maddelik testler için sırasıyla 189, 231 ve 289 madde kullanılmıştır. Tablo 22'de görüldüğü gibi, tüm havuz büyüklüğü ve test uzunluğu koşullarında, modül ve paneller oluşturulurken kullanılan bu maddeler arasında hiç uygulanmayan madde bulunmamaktadır. Bu bulgu, üç panelde yer alan modüllerin tamamının (21 modül) en az bir kez uygulandığını göstermektedir.

BST uygulamalarında madde havuzu kullanımın havuz büyüklüğü ve test uzunluğuna göre nasıl bir değişim gösterdiğini belirlemek için hiç kullanılmayan madde sayı ve yüzdesi hesaplanmıştır. Hem **kesme puanında** maksimum bilgi veren maddenin seçildiği BST (BST_KP) hem de **geçici yetenek düzeyinde** maksimum bilgi veren maddenin seçildiği BST (BST_GY) uygulamaları için hiç kullanılmayan madde sayı ve yüzdesi hesaplanmıştır. BST_KP simülasyon koşulları için elde edilen bulgular Tablo

23’te, BST_GY simülasyon koşullarında elde edilen bulgular ise Tablo 24’te verilmiştir.

Tablo 23

BST_KP Uygulamalarında Havuz Kullanımı

		Hiç Kullanılmayan Madde	
Havuz Büyüklüğü	Test Uzunluğu	N	%
400	27	333	83.25
	33	329	82.25
	39	323	80.75
600	27	535	89.17
	33	531	88.50
	39	525	87.50
800	27	735	91.88
	33	731	91.38
	39	725	90.63

Tablo 23’te görüldüğü üzere, BST_KP simülasyon koşullarında hiç kullanılmayan madde yüzdeleri % 80’in üzerindedir. Bu durum, BST_KP simülasyonlarında kullanılan madde seçim yöntemiyle ilişkilendirilebilir. Kesme puanında en yüksek bilgiyi veren maddenin seçilmesi yaklaşımı, havuzdaki maddelerin sınırlı bir kısmının kullanımına yol açmış olabilir. Tüm bireylere kesme puanında en yüksek bilgiyi veren maddelerin uygulanması sonucunda, kesme puanı ve çevresinde maksimum bilgi veren maddeler dışındaki maddeler uygulama dışında kalmış olabilir. Madde havuzu kullanımının havuz büyüklüğü ve test uzunluğuna göre değişimine ilişkin bulgular, BBT simülasyonlarında elde edilen bulgularla benzerdir. Havuz büyüklüğü fark etmeksizin test uzunluğu arttıkça hiç kullanılmayan madde sayısı ve yüzdesi azalmaktadır. Havuz büyüklüğünü arttırılması sonucunda hiç kullanılmayan madde yüzdelerinin arttığı görülmektedir.

Tablo 24

BST_GY Uygulamalarında Havuz Kullanımı

		Hiç Kullanılmayan Madde	
Havuz Büyüklüğü	Test Uzunluğu	N	%
400	27	161	40.25
	33	147	36.75
	39	133	33.25
600	27	335	55.83
	33	320	53.33
	39	303	50.50
800	27	521	65.13
	33	501	62.63
	39	484	60.50

Tablo 24'te görüldüğü üzere, BST_GY uygulamalarında havuz büyüklüğünden bağımsız olarak test uzunluğu arttıkça hiç kullanılmayan madde sayısı ve yüzdesi azalmaktadır. Bireylere uygulanan madde sayısı arttıkça kullanılan madde oranı artmıştır. Sabit test uzunluğunda havuz büyüklüğünü arttırılması sonucunda, hiç kullanılmayan madde yüzdeleri önemli derecede artmaktadır.

ÇABT uygulamalarına benzer şekilde, bilgisayarlı lineer test (BLT) uygulamalarında bireylere önceden hazırlanmış test formları uygulanmaktadır. Formların oluşturulması sırasında kullanılmayan maddelerin uygulanma olasılığı yoktur. Bu nedenle madde havuzu kullanımı sadece formlarda yer alan maddelere dayalı olarak değerlendirilmiştir. Farklı BLT koşulları için hiç kullanılmayan (uygulanmayan) madde sayı ve yüzdesi Tablo 25'te verilmiştir.

Tablo 25

BLT Uygulamalarında Havuz Kullanımı

		Hiç Kullanılmayan Madde	
Havuz Büyüklüğü*	Test Uzunluğu	N	%
400	27	0	0.00
	33	0	0.00
	39	0	0.00
600	27	0	0.00
	33	0	0.00
	39	0	0.00
800	27	0	0.00
	33	0	0.00
	39	0	0.00

*BLT simülasyonlarında hiç kullanılmayan madde sayı ve yüzdesi, test formları oluşturulurken kullanılan maddelere dayalı olarak hesaplanmıştır.

BLT simülasyonlarında **formlar oluşturulurken** 27, 33 ve 39 maddelik testler için sırasıyla 81, 99 ve 117 madde kullanılmıştır. Tablo 25'te görüldüğü üzere, tüm havuz büyüklüğü ve test uzunluğu koşullarında formlar oluşturulurken kullanılan bu maddeler arasında hiç uygulanmayan madde bulunmamaktadır. Bireyler üç formdan birine rastgele atandığı için uygulanmayan form bulunmamaktadır. Bu bulguya göre, form sayısı arttırıldığında orijinal havuzdaki maddelerin daha etkili bir şekilde kullanılması sağlanabilir.

BBT, ÇABT, BST_KP, BST_GY ve BLT yaklaşımlarını, madde havuzu kullanımı açısından karşılaştırabilmek amacıyla farklı test uzunluğu koşulları için hesaplanan hiç kullanılmayan madde sayılarının ortalaması alınmış ve bu ortalama değerlere göre yüzdeler belirlenmiştir. Test uzunluğu koşullarından bağımsız olarak, bu test yaklaşımlarının havuz kullanımına ilişkin bulgular Tablo 26'da özetlenmiştir.

Tablo 26

Madde Havuzu Kullanımının Test Yaklaşımına Göre Değişimi

Test Yaklaşımı	Havuz* Büyüklüğü	Hiç Kullanılmayan Madde	
		N	%
BBT	400	148	37.00
	600	319	53.17
	800	502	62.75
ÇABT	400	0	0.00
	600	0	0.00
	800	0	0.00
BST_KP	400	328	82.00
	600	530	88.33
	800	730	91.25
BST_GY	400	147	36.75
	600	319	53.17
	800	502	62.75
BLT	400	0	0.00
	600	0	0.00
	800	0	0.00

*Hiç kullanılmayan madde sayı ve yüzdeleri; ÇABT uygulamalarında modül ve paneller oluşturulurken kullanılan maddelere, BLT uygulamalarında ise test formları oluşturulurken kullanılan maddelere dayalı olarak hesaplanmıştır.

Tablo 26'da görüldüğü üzere, tüm havuz büyüklüğü koşullarında hiç kullanılmayan madde sayı ve yüzdesinin en yüksek olduğu test yaklaşımı BST_KP'dir. Hiç kullanılmayan madde sayı ve yüzdesinin en düşük olduğu test yaklaşımları ÇABT

ve BLT'dir. ÇABT uygulamalarında, modül ve paneller oluşturulurken kullanılan maddeler arasında hiç uygulanmayan madde bulunmamaktadır. BLT uygulamalarında ise test formları oluşturulurken kullanılan maddelerin tamamı uygulanmıştır. Bilgisayarlı sınıflama testi yaklaşımlarından BST_GY uygulamalarında hiç kullanılmayan madde sayı ve yüzdesi, BST_KP'ye göre düşüktür. BST_KP uygulamalarında tüm bireylere kesme puanında en yüksek bilgiyi veren maddelerin uygulanması sonucunda, havuzdaki maddelerin çoğunluğunun kullanım dışında kaldığı görülmektedir.

BÖLÜM 5

SONUÇ VE ÖNERİLER

Bu bölümde araştırma bulgularına dayalı olarak ulaşılan sonuçlara ve bu sonuçlara dayalı olarak önerilere yer verilmiştir. Öneriler, uygulayıcılara yönelik öneriler ve araştırmacılara yönelik öneriler olmak üzere iki başlık altında ele alınmıştır.

Sonuçlar

Bu çalışmada bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT) ve bilgisayarlı sınıflama testi (BST) bireyselleştirilmiş test yaklaşımları, Yüksek Öğretim Kurumları Dil Sınavı (YÖKDİL) bağlamında karşılaştırılmıştır. Bireyselleştirmenin olmadığı bilgisayarlı lineer test (BLT) ise bireyselleştirilmiş testlerin geleneksel testlere göre ne derece avantajlı olduğunu belirlemek için referans olarak kullanılmıştır. Madde seçim yönteminin BST uygulamalarının sınıflama doğruluğu üzerindeki etkisini araştırabilmek için bu test yaklaşımı iki farklı madde seçim yöntemiyle BST_KP ve BST_GY olarak uygulanmıştır. BST_KP uygulamalarında **kesme puanında** maksimum bilgi veren madde seçilirken, BST_GY uygulamalarında **geçici yetenek düzeyinde** maksimum bilgi veren madde seçilmiştir. Bu test yaklaşımları madde havuzu büyüklüğü ve test uzunluğunun değiştirildiği koşullarda sınıflama doğruluğu, test güvenliği ve madde havuzu kullanımı açısından karşılaştırılmıştır. BBT, ÇABT ve BLT ölçme kesinliği açısından da karşılaştırılmıştır. Araştırmada kullanılan madde havuzları, YÖKDİL'in içerik ve madde sayısına dayalı olarak oluşturulmuştur. Araştırma sonucunda aşağıdaki sonuçlara ulaşılmıştır;

1. Bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT) ve bilgisayarlı lineer test (BLT) test yaklaşımlarının hepsinde, test uzunluğu artıkça ölçme kesinliği artmaktadır. Test uzunluğu arttırıldığında gerçek değere daha yakın yetenek kestirimleri yapılmaktadır. Fakat test uzunluğundaki artışın gerçek ve kestirilen yetenek değerleri arasındaki korelasyon üzerindeki etkisi görece daha azdır.

2. BBT uygulamalarında havuz büyüklüğü 800 ve 600 olduğu koşullardaki ölçme kesinliği, havuz büyüklüğünün 400 olduğu koşullardaki ölçme kesinliğinden yüksektir. Fakat 600 ve 800 maddelik havuzlar kullanıldığı koşullarda ölçme kesinliği açısından önemli farklılık bulunmamaktadır. BBT uygulamalarında havuz büyüklüğünü arttırmak, ölçme kesinliğini her zaman arttırmayabilmektedir. Yeterli büyüklükteki bir havuza yeni maddeler eklenmesi, ölçme kesinliği üzerinde önemli değişikliğe yol açmamaktadır. ÇABT uygulamalarında havuz büyüklüğü arttıkça ölçme kesinliği artmaktadır. Fakat ölçme kesinliğindeki artışın büyüklüğü, havuz büyüklüğündeki artış ve test uzunluğuna göre değişmektedir. Görece kısa test uzunluğunda (27 ve 33), 400 ve 600 maddelik havuzların kullanıldığı koşullar arasındaki fark azdır. Test uzunluğu 39 iken ise 600 ve 800 maddelik havuzların kullanıldığı koşullar arasındaki fark azdır. Havuz büyüklüğünün 400 ve 800 olduğu koşullar arasındaki fark ise daha açık bir şekilde görülmektedir. BLT uygulamalarında havuz büyüklüğü arttıkça ölçme kesinliği genel olarak artmaktadır. Fakat havuz büyüklüğü 600 ve 800 olduğu koşullarda ölçme kesinliği açısından önemli farklılık yoktur.

3. Ölçme kesinliği açısından en iyi test yaklaşımı BBT'dir. Madde düzeyinde adaptasyon olduğu için BBT uygulamalarında, gerçek değerlere daha yakın yetenek kestirimleri yapılmıştır. Ölçme kesinliği açısından BBT'den sonra gelen test yaklaşımı, modül düzeyinde adaptasyon yapılan ÇABT'dır. BLT ise ölçme kesinliği açısından son sırada yer almaktadır. Bireylerin yeteneklerine göre adaptasyon düzeyi arttıkça ölçme kesinliği artmıştır.

4. Bireyselleştirilmiş bilgisayarlı test (BBT), çok aşamalı bireyselleştirilmiş test (ÇABT) , bilgisayarlı sınıflama testi (BST) ve bilgisayarlı lineer test (BLT) test yaklaşımlarının hepsinde, test uzunluğu artıkça sınıflama doğruluğu artmaktadır. BBT, ÇABT ve BLT uygulamalarında test uzunluğu arttırıldığında daha doğru kestirimler yapılması, sınıflama doğruluğunu arttırmıştır.

5. BBT ve BST uygulamalarında, havuz büyüklüğünün arttırıldığında sınıflama doğruluğu her zaman artmayabilmektedir. Havuz büyüklüğü 800 ve 600 olduğu koşullardaki sınıflama doğruluğu, havuz büyüklüğünün 400 olduğu koşullardaki sınıflama doğruluğundan yüksektir. Fakat 600 maddelik havuzun kullanıldığı koşullardaki sınıflama doğruluğu, 800 maddelik havuzun kullanıldığı koşullardakinden yüksek çıkmıştır. Bu beklenmedik sonuç, havuzların maksimum bilgi verdiği yetenek

düzeyinin kesme puanına olan uzaklıklarının farklı olmasından kaynaklanmıştır. Büyüklüğü 600 olan havuzun bilgi fonksiyonunun kesme puanı olan θ =0.13'e daha yakın bir yetenek düzeyinde maksimum değere sahip olması, sınıflama doğruluğunu arttırmıştır. Sonuç olarak, bu iki test yaklaşımında havuz bilgi fonksiyonun şeklinin sınıflama doğruluğu üzerindeki etkisi, havuz büyüklüğünün etkisini gölgeleyebilmektedir. ÇABT ve BLT uygulamalarında ise havuz büyüklüğü arttıkça sınıflama doğruluğu artmaktadır.

6. Sınıflama doğruluğu açısından en iyi test yaklaşımı BST_KP'dir. Daha sonra sırasıyla BBT, BLT, BST_GY ve ÇABT gelmektedir. BST_KP uygulamalarında kesme puanında maksimum bilgi veren maddenin seçilmesi, sınıflama doğruluğunu arttırmıştır. BBT uygulamalarında ise ölçme kesinliğinin yüksek olması, sınıflama doğruluğuna yansımıştır. Ölçme kesinliği açısından daha iyi olmasına rağmen, ÇABT'ın sınıflama doğruluğu BLT'den düşük çıkmıştır. Hedef bilgi fonksiyonları θ = 0.00 noktasında maksimum bilgi verecek şekilde oluşturulan BLT formları, ÇABT ile karşılaştırıldığında kesme puanı (θ = 0.13) etrafında daha fazla bilgi vermektedir. ÇABT uygulamalarındaki kolay ve zor modüllerin sırasıyla θ = -1.00 ve θ = 1.00 yetenek düzeyinde maksimum bilgi verecek şekilde oluşturulması, bu test yaklaşımının sınıflama doğruluğunu olumsuz etkilemiştir.

7. BBT, BST_KP ve BST_GY uygulamalarında test uzunluğu arttıkça bireylere uygulanan ortak madde oranı artmaktadır. Bu durum test güvenliğine zarar vermektedir. BBT ve BST_GY uygulamalarında havuz büyüklüğünün artması, test güvenliği üzerinde kısmen olumlu bir etkiye sahiptir. Havuz büyüklüğü arttıkça bireylere uygulanan ortak madde oranı azalmaktadır. Fakat havuz büyüklüğündeki artışa rağmen maksimum madde kullanım sıklığı değerleri yüksektir. Geçici yetenek düzeyinde maksimum bilgi veren maddenin seçildiği bu iki test yaklaşımında bazı maddeler çok sık kullanılmıştır. Madde kullanım sıklığını kontrol etmek için kullanılan bilgi koşullu tesadüfi yöntemi, yüksek bilgi veren bazı maddelerin aşırı kullanımını kontrol etmede yetersiz kalmıştır. BST_KP uygulamalarında ise madde seçim yönteminden dolayı, havuz büyüklüğü fark etmeksizin tüm bireylere kesme puanı etrafında maksimum bilgi veren sınırlı sayıdaki madde uygulanmıştır. Bilgi koşullu tesadüfi yöntemi, her seferinde kesme puanında maksimum bilgi veren beş madde arasından birinin rastgele seçmesine rağmen bazı maddelerin aşırı kullanılmasını engelleyememiştir. Sonuç olarak, BST uygulamalarında kesme puanı etrafında maksimum bilgi veren maddeyi

seçme yöntemi bilgi koşullu tesadüfi madde sıklığı kontrol yöntemiyle birlikte kullanıldığı durumlarda, havuz büyüklüğünün test güvenliği üzerinde hiçbir etkisi yoktur.

8. ÇABT ve BLT uygulamalarında test uzunluğu ve havuz büyüklüğünün test güvenliği üzerinde doğrudan bir etkisi bulunmamaktadır. Bu iki test yaklaşımında, madde kullanım sıklığı oranı ve test çakışma oranı test uzunluğu ve havuz büyüklüğüne göre değişmemektedir. ÇABT uygulamalarında panel sayısı, BLT uygulamalarında ise form sayısı test güvenliğinin sağlanmasında etkili olmuştur.

9. Test güvenliği açısından en iyi bireyselleştirilmiş test yaklaşımı ÇABT'tır. Bireyler üç panelden birine rastgele atandıkları için ÇABT uygulamalarında maksimum madde kullanım sıklığı 0.35'i geçmemektedir. Ayrıca, test çakışma oranı en düşük olan test yaklaşımı da ÇABT'tır. BBT ve BST_GY uygulamalarında bilgi koşullu tesadüfi madde kullanım sıklığı kontrol yöntemi kullanılmasına rağmen, bazı maddeler sık kullanılmıştır. Fakat bu iki test yaklaşımında, test çakışma oranı BST_KP ve BLT'ye göre düşüktür. Test güvenliği açısından en kötü test yaklaşımı ise BST_KP'dir. Hem maksimum madde kullanım sıklığı hem de test çakışma oranı açısından en kötü sonuçlar BST_KP uygulamalarında elde edilmiştir. Bireyselleştirmenin olmadığı BLT uygulamalarında ise bireylere üç test formundan birisinin rastgele uygulanması, maksimum madde kullanım sıklığını düşürmüştür. Fakat test çakışma oranı yüksektir.

10. BBT, BST_GY ve BST_KP uygulamalarında test uzunluğu arttıkça hiç kullanılmayan madde oranı azalmakta ve havuz daha etkili bir şekilde kullanılmaktadır. Havuz büyüklüğü arttıkça ise havuz kullanımı kötüleşmekte ve hiç kullanılmayan madde oranı artmaktadır. Fakat BST_KP uygulamalarında havuz büyüklüğü fark etmeksizin dar bir yetenek aralığındaki maddeler kullanıldığı için havuz büyüklüğündeki artışa bağlı olarak hiç kullanılmayan madde oranındaki artış BBT ve BST_GY'ye göre daha az olmuştur. ÇABT uygulamalarında, test uzunluğu ve havuz büyüklüğü fark etmeksizin modül ve panellerin oluşturulmasında kullanılan maddelerin hepsi uygulanmıştır. BLT uygulamalarında ise test uzunluğu ve havuz büyüklüğü fark etmeksizin formları oluşturan maddelerin tamamı uygulanmıştır.

11. Havuz kullanımı açısından en iyi bireyselleştirilmiş test yaklaşımı ÇABT'tır. Panellerin oluşturulmasında kullanılan maddelerin tamamı en az bir kez uygulanmıştır. Bu sonuca göre, ÇABT uygulamalarında panel sayısı artırıldığında orijinal havuzdaki

maddelerin çoğundan faydalanılabilir. Havuz kullanımı açısından en kötü bilgisayarlı test yaklaşımı ise BST_KP'dir. Tüm bireylere kesme puanında maksimum bilgi veren maddeler uygulandığı için havuzların çok büyük bir bölümü kullanılmamıştır. Farklı test uzunluğu ve havuz büyüklüğü koşullarında hiç kullanılmayan madde yüzdesi % 80.75 ile % 91.88 arasında değişmektedir. BBT ve BST_GY test yaklaşımlarının havuz kullanımı, BST_KP'ye göre daha iyidir.

Öneriler

Bu bölümde sonuçlara dayalı olarak bazı önerilere yer verilmiştir. Öneriler, uygulayıcılara yönelik öneriler ve araştırmacılara yönelik öneriler olmak üzere iki başlık altında sunulmuştur.

Uygulayıcılara Yönelik Öneriler

1. YÖKDİL'e dayalı olarak yapılan simülasyonların sonuçlarına göre en doğru yetenek kestirimleri BBT uygulamalarında elde edilmektedir. Fakat BBT uygulamalarında bazı maddeler çok sık kullanılmaktadır. Madde kullanım sıklığını kontrol etmek için kullanılan bilgi koşullu tesadüfi yöntemi (randomesque), bazı maddelerin sık kullanılmasını engellemekte yetersiz kalmaktadır. Bazı simülasyon koşullarında bireylerin yarısından fazlasına uygulanan madde ya da maddeler bulunmaktadır. Sık kullanılan bu maddelerin açığa çıkma olasılığı yüksektir. Bu durum test güvenliğini tehlikeye atmaktadır. Bu nedenle karar vericiler YÖKDİL ve YDS gibi dil sınavlarını BBT olarak uygulamaya karar verirse, uygulamadan önce en uygun madde kullanım sıklığı kontrol yöntemi belirlenmelidir. Madde kullanım sıklığını kontrol etmek için maksimum madde kullanım sıklığını sınırlayan Sympson Hetter gibi koşullu yöntemler kullanılabilir. Eğer bilgi koşullu tesadüfi yöntemi kullanılacaksa, madde seçim algoritmasına bazı eklemeler yapılmalıdır. Madde seçim algoritmasında yapılacak değişikliklerle, kullanım sıklığı önceden belirlenen sınır değere ulaşan maddelerin tekrar uygulanması engellenmelidir.

2. Dil becerisinin değerlendirilmesinde bireyleri ileri, orta ve başlangıç gibi yeterlik düzeylerine göre sınıflamak yaygın bir uygulamadır. Bireylerin yeterlik düzeylerini doğru bir şekilde belirleyebilmek için kullanılan test yaklaşımının sınıflama doğruluğunun yüksek olması önemlidir. Araştırma sonuçlarına göre bireyleri en doğru

şekilde sınıflayan test yaklaşımı BST-KP'dir. Fakat bu test yaklaşımında, kesme puanı civarında maksimum bilgi veren maddelerin çok sık kullanılması test güvenliğine zarar verebilir. Ayrıca, tüm bireylere kesme puanı civarında maksimum bilgi veren maddeler uygulandığı için havuzdaki maddelerin çoğunluğu kullanılmamaktadır. Madde yazım sürecinin maliyeti dikkate alındığında bu durum büyük bir israfa yol açacaktır. Tüm bu nedenlerden dolayı bilgisayarlı sınıflama testinin bu madde seçim yöntemiyle YÖKDİL ve YDS gibi dil sınavlarında kullanılması uygun değildir. Fakat yetkili kurumlar bu sınavları sınıflama testi olarak uygulamaya karar verirse BST-KP test yaklaşımı yapılacak bazı değişikliklerle uygulanabilir. Bu test yaklaşımında, geniş bir yetenek aralığı dikkate alınarak oluşturulan bir havuz işlevsel değildir. Bireylere kesme puanında maksimum bilgi veren maddeler uygulandığı için havuzdaki maddeler kesme puanlarına dayalı olarak hazırlanmalıdır. Madde yazımına başlamadan önce yeterlik düzeyi sayısı ve standart kesme puanları belirlenmelidir. Daha sonra belirlenen her bir kesme puanına karşılık gelen yetenek düzeyine uygun yeterince madde yazılarak havuz oluşturulmalıdır. Böylelikle madde havuzunun daha etkili bir şekilde kullanımı sağlanacaktır. Fakat havuzun böyle bir yaklaşımla oluşturulması test güvenliğini sağlamak için yeterli olmayabilir. Bilgi koşullu tesadüfi yöntemi, kesme puanında en çok bilgiyi veren bazı maddelerin sık kullanılmasını engelleyemeyebilir. Bu nedenle madde kullanım sıklığının kontrol edilmesi sürecinde de bazı değişiklikler yapılmalıdır. Bilgi koşullu tesadüfi yöntemi yerine, maksimum madde kullanım sıklığını sınırlayan koşullu yöntemler kullanıldığında test güvenliği açısından daha iyi sonuçlar elde edilebilir. Eğer bilgi koşullu tesadüfi yöntemi kullanılacaksa, madde seçim algoritmasında bazı değişiklikler yapılmalı ve kullanım sıklığı önceden belirlenen maksimum değere ulaşan maddelerin tekrar uygulanmasına izin verilmemelidir.

3. Bu araştırmanın sonuçlarına göre BBT hem ölçme kesinliği hem de sınıflama doğruluğu açısından; BST ise sınıflama doğruluğu açısından ÇABT'a göre daha iyidir. Fakat ÇABT, test güvenliği ve havuz kullanımı açısından bu iki test yaklaşımına göre daha iyidir. Madde yazımının maliyeti, test hazırlayan kurumlar için ekonomik bir yüktür. Bu maliyet dikkate alındığında, havuzdaki maddeleri etkili bir şekilde kullanan ÇABT ekonomik açıdan en avantajlı bireyselleştirilmiş test yaklaşımıdır. YÖKDİL ve YDS gibi dil sınavlarını uygulayan yetkili kuruluşlar, bireyselleştirilmiş test uygulamalarına geçiş yaparken ÇABT'ın ekonomik açıdan daha avantajlı olmasını dikkate alarak karar vermelidirler.

Araştırmacılara Yönelik Öneriler

1. YÖKDİL ortak köklü maddeler içermektedir. Madde takımları (testlet) olarak da adlandırılan ortak köklü maddeler, MTK'nın yerel bağımsızlık varsayımını ihlal etmektedir. Bu çalışmada, yerel bağımsızlığın ihlal edilmesinin sebep olacağı problemleri ortadan kaldırmak için madde takımları çoklu puanlanan madde olarak ele alınmıştır. Madde takımları çoklu puanlanan madde olarak ele alındığında madde havuzunda hem ikili hem de çoklu puanlanan maddeler yer aldığı için MTK modeli olarak genelleştirilmiş kısmi puan modeli (GKPM) kullanılmıştır. Gelecekteki araştırmalarda yerel bağımlılık problemine karşı madde takımları tepki modeli (teslet response model) kullanılabilir. Araştırmacılar, bireyselleştirilmiş test yaklaşımlarını madde takımları tepki modeli bağlamında karşılaştırabilirler.

2. Bu araştırmada kesme puanı belirlemek için standart belirleme çalışması yapılmamıştır. Kesme puanı, yeteneğin normal dağıldığı varsayımına dayalı olarak belirlenmiştir. Yükseköğretim Kurumunun doçentlik başvurusu için belirlediği kesme puanı olan 55, θ yetenek ölçeğine dönüştürülmüştür. Kesme puanı standart belirleme çalışmalarına dayalı olarak belirlendiğinde, bireyselleştirilmiş test yaklaşımlarının sınıflama doğruluğu hakkında daha gerçeğe yakın sonuçlar elde edilebilir. Araştırmacılar, gelecekteki araştırmalarda standart belirleme çalışmalarına dayalı olarak belirledikleri kesme puanlarını kullanabilirler. Ayrıca, bu araştırmada sadece bir kesme puanı kullanılmıştır. Bu kesme puanına göre bireyler başarılı/başarısız şeklinde iki kategoriye sınıflanmaktadır. Fakat dil becerisinin değerlendirilmesinde ileri (advanced), orta-üstü (upper-intermediate), orta (intermediate), temel (elementary) ve başlangıç (beginner) gibi yeterlilik düzeyleri sıklılıkla kullanılmaktadır. Bireyleri bu yeterlik düzeylerine göre sınıflayabilmek için birden fazla kesme puanının kullanılması gerekmektedir. Gelecekteki araştırmalarda birden fazla kesme puanının kullanıldığı koşullarda bireyselleştirilmiş test yaklaşımları sınıflama doğruluğu açısından karşılaştırılabilir.

3. Bu araştırmada, BBT, BST-GY ve BST-KP test yaklaşımlarının madde kullanım sıklığını kontrol etmek için bilgi koşullu tesadüfi yöntemi kullanılmıştır. Bu yöntem, madde seçim yönteminin seçtiği beş maddeden birisi bireylere rastgele olarak uygulamasına rağmen bazı maddelerin sık kullanılmasını engelleyememiştir. BBT ve BST-GY uygulamalarında, yüksek bilgi veren bazı maddeler sık kullanılmıştır. BST-KP

uygulamalarında ise maksimum bilgiyi kesme puanı etrafında veren bazı maddeler sık kullanılmıştır. Sonuç olarak, bilgi koşullu tesadüfi yöntemi maddelerin maksimum kullanım sıklığını sınırlamakta yetersiz kalmış ve BBT ve BST uygulamalarının madde kullanım sıklığını kontrol etmede yeterince etkili olamamıştır. Bu nedenle gelecekteki araştırmalarda BBT ve BST uygulamalarının madde kullanım sıklığı kontrol etmek için farklı yöntemler kullanılabilir. Araştırmacılar, maksimum madde kullanım sıklığını sınırlayan koşullu yöntemleri ya da madde havuzunu madde ayırt edicilik parametresine göre tabakalara ayıran tabakalı yöntemleri kullanabilirler.

4. Bu araştırmada, ÇABT simülasyonları için 1-3-3 panel deseninde 3 panel oluşturulmuştur. Her biri yedi modül içeren bu paneller oluşturulurken yukarıdan aşağı (top-down) ve aşağıdan yukarı (bottom-up) panel oluşturma yöntemlerinin birlikte kullanıldığı karma (mixture) panel oluşturma yöntemi kullanılmıştır. Bireyleri sonraki aşamalara yönlendirmek için maksimum Fisher bilgisi yönlendirme yöntemi kullanılmıştır. Araştırmanın ÇABT ile ilgili sonuçları, araştırmada belirlenen aşama sayısı, panel deseni, panel sayısı, panel oluşturma yöntemi ve yönlendirme yöntemiyle sınırlıdır. Araştırmacılar, gelecekteki araştırmalarda farklı aşama sayısı, panel deseni, panel sayısı, panel oluşturma yöntemi ve yönlendirme yöntemi kullanabilirler.

5. Bu araştırmada, kolay, orta ve zor güçlük düzeyindeki modüller test bilgi fonksiyonları sırasıyla -1.00, 0.00 ve 1.00 yetenek düzeylerinde maksimum değeri verecek şekilde oluşturulmuştur. Kolay ve zor modüllerin maksimum bilgi verdiği yetenek düzeylerinin kesme puanından ($\theta = 0.13$) uzak olması, ÇABT uygulamalarının sınıflama doğruluğunu olumsuz etkilemiştir. Araştırmacılar, gelecekteki araştırmalarda modüllerin hedef bilgi fonksiyonlarını kesme puanına dayalı olarak belirlerse sınıflama doğruluğu açısından daha iyi sonuçlar elde edebilirler.

KAYNAKLAR

Aybek, E. C., & Çıkrıkçı, R.N. (2018). Kendini Değerlendirme Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Test Olarak Uygulanabilirliği. *Türk Psikolojik Danışma ve Rehberlik Dergisi*, *8*(50), 117-141.

Boztunc Ozturk, N., & Dogan, N. (2015). Investigating item exposure control methods in computerized adaptive testing. *Educational Sciences: Theory & Practice*, *15*(1), 85–98. doi: 10.12738/estp.2015.1.2593

Briel, J. B., & Michel, R. (2014). Revisiting the GRE general test. In C. Wendler & B. Bridgeman (Eds.) .*The research foundation for the GRE revised General Test: A compendium of studies.* Princeton, NJ: Educational Testing Service. http://www. ets. org/s/research/pdf/gre_compendium. pdf adresinden alınmıştır.

Bulut, O., & Kan, A. (2012) Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Egitim Arastirmalari- Eurasian Journal of Educational Research*, *49*, 61–80.

Chalhoub–Deville, M., & Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, *19*, 273–299. doi:10.1017/S0267190599190147

Chang, H., & Ying, Z. (1999). a-stratified multistage computerized adaptive testing. *Applied Psychological Measurement, 23*, 211–222.

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.

Cheng, Y., Patton, J. M., & Shao, C. (2015). A-stratified computerized adaptive testing in the presence of calibration error. *Educational and Psychological Measurement*, *75*(2), 260–283. doi: 10.1177/0013164414530719

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory.* Belmont: Wadsworth Group.

Davis, L. L. (2004). Strategies for controlling item exposure in computerized adaptive testing with the generalized partial credit model. *Applied Psychological Measurement*, *28*(3), 165–185. doi: 10.1177/0146621604264133

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. doi: 10.1177/0146621603256804

Davis, L. L., Pastor, D. A., Dodd, B. G., Chiang, C., & Fitzpatrick, S. J. (2003). An examination of exposure control and content balancing restrictions on item selection in CATs using the partial credit model. *Journal of Applied Measurement*, *4*, 24–42.

DeMars, C. (2010). *Item response theory.* New York, NY: Oxford University Press.

Deng, H., Ansley, T., & Chang, H. H. (2010). Stratified and maximum information item selection procedures in computer adaptive testing. *Journal of Educational Measurement*, *47*(2), 202-226.

Dodd, B. G., De Ayala, R. J., & Koch, W. R. (1995). Computerized adaptive testing with polytomous items. *Applied Psychological Measurement*, *19*(1), 5-22.

Doong, S. H. (2009). A knowledge-based approach for item exposure control in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, *34*(4), 530-558. doi: 10.3102/1076998609336667

Eggen, T. J. H. M. (1999). Item selection in adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement*, *23*(3), 249–261.

Eggen, T .J. H. M. (2010). Three-category adaptive classification testing. In W. J. van der Linden & C. A.W. Glas (Eds.), *Elements of adaptive testing* (pp.373-387). New York, NY: Springer.

Eggen, T. J. H. M. (2011). Computerized classification testing with the Rasch model. *Educational Research and Evaluation*, *17*(5), 361–371. doi: 10.1080/13803611.2011.630528

Eggen, T. J. H. M., & Straetmans, G. J. J. M. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement, 60*(5), 713-734.

Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum Associate.

Eroğlu, M. G., & Kelecioğlu, H. (2015). Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Farklı Sonlandırma Kurallarının Ölçme Kesinliği Ve Test Uzunluğu Açısından Karşılaştırılması. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, *28*(1), 31-52.

Flaugher, R. (2000). Item pools. In H. Wainer (Eds.), *Computerized adaptive testing: A primer* (2nd ed.) (pp. 37–59). Mahwah, NH: Lawrence Erlbaum Associates.

Georgiadou, E. G., Triantafillou, E., & Economides, A. A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *The Journal of Technology, Learning and Assessment*, *5*(8), 3-37. http://www.jtla.org sitesinden alınmıştır.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, *21*(4), 347-360.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

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.

Han, K. T. (2013). Item pocket method to allow response review and change in computerized adaptive testing. *Applied Psychological Measurement*, *37*(4), 259–275. doi: 10.1177/0146621612473638

Harwell, M., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement*, *20*(2), 101-125.

He, W. (2010). *Optimal item pool design for a highly constrained computerized adaptive test*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3435100).

He, W., Diao, Q., & Hauser, C. (2014). A comparison of four item-selection methods for severely constrained CATs. *Educational and Psychological Measurement*, *74*(4), 677–696. doi: 10.1177/0013164413517503

Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice, 26*(2), 44-52.

Ho, T.-H. & Dodd, B. G. (2012). Item selection and ability estimation procedures for a mixed-format adaptive test. *Applied Measurement in Education, 25*(4), 305-326, doi: 10.1080/08957347.2012.714686

Jodoin, M. G. (2003). *Psychometric properties of several computer-based test designs with ideal and constrained item pool*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3096288)

Jodoin, M. G., Zenisky, A. L., & 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. doi: 10.1207/s15324818ame1903

Kalender, İ. (2012).Computerized adaptive testing for student selection to higher education. *Journal of Higher Education, 2*(1), 13-19.

Kang, H.-A., & Chang, H.-H. (2016). Parameter drift detection in multidimensional computerized adaptive testing based on informational distance/divergence measures. *Applied Psychological Measurement*, *40*(7), 534–550.

Keng, L. (2008). *A Comparison of the performance of testlet-based computer adaptive tests and multistage tests.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3315089).

Kezer, F. & Koç, N. (2014). Bilgisayar Ortamında Bireye Uyarlanmış Test Stratejilerinin Karşılaştırılması [A comparison of computerized adaptive testing strategies]. *Eğitim Bilimleri Araştırmaları Dergisi - Journal of Educational Sciences Research, 4*(1), 145-174.

Kim, J. (2010). *A comparison of computer-based classification testing approaches using mixed- format tests with the generalized partial credit model*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3429006).

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

Kingsbury, G. G., & Weiss, D. J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D. J. Weiss (Eds.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 257-283). New York: Academic Press.

Kingsbury, G. G., & Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education, 2*(4), 359-375.

Konis, K. (2014). *IpSolveAPI: R Interface for Ip_solve version 5.5.2.0. R package version 5.5.2.0-14*. https://cran.r-project.org/ web/packages/lpSolveAPI/index.html adresinden alınmıştır.

Leroux, A. J., Lopez, M., Hembry, I., & Dodd, B. G. (2013). A comparison of exposure control procedures in CATs using the 3PL model. *Educational and Psychological Measurement, 73*(5), 857-874. doi: 10.1177/001316441348680

Lin, C.-J. (2011). Item selection criteria with practical constraints for computerized classification testing. *Educational and Psychological Measurement, 71*(1), 20–36. doi: 10.1177/0013164410387336

Luecht, R. M. (2003, April). *Exposure control using adaptive multi-stage item bundles*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Luecht, R. M., Brumfield, T., & Breithaupt, K. (2006). A testlet assembly design for adaptive multistage tests. *Applied Measurement in Education, 19*(3), 189-202. doi:10.1207/s15324818 ame1903_2

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 W. J. van der Linden & C. A. W. Glas (Eds.), *Computer-adaptive testing:Theory and practise* (pp. 117-128). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Luecht, R. M., & Sireci, S. G. (2011). *A review of models for computer-based testing* (Research Report RR-2011-12). New York, NY: The College Board.

Luo, X., & Kim, D. (2018). A top-down approach to designing the computerized adaptive multistage test. *Journal of Educational Measurement, 55*(2), 243-263.

Magis, D., & Raîche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software*, *48*(8), 1–31.

Magis, D., Yan, D., & von Davier, A. (2018). *mstR: Procedures to generate patterns under multistage testing.* R package version 1.2. https://cran.r-project.org/web/packages/mstR/index.html adresinden alınmıştır.

Melican, G. J., Breithaupt, K., & Zhang, Y. (2010). Designing and implementing an multistage adaptive test: The Uniform CPA Exam. In W. J. van der Linden & C.A.W. Glas (Eds.) *Elements of adaptive testing*, (pp. 167-189). New York, NY: Springer.

Murphy, D. L., Dodd, B. G., & Vaughn, B. K. (2010). A comparison of item selection techniques for testlets. *Applied Psychological Measurement*, *34*(6), 424–437. doi: 10.1177/0146621609349804

Nogami, Y., & Hayashi, N. (2010). A Japanese adaptive test of English as a foreign language: Developmental and operational aspects. In W. J. van der Linden & C.A.W. Glas (Eds.) *Elements of adaptive testing*, (pp. 191-211). New York, NY: Springer.

Nydick, S. (2014). *catIrt: An R package for simulating IRT-based computerized adaptive tests.* R package version 0.5-0. https://cran.r-project.org/web/packages/catIrt/index.html adresinden alınmıştır.

Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, *70*, 351-356.

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York, NY: Springer

Pastor, D.A., Dodd, B.G., & Chang, H. H. (2002). A comparison of item selection techniques and exposure control mechanisms in CATs using the generalized partial credit model. *Applied Psychological Measurement*, *26*(2), 147-163.

Patsula, L. N. (1999). *A comparison of computerized-adaptive testing and multi-stage testing.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9950199).

Reckase, M. D. (1981). *Final report: Procedures for criterion referenced tailored testing*. Columbia: University of Missouri, Educational Psychology Department.

Reckase, M. D. (1983). A procedure for decision making using tailored testing. In D. J. Weiss (Eds.), *New horizons in testing: Latent trait theory and computerized adaptive testing* (pp. 237-254). New York, NY: Academic Press.

Revuelta, J. & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement, 35*(4), 311-327.

Samejima, F. (1969). Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17

Sari, H. I. & Huggins-Manley, A. C. (2017). Examining content control in adaptive tests: Computerized adaptive testing vs. computerized adaptive multistage testing. Educational Sciences: Theory & Practice, *17*, 1759-1781. doi:10.12738/estp.2017.5.0484

Segall, D. O. (2005). Computerized adaptive testing. *Encyclopedia of Social Measurement, 1*, 429–438.

Shin, C. D., Chien, Y., Way, W. D., & Swanson, L. (2009). *Weighted penalty model for content balancing in CATs*. San Antonio, TX: Pearson. http:// images.pearsonassessments.com/images/tmrs/tmrs_rg/WeightedPenaltyModel.pdf ?WT.mc_ id=TMRS_Weighted_Penalty_Model_for_Content_Balancing adresinden alınmıştır.

Sireci, S.G., Thissen, D., & Wainer, H. (1991).On the reliability of testlet-based tests. *Journal of Educational Measurement, 28*(3), 237–247.

Spray, J. A., & Reckase, M. D. (1994, April). *The selection of test items for decision making with a computerized adaptive test*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Spray, J. A., & Reckase, M. D. (1996). Comparison of SPRT and sequential Bayes procedures for classifying examinees into two categories using a computerized test. *Journal of Educational and Behavioral Statistics, 21*(4), 405–414.

Stocking, M. L. (1994). *Three practical issues for modern adaptive testing item pools* (ETS Research Report 94-5). Princeton, NJ: Educational Testing Service.

Stocking, M. L., & Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement, 17*, 277-292.

Stone, E., & Davey, T. (2011). *Computer-adaptive testing for students with disabilities: A Review of the literature* (ETS Research Report 11-32). New Jersey , NJ: Educational Testing Service.

Thissen, D., Steinberg, L., & Mooney, J. (1989). Trace lines for testlets: A use of multiple- category-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*(5), 778-793. doi: 10.1177/0013164408324460

Thompson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research, and Evaluation, 16*(1), http://pareonline.net/getvn.asp?v=16&n=1 adresinden alınmıştır.

van der Linden, W. J. (1998). Bayesian item vselection criteria for adaptive testing. *Psychometrika, 63*(2), 201-216.

van der Linden, W. J., & Pashley, P. J. (2010). Item selection and ability estimation in adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 3-30). New York, NY: Springer.

van der Linden, W. J., & Reese, L. M. (1998). A model for optimal constrained adaptive testing. *Applied Psychological Measurement*, 22(3), 259-270.

van Groen, M. M., Eggen, T. J. H. M., & Veldkamp, B. P. (2014). Item selection methods based on multiple objective approaches for classifying respondents into multiple levels. *Applied Psychological Measurement*, 38(3) 187–200. doi: 10.1177/0146621613509723

van Groen, M. M., Eggen, T. J. H. M., & Veldkamp, B. P. (2016). Multidimensional computerized adaptive testing for classifying examinees with within-dimensionality. *Applied Psychological Measurement*, 40(6) 387–404. doi: 10.1177/0146621616648931

van Rijn, P. W., Eggen, T. J. H. M., Hemker, B. T., & Sanders, P. F. (2002). Evaluation of selection procedures for computerized adaptive testing with polytomous items. *Applied Psychological Measurement, 26*(4), 393–411. doi: 10.1177/014662102237796

Verschoor, A. J., & Straetmans, G. J. J. M. (2010). MATHCAT: A flexible testing system in mathematics education for adults. In W. J. van der Linden & C.A.W. Glas (Eds.), *Elements of adaptive testing* (pp. 137–149). New York, NY: Springer.

Vos, H. J. (2000). A Bayesian procedure in the context of sequential mastery testing. *Psicologica, 21*, 191-211.

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications.* New York, NY: Cambridge University Press.

Wainer, H., & Kiely, G. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24*, 185-202.

Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement, 37*(3), 203–220.

Wang, K. (2017). *A fair comparison of the performance of computerized adaptive testing and multisatege adaptive testing*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 10273809).

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*, 427–450.

Weiss, D.J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement, 6*(4), 473-492.

Weissman, A. (2014). IRT-based multistage testing. In D.Yan, A.A. von Davier & C.Lewis (Eds.),Computerized multistage testing: Theory and applications (pp.153-168). New York, NY: CRC Press.

Widiatma, H. S. (2004). *A simulation and evaluation of computerized adaptive testing designs for the verbal battery of the cognitive abilities test (Cogat).* Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3129356).

Xing, D., & Hambleton, R. K. (2004). Impacts of test design, item quality, and item bank size on the psychometric properties of computer-based credentialing examinations. *Educational and Psychological Measurement, 64*(1), 5-21. doi: 10.1177/0013164403258393

Yao, L. (2013). Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. *Applied Psychological Measurement*, *37*(1), 3–23. doi: 10.1177/0146621612455687

Yao, L. (2014). Multidimensional CAT item selection methods for domain scores and composite scores with item exposure control and content constraints. *Journal of Educational Measurement, 51*(1), 18–38.

Zhang, J., & Li, J. (2016). Zhang, J., & Li, J. (2016). Monitoring items in real time to enhance CAT. *Journal of Educational Measurement*, *53*(2), 131–151.

Zheng, Y., & Chang, H. H. (2015). On-the-fly assembled multistage adaptive testing. *Applied Psychological Measurement*, *39*(2*)*, 104–118.doi: 10.1177/0146621614544519

Zheng, Y., Nozawa, Y., Gao, X., & Chang, H. H. (2012). *Multistage adaptive testing for a large-scale classification test: The designs, automated heuristic assembly, and comparison with other testing modes* (Research Report 2012–6). Iowa City, IA: ACT Inc.

EKLER

Ek 1. Etik Kurul Muafiyet Belgesi

ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ALT ETİK KURULU
KARAR ÖRNEĞİ

Karar Tarihi : **22/04/2019**

Toplantı Sayısı : **05**

Karar Sayısı : **165**

165- Üniversitemiz Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı doktora öğrencisi **Ercan Çoban**'ın "Bilgisayarda Uygulanan Farklı Bireyselleştirilmiş Test Yaklaşımlarının Yükseköğretim Kurumları Dil Sınavında Uygulanabilirliğinin İncelenmesi" başlıklı tezi ile ilgili 28/03/2019 tarihli "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

Üniversitemiz Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı doktora öğrencisi **Ercan Çoban**'ın "Bilgisayarda Uygulanan Farklı Bireyselleştirilmiş Test Yaklaşımlarının Yükseköğretim Kurumları Dil Sınavında Uygulanabilirliğinin İncelenmesi" başlıklı tezinin, insan üzerinde yapılan bir çalışma olmaması nedeniyle etik kurul onayına ihtiyaç duymadığına oy birliği ile karar verildi.

ASLININ AYNIDIR
22/04/2019

Prof. Dr. Muharrem ÖZEN
Ankara Üniversitesi
Etik Kurulu Başkanı

BENZERLİK BİLDİRİMİ

"Bilgisayar Temelli Bireyselleştirilmiş Test Yaklaşımlarının Türkiye'deki Merkezi Dil Sınavlarında Uygulanabilirliğinin Araştırılması" başlıklı tezimin ana bölümü (ön bölüm, kaynaklar ve ekler hariç) Turnitin İntihali Engelleme Programı aracılığıyla incelenmiş ve ilgili rapor danışmanım tarafından da kontrol edilmiştir. Kontrol sırasında (1) "Kapak, Önsöz, Özet, İçindekiler, Kısaltmalar/Simgeler Sayfası, Tablolar Dizini, Şekiller Dizini, Dipnotlar, Kaynaklar, Ekler, Özgeçmiş" (2) "Tırnak içinde ve blok olarak verilen (doğrudan) alıntılar" (3) "Yedi sözcüğe kadar olan benzerlik/örtüşme içeren metin kısımları" (4) "Öğrencinin lisansüstü tezi ile ilgili yapmış olduğu yayınlar ile yararlandığı mevzuat metinleri" dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıda sunulmuştur.

Rapor Tarihi	:19 Ağustos 2020
Gönderim Numarası	: 1371402292
Sayfa Sayısı	:72
Sözcük Sayısı	: 18194
Karakter Sayısı	: 126630
Benzerlik Oranı	: %2
Savunma Tarihi	: 28 Temmuz 2020

Yukarıda belirtilen sonuçları gösteren Turnitin İntihali Engelleme Programı'na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu bildirimin ekinde Enstitüye teslim ettiğimi, tezimin **%10**'dan fazla benzerlik oranı içerdiğinin ve tek bir kaynakla eşleşme oranının **%2**'yi geçtiğinin belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı Soyadı: Ercan ÇOBAN

Tarih: 19.08.2020

İmza: E. Çoban

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı : Ercan ÇOBAN

E-Posta Adresi: coban.ercan@gmail.com

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	Matematik Öğretmenliği	Boğaziçi Üniversitesi	2006-2011
Yüksek Lisans	Ölçme ve Değerlendirme	Ankara Üniversitesi	2013-2015
Doktora	Ölçme ve Değerlendirme	Ankara Üniversitesi	2015-2020

Yabancı Dil:

Almanca (Başlangıç Düzeyi)
İngilizce (İleri Düzey)