

Kendini Değerlendirme Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Test Olarak Uygulanabilirliği

Applicability of the Self Assessment Inventory as a Computerized Adaptive Test

Eren Can Aybek^{ORCID}, R. Nükhet Çıkrıkçı^{ORCID}

Öz. Bu çalışmada, Kendini Değerlendirme Envanteri –KDE–'nin Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) olarak uygulanabilirliği araştırılmıştır. Bu amaçla simülatif ve canlı BOBUT uygulamaları, 1144 lise öğrencisinden toplanan verilere dayalı olarak yürütülmüştür. Simülasyonlarda, farklı test sonlandırma kurallarına ve çoklu puanlanan maddelere uygun farklı madde tepki kuramı modelleri ve madde seçim yöntemlerine göre karşılaştırmalar yapılmıştır. Bu karşılaştırmalara göre; madde tepki modeli olarak Genelleştirilmiş Kısmi Puan Modeli, madde seçim yöntemi olarak Fisher En Yüksek Bilgi, durdurma kuralı olarak .40 ölçme standart hatasının KDE için en uygun BOBUT uygulaması ortaya koyduğuna karar verilmiştir. Bu uygulamayla, kağıt-kalem formu 230 madde olan KDE, simülasyonda 113 madde ile sonlanmış ve kağıt-kalem formundan alınan puanlar ile simülasyonla kestirilen özellik düzeyleri arasında .90 ile .96 arasında korelasyon katsayıları elde edilmiştir. Bunun yanında, 25 öğrencinin katıldığı canlı BOBUT uygulaması gerçekleştirilmiştir. Sonuçlar, KDE'nin BOBUT olarak uygulanabilir olduğunu ve KDE'nin bu uygulama yöntemiyle öğrenci rehberlik hizmetlerinde daha etkili kullanılabileceğini göstermektedir.

Anahtar Kelimeler. Bilgisayar Ortamında Bireye Uyarlanmış Test, Madde Tepki Kuramı, Mesleki Rehberlik

Abstract. In this study, the applicability of Self-Assessment Inventory (SAI) as a Computerized Adaptive Test (CAT) form has been investigated. According to this purpose, simulated and live CAT applications were conducted based on data collected from 1144 high school students. In simulations, different item response theory models and item selection methods are applied to different test termination rules. According to these comparisons; Generalized Partial Credit Model as the item response model, Maximum Fisher Information as the item selection method, and .40 standard error of measurement as stopping rule were determined to provide the most appropriate CAT application for SAI. With this application, SAI with 230 items of paper-pencil ended with 113 items in simulation and correlation coefficients varies .90 to .96 between the paper-pencil form scores and the simulation estimates. In addition, a live CAT application conducted with participation of 25 students. The students' SAI profiles they obtained from the CAT application were compared with paper-pencil form and found that the same students had similar profiles from these two applications. The results show that SAI is applicable as a CAT form.

Keywords. Computerized Adaptive Test, Item Response Theory, Vocational Counseling

Eren Can Aybek

Pamukkale Üniversitesi, Eğitim Fakültesi, Denizli, Türkiye
e-mail: erencan@aybek.net

R. Nükhet Çıkrıkçı

İstanbul Aydın Üniversitesi, Ölçme ve Değerlendirme Uygulama ve Araştırma Merkezi, İstanbul, Türkiye
e-mail: rnukhet@yahoo.com

Geliş / Received: 7 Eylül/September 2017

Düzeltilme / Revision: 22 Mart/March 2018

Kabul / Accepted: 29 Mart/March 2018



Türk Psikolojik Danışma ve Rehberlik Derneği
Turkish Psychological Counseling and Guidance Association

Eğitimde ve psikolojide ölçme işlemleri; öğrenci/personel seçme, öğretimi değerlendirme, eğitim politikalarına yön verme, bireyi tanıma ve yönlendirme hizmetleri gibi çeşitli amaçlarla yapılmaktadır. Eğitim kademeleri arası geçişlerde maksimum performans testlerinin yanı sıra gelecekteki kariyer planlarının şekillendirilmesinde, bireyi tanıma hizmetlerinde, ilgi ve kişilik envanterlerinin uygulanması da önemlidir. Bu türden bireyi tanıma hizmetleri mesleki rehberlik kavramı altında tanımlanmaktadır.

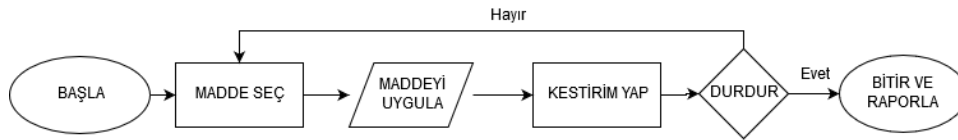
Kuzgun (2004), meslek seçiminin gelişmesinde yetenek, ilgi, mesleki değer ve ihtiyaçlar gibi psikolojik özelliklerin etkili olduğunu ifade etmiştir. İlgilerin ölçülmesi ile ilgili ilk çalışmalar 1914 yılına dayanmakta ve günümüzde, mesleki rehberlik çalışmaları kapsamında Strong İlgi Envanteri (Strong Interest Inventory), Holland'ın Kendini İnceleme Envanteri (Self-Directed Search) ve Türkiye'de; Kuder İlgi Alanları Tercih Envanteri, Kendini Değerlendirme Envanteri (KDE) ve Akademik Benlik Kavramı Ölçeği (ABKÖ) gibi araçlar sıklıkla kullanılmaktadır.

Kuzgun tarafından 1989 yılında geliştirilen Kendini Değerlendirme Envanteri (KDE), lise öğrencilerine meslek seçiminde yol göstermek üzere, öğrencilerin kendilerini daha iyi tanımlarında yardımcı olmak amacıyla algılanan yetenek, ilgiler ve mesleki değerler temel boyutlarındaki özelliklerini ölçmek üzere geliştirilmiştir. KDE'nin geliştirilmesinde Holland'ın kişilik kuramı dikkate alınmıştır (Kuzgun, 2004). Milli Eğitim Bakanlığı Özel Eğitim ve Rehberlik Hizmetleri Genel Müdürlüğü de rehberlik ve psikolojik danışma hizmetleri amacıyla kullanılan ölçme araçları arasında Kendini Değerlendirme Envanteri'ne yer vermiştir (MEB, 2015). Dolayısıyla KDE, bireyi tanıma amacıyla okullarda halen kullanılan bir ölçme aracı özelliği taşımaktadır.

Duyuşsal bir özellik olarak ilgilerin ölçülmesinde kullanılan ölçekler/envanterler, klasik olarak kağıt-kalem formunda uygulandığı gibi, bilgisayar ortamında da uygulanmaktadır. Mesleki ilgi envanterlerinin bilgisayar ortamında uygulanmasıyla birlikte bireyler sonuçları anında görebilmekte ve internet aracılığı ile kendilerine uygun meslekler hakkında bilgiye hızlı bir şekilde ulaşabilmektedir (Betz ve Turner, 2011). Türkiye'de geliştirilen ilk Bilgisayar Destekli Meslek Rehberliği (BİLDEMER) yazılımının çalışmalarına 1991 yılında başlanmıştır (Kuzgun, 2004). Bilgisayar ortamında uygulanan bu uygulama, kağıt kalem uygulamasının bilgisayar ortamına aktarılmış bir uygulaması olup, bireye uyarlanmış (adaptive) özellikte değildir. Milli Eğitim

Bakanlığı Özel Eğitim Rehberlik ve Danışma Hizmetleri Genel Müdürlüğü tarafından geliştirilen Mesleki Bilgi Sistemi (MBS) ise, 13 yaş ve üstü genç ve yetişkinler için hazırlanmıştır (Kaya, 2011). Web tabanlı olarak uygulanan bu sistem kağıt kalem testi uygulamalarının bilgisayar ortamında uygulanmış halidir ve bireye uyarlanma amacı taşımamaktadır. Benzer şekilde KDE'nin de bilgisayar ortamında uygulanan bir formu bulunmaktadır. Bu formda birey yanıtlarını yazılıma girmekte ve bireyin profili yazılım tarafından otomatik olarak oluşturulmaktadır. KDE'nin bilgisayar ortamında uygulanan bu formu da doğrusal bir uygulama olup bireylerden tüm maddeleri yanıtlamaları beklenmektedir (Kuzgun, 2014).

Günümüzde bilgisayar ve yazılım teknolojisinin ulaştığı düzey, eğitim ve psikolojide kullanılan testlerin geçerlik ve güvenilirliklerinden ödün vermeden daha kullanışlı uygulanma olanağı tanıyan bilgisayar ortamında bireye uyarlanmış testlerin (BOBUT) (Computerized Adaptive Testing-CAT) giderek yaygınlaşmasına yol açmıştır. Bilgisayar Ortamında Bireye Uyarlanmış Testlerde (BOBUT) bireyin hangi maddeyi yanıtlayacağı onun ölçülen yetenek / özellik (θ) düzeyine göre belirlenmektedir. Bireye uyarlanmış testlerin ardında yatan temel mantık, testi alan bireye çok zor ya da çok kolay gelebilecek maddeleri uygulamadan, bireyin yeteneğine/özelliğine en uygun maddelerin seçilmesidir (Davey, 2010; Weiss ve Kingsbury, 1984). Bunun için bireyin bir maddeye verdiği yanıtı göre, madde havuzundan yeni bir madde seçilmesi gerekmektedir. Bu işlemin ise kağıt-kalem uygulaması ile kullanışlı olmayacağı açıktır. Bilgisayar ile yapılan uygulamalar düşünüldüğünde ise bireyin verdiği yanıtlara göre yetenek / özellik düzeyinin kestirilmesi ve bu düzeye göre yeni bir maddenin seçimini sağlayan bir sürecin işlenmesi gerekmektedir (Linden ve Glas, 2002). Bu bağlamda, test kuramlarından biri olan Madde Tepki Kuramı (MTK)'nda madde parametrelerinin grubun testle ölçülen özelliğine ait düzeyden; bireye ait yetenek kestiriminin de bireyin aldığı madde örnekleminde bağımsız olması, BOBUT uygulamalarının MTK temelinde geliştirilmesine neden olmaktadır. Bir BOBUT uygulamasının işleyişi temel olarak Şekil 1'de görüldüğü gibidir.



Şekil 1. BOBUT Akış Şeması (Davey, 2011).

Tipik bir BOBUT uygulamasında madde havuzunun, madde havuzundan maddelerin nasıl seçileceğinin, puanlamanın nasıl yapılacağı ve uygulamanın hangi durumlarda durdurulacağı belirlenmesi gerekmektedir. BOBUT uygulamasında geniş bir madde havuzu kullanılmalı, madde havuzunda her yetenek/özellik düzeyine uygun yeter sayıda madde yer almalıdır (Wainer ve Dorans, 2000). İlk maddenin nasıl seçileceği, bireyin maddeye verdiği yanıtı göre yetenek/özellik kestiriminin nasıl yapılacağı ve buna göre havuzdan diğer maddelerin hangi kurala göre seçileceği birçok araştırmacının (Babcock ve Weiss, 2012; Cheng, 2008; Choi ve Swartz, 2009; Downing ve Haladyna, 2006; Linacre, 2000) konusu olmaktadır.

Geçmiş araştırmalar incelendiğinde, BOBUT uygulamalarının ya da simülasyonlarının kağıt-kalem formlarına göre çok daha az sayıda madde ile durdurma kuralını sağladıkları, yaklaşık %45-%50 daha az maddeyle uygulamanın sonlandırıldığı bulunmuştur (Betz ve Turner, 2011; Chien, Lai, Lu, Wang, Chen, Wang ve Su, 2011; Chien, Wang, Wang ve Lin, 2009; Choi, Grady ve Dodd, 2011; Elhan, Öztuna, Kutlay, Küçükdeveci ve Tennant, 2008; Gardner ve diğerleri, 2004; Gibbons ve diğerleri, 2008; Gibbons ve diğerleri, 2012; Haley, Coster, Andres, Kosinski ve Ni, 2004; Haley ve diğerleri, 2009; Hart, Mioduski ve Stratford, 2004; Hart, Cook, Mioduski, Teal ve Crane, 2006; Hol, Vorst ve Mellenbergh, 2007; Kocalevent ve diğerleri, 2009; Oswald, Shaw ve Farmer, 2015; Özbaşı ve Demirtaşlı, 2015; Scullard, 2007; Simms, Goldberg, Roberts, Watson, Welte ve Rotterman, 2011; Smits, Cuijpers ve van Straten, 2011; Walter, Becker, Bjorner, Fliege, Klapp ve Rose, 2007).

Araştırmada yer alan Kendini Değerlendirme Envanteri halen okullarda mesleki rehberlik hizmetlerinde sık kullanılan, ayrıca eğitim fakültelerinde, psikolojik danışman adaylarına tanıtılmaya devam eden bir araçtır. Bu envantere 230 madde yer almakta ve maddelerin yanıtlanması uzun zaman alabilmektedir. Ayrıca uygulama sonrasında envanterin el kitabından maddelerin hangi alt ölçeğe ait olduğunun belirlenmesi, buna göre ham puanların hesaplanması ve daha sonra norm tablosundan yararlanarak puanların yüzdelik puanlara dönüştürülerek profil kağıdına işaretlenmesi zaman alıcı olmaktadır.

Türkiye'de mesleki rehberlik alanında kullanılan ve duyuşsal özellikleri ölçmek üzere geliştirilmiş bilgisayar ortamında uygulanan ölçme araçları olmakla birlikte, ilgileri ölçme işini hem bilgisayarlı hem de bireye uyarlanmış yöntemle

gerçekleştiren bir ilgi ölçeği bulunmamaktadır. KDE'nin yerli bir araç olarak BOBUT olarak uygulanabilirliğini araştırmak ve bu yolla mesleki rehberlik uygulamalarında öğrenci tanıma hizmetlerini kolaylaştıracak yeni bir yol açmak, bu çalışmayı anlamlı ve önemli kılmaktadır.

Bu çalışmada aşağıdaki araştırma sorularına yanıt aranmıştır:

1. Ölçek uygulamasının sonlandırılmasında farklı ölçmenin standart hata değerleri (.30, .40 ve .50) alındığında:
 - a. KDE'nin orijinal formundan ve BOBUT simülasyonundan kestirilen özellik düzeyleri arasındaki ilişkinin miktarı ve yönü nedir?
 - b. KDE BOBUT simülasyonu ortalama kaç madde ile sonlanmaktadır?
 - c. Farklı madde seçim yöntemlerine göre KDE BOBUT simülasyonu ortalama kaç madde ile sonlanmaktadır?
2. Simülasyon sonuçlarına göre en uygun çıkan MTK modeli ve ölçek durdurma kuralına göre, KDE canlı BOBUT olarak uygulandığında:
 - a. BOBUT uygulaması ortalama kaç madde ile sonlanmıştır?
 - b. BOBUT uygulaması ortalama ne kadar sürede sonlanmıştır?
 - c. Öğrencilerin KDE'nin kâğıt-kalem formundan aldıkları puanlar ile BOBUT formundan kestirilen özellik düzeyleri arasındaki ilişkinin miktarı ve yönü nedir?

YÖNTEM

Araştırma Grubu

Araştırmaya iki ayrı çalışma grubu dahil edilmiştir. Birinci çalışma grubunu, 2013 – 2014, 2014 – 2015 öğretim yıllarında Eskişehir ve Bilecik illerinde, 9, 10 ve 11. sınıfta öğrenim görmekte olan 1367 öğrenci oluşturmaktadır. Bu grupta bazı ölçekler eksik doldurulduğu için net 1144 öğrencinin verisiyle çalışılmıştır. Bu öğrencilerden 653'ü kız, 490'ı erkek ve 1 (cinsiyet bilgisi belirtmeyen); 704'ü 9. sınıf, 267'si 10. sınıf ve 173'ü 11. sınıfa devam etmektedir. Bu verilerden post-hoc simülasyon çalışmaları için yararlanılmıştır.

İkinci çalışma grubu ile hem kâğıt-kalem hem de geliştirilen canlı BOBUT uygulaması yürütülmüştür. Birinci çalışma grubundaki öğrenciler, ikinci çalışma grubuna dahil edilmemiştir. Araştırma süreci hakkında öğrencilere duyuru yapılmış ve araştırmaya katılmak için gönüllü olan 35 öğrenci ile çalışılmıştır.

Ancak her iki uygulamaya katılmayan üç öğrenci ve uygulamalardan yalnızca birisine katılan yedi öğrenciden elde edilen veriler çalışmaya dahil edilmemiştir. Buna göre, canlı BOBUT uygulaması 2015 – 2016 öğretim yılında Eskişehir ilinde bir sosyal bilimler lisesinde öğrenim görmekte olan 22 kız, 3 erkek toplam 25 öğrenci ile yürütülmüştür. Öğrencilerden 12'si 9. sınıf, 10'u 10. sınıf ve 3'ü 11. sınıfa devam etmektedir.

Veri Toplama Araçları

Kendini Değerlendirme Envanteri Kağıt-Kalem Formu (KDE-KK).

Kuzgun (2014) tarafından ilk olarak 1989 yılında geliştirilen bu envanter, bireyin algıladığı yeteneklerini, ilgilerini ve mesleki değerlerini ölçmeyi amaçlamaktadır. Envanterde üç boyut altında toplanmış 23 alt ölçek ve toplam 230 madde yer almaktadır. Her alt ölçek 10 maddeden oluşmaktadır. Envanter, yetenekler, ilgiler ve değerler boyutlarından oluşmaktadır.

KDE, yetenekler boyutunda Sözel Yetenek, Sayısal Yetenek ve Şekil-Uzay İlişkileri Yeteneği olmak üzere üç alt ölçek altında öğrencinin kendi algısına göre yeteneklerini ölçmektedir (Kuzgun, 2014). Maddeler dört dereceli olarak puanlanmaktadır. A, hiçbir zaman ya da hiç önemli değil (1 puan); B, ara sıra ya da pek önemli değil (2 puan); C, sıklıkla ya da önemli (3 puan) ve D, her zaman ya da çok önemli (4 puan) anlamına gelmektedir. Her bir alt ölçekte 10 madde yer almasından dolayı, bir alt ölçekten alınabilecek en küçük ham puan 10 ve en büyük ham puan 40'tır. Elde edilen ham puanlar, envanterin el kitabında yer alan norm tablosuna göre yüzdellik standart puanlara dönüştürülmektedir. Norm tablosu, cinsiyete göre ve cinsiyet farkı gözetmeksizin ayrı ayrı hazırlanmıştır.

Envanter maksimum performansı değil, tipik tepkileri ölçmeyi amaçlamaktadır. Dolayısıyla maddeler için doğru ya da yanlış yanıtlardan bahsetmek mümkün değildir. Bunun yerine, birey kendi durumuna en uygun olan tepki seçeneğini işaretlemekte, yani kendini rapor etmektedir. Envanterden aldığı puan, öğrenci için geçme-kalma kararında kullanılmamakta, yalnızca bireyi tanıma amacı taşımaktadır. Bu nedenle Kendini Değerlendirme Envanteri'nin geçerlilik ve güvenilirliği bireyin verdiği yanıtların samimi olması ve kişinin kendisini ne derece iyi tanıdığı ile ilişkilidir. Bu araştırma kapsamında toplanan veriler ile KDE'nin alt ölçeklerine ait iç tutarlılık katsayıları hesaplanmış ve Değerler ölçeğindeki işbirliği ile değişiklik alt ölçekleri dışındaki tüm alt ölçeklerin .71

ile .88 arasında değişen iç tutarlılığa sahip olduğu bulunmuştur. İşbirliği ve değişiklik alt ölçeklerine ait iç tutarlılık katsayıları ise .60'tan büyüktür. Schmitt (1996), bir ölçme aracına ait tekboyutluluk, kapsam geçerliliği gibi özelliklerin sağlandığı durumlarda .70'in altında kalan düşük iç tutarlılık katsayılarının ölçme sonuçlarının kullanılmasına bir engel oluşturmayacağını ifade etmiştir.

Kendini Değerlendirme Envanteri Bilgisayar Ortamında Bireye Uyarlanmış Test Formu (KDE-BOBUT). Post-hoc simülasyonlar sonucunda karar verilen parametrelere göre Concerto (Scalise ve Allen, 2015) platformunda uygulanan KDE-BOBUT geliştirilmiştir. KDE-BOBUT'un geliştirilmesinde catR (Magis, Raiche ve Barrada, 2016) paketi ve R betikleri, HTML ve CSS kodları ve MySQL veritabanından yararlanılmıştır. Uygulama ile ilgili daha detaylı bilgiler Canlı BOBUT uygulaması ile verilerin toplanması başlığında verilmiştir.

Verilerin Toplanması

Post-hoc Simülasyonlar için Verilerin Toplanması. Veriler, Milli Eğitim Bakanlığı Eğitim Teknolojileri Genel Müdürlüğü'nden ve okul müdürlerinden alınan izinlere göre beş lisede, psikolojik danışmanların gözetiminde 2013-2014 bahar, 2014-2015 güz ve bahar yarıyıllarında 9, 10 ve 11. sınıf öğrencilerinden toplanmıştır. her öğrenciye bir adet KDE ve bir adet cevap kâğıdı verilerek, sınıf ortamında uygulanmış ve her sınıftaki uygulama yaklaşık 45-50 dakika sürmüştür. Uygulama sırasında öğrencilerin soruları psikolojik danışman veya araştırmacı tarafından yanıtlanmıştır. Veri toplama sürecinin ardından araştırmacı tarafından puanlanan ve buna göre elde edilen profil kâğıdı okul psikolojik danışmanı aracılığı ile öğrencilere ulaştırılmıştır.

Canlı BOBUT Uygulaması ile Verilerin Toplanması. KDE-BOBUT uygulaması ve KDE'nin kağıt-kalem formundan elde edilen profilleri karşılaştırmak için veriler çalışma grubundaki 25 öğrenciden iki aşamalı bir süreç izlenerek toplanmıştır. Sıra etkisinin önüne geçmek için ilk oturumda 15 öğrenci kağıt-kalem formunu, diğer 10 öğrenci BOBUT formunu; ikinci oturumda ise 15 öğrenci BOBUT formunu ve 10 öğrenci kağıt-kalem formunu almıştır. Her iki oturum arasında iki hafta ara verilmiştir.

Uygulama özel bir seminer odasında gerçekleştirilmiş, çalışma grubundaki öğrenciler kendi dizüstü bilgisayarlarını getirmiş ve okulun kablosuz ağ altyapısı

kullanılmıştır. KDE-BOBUT uygulaması yaklaşık 20 dakikada sonlanmıştır. Yalnızca bir öğrencinin bilgisayarında çözülemeyen teknik bir sorun nedeniyle uygulama yaklaşık 44 dakika sürmüştür.

Öğrenci bütün alt ölçeklere ait maddelere yanıt verdikten sonra KDE-BOBUT uygulaması tamamen sonlanmakta ve öğrenciye görülen bir rapor sunulmaktadır. Bu raporda, uygulamanın kaç maddede sonlandığı ve KDE-BOBUT ile oluşturulan profil grafiği bulunmaktadır. Öğrenciler profillerini kaydedebilmekte ya da Profili Yazdır düğmesine tıklayarak yazdırabilmektedir.

Verilerin Analizi

KDE Verileri için MTK Model-Veri Uyumuna İlişkin Ön Analizler. Verilerin analizinde LISREL 8.7, IBM SPSS Statistics v21, PARSCALE, Firestar ve R yazılımları kullanılmıştır. Madde Tepki Kuramı için analizlere başlamadan önce tek boyutluluk varsayımı doğrulayıcı faktör analizi ile incelenmiştir. Envanterin yapısı belli olduğu için elde edilen verilerin envanterin yapısına uygun olup olmadığını belirlemek amacıyla doğrulayıcı faktör analizi tercih edilmiştir. Doğrulayıcı faktör analizi LISREL 8.7 yazılımı ile Yetenekler, İlgiler ve Mesleki Değerler boyutları için ayrı ayrı yapılmıştır.

KDE Verilerine Dayalı Post-Hoc Simülasyonlar. MTK için varsayımların test edilmesinin ardından KDE'nin madde parametreleri hesaplanmış ve Firestar (Choi, 2009) ile R yazılımında yapılacak simülasyon için gerekli sözdizimi oluşturulmuştur.

Simülasyonlar aşağıdaki kurallara göre gerçekleştirilmiştir:

1. Durdurma kuralı: En az 3 madde ve .30, .40 ve .50 standart hata
2. İlk madde seçimi: $\theta = .00$
3. MTK modeli: Genelleştirilmiş Kısmi Puan Modeli (GKPM) ve Kademeli Tepki Modeli (KTM)
4. Madde seçimi: Fisher En Yüksek Bilgi (FEYB), Beklenen En Yüksek Bilgi (BEYB), Beklenen En Düşük Sonsal Varyans (BEDSV), Beklenen En Yüksek Sonsal Ağırlıklandırılmış Bilgi (BEYSAB)
5. Madde kullanım kontrolü (Exposure control): Kullanma
6. En düşük θ : -3.00
7. En büyük θ : 3.00
8. θ artışı: .10
9. θ kestirim modeli: Beklenen Sonsal (BS – Expected a Posteriori, EAP)

10. Önsel dağılım (prior distribution): Ortalama = .00; Standart Sapma = 1.00
11. Standart hata hesaplama yöntemi: Sonsal (Posterior)
12. Ölçekleme (D): 1.7

Simülasyon sonuçlarına göre her bir alt ölçek için uygulamanın ortalama kaç maddede sonlandığı, ortalama standart hata ve tüm test ile simülasyon sonucu elde edilen θ değerleri arasındaki korelasyon katsayıları elde edilmiştir. Buna göre hangi durdurma kuralının, madde seçim yönteminin ve MTK modelinin envanter için en uygun olduğu belirlenmeye çalışılmıştır.

Canlı BOBUT Verilerinin Analizi. KDE-KK formu puanları ile KDE-BOBUT formundan kestirilen θ düzeyleri arasındaki korelasyonun hesaplanmasında, araştırma grubunun küçüklüğünden dolayı Spearman Sıra Farkları Korelasyon Katsayısı kullanılmıştır. KDE-KK ve KDE-BOBUT profillerinin karşılaştırmasının yapılabilmesi için araştırma grubundaki öğrencilerin KDE-KK formundan aldıkları ham puanlar standart z puanına dönüştürülmüştür. KDE-BOBUT uygulaması sonucunda her öğrenci için bir θ kestirimi yapılmıştır. θ kestirimleri genellikle -3 ve +3 arasında değer aldığından dolayı, z puanına dönüştürülen KDE-KK formu puanları ile aynı profil üzerinde karşılaştırma yapılması mümkün olmuştur. KDE-KK ve KDE-BOBUT profillerinin uyum gösterip göstermediğinin belirlenmesi için χ^2 uyum istatistiğinden yararlanılmış ve her öğrencinin 23 alt ölçeğinden bir χ^2 değeri hesaplanmış ve 22 serbestlik derecesine göre hesaplanan χ^2 değerinin manidar olup olmadığına karar verilmiştir.

BULGULAR

Tek Boyutluluğun Belirlenmesi

Kendini Değerlendirme Envanteri'nin faktörleri için tek boyutluluğun test edilmesi için yapılan doğrulayıcı faktör analizine ait uyum indeksleri Çizelge 1'de verilmiştir. Doğrulayıcı faktör analizinin yapılmasında maddelerin paralel, eşdeğer ya da konjenerik madde olup olmadığı da test edilmiştir.

Tabachnick ve Fidell (2007) RMSEA için .080 ve altı değerleri kabul edilebilir uyum olarak nitelendirmiştir. RMR için Hu ve Bentler (1995) .080 ve altı; CFI için .90 ve üstü değerleri kabul edilebilir uyum olarak ifade etmiştir. AGFI indeksi için .90 ve üstü değerler iyi uyum olarak kabul edilmiştir (Kline, 2005). Kline (2005), χ^2/sd oranının 3.00 ve altında olmasının iyi uyum olduğunu belirtirken, Wheaton, Muthen, Alwin ve Summers (1977) 5.00 ve altındaki değerlerin kabul edilebilir uyuma işaret ettiğini belirtmiştir. Bu ölçütler doğrultusunda, RMSEA ve RMR için .080; CFI ve AGFI için .90 ve χ^2/sd için 5.00 değerleri ölçüt olarak alınmış; yapılan iki aşamalı doğrulayıcı faktör analizi sonuçlarına göre maddeler, alt ölçekler ve boyutlar (yetenekler, ilgiler ve mesleki değerler) için faktör yapısının doğrulandığı bulunmuştur (Tablo 1).

Tablo 1. Doğrulayıcı Faktör Analizi Uyum İndeksleri

	χ^2/sd	RMSEA	RMR	CFI	AGFI
Yetenekler					
Paralel	4.60	.059	.055	.93	.88
Eşdeğer	3.67	.051	.049	.95	.90
Konjenerik	3.53	.047	.039	.96	.91
İlgiler					
Paralel	4.69	.074	.013	.90	.58
Eşdeğer	3.32	.056	.076	.94	.69
Konjenerik	3.14	.053	.068	.94	.71
Mesleki Değerler					
Paralel	-	-	-	-	-
Eşdeğer	-	-	-	-	-
Konjenerik	3.01	.052	.053	.95	.75

Tablo 1'de yer alan değerler incelendiğinde yalnızca AGFI değerinin veri model uyumunun olmadığını işaret ettiği görülmektedir. Ancak bu değer örneklem büyüklüğünden etkilenmektedir (Hooper, Coughlan ve Mullen, 2008). Diğer indeksler ise kabul edilebilir sınırlar arasında yer almaktadır. Ayrıca tüm boyutlar için maddelerin konjenerik olduğu sonucuna ulaşılmıştır.

Post-hoc Simülasyon Bulguları

Post-hoc simülasyonda kullanılan madde sayısı, tüm test ve simülasyon kestirimleri arasındaki korelasyonlar ile ortalama standart hata deęerinin ne olduęu araştırılmış ve KDE 23 alt ölçek verileri için farklı simülasyon kuralları ile elde edilen bulgular Tablo 2’de paylaşılmıştır. Tablo 2 incelendiğinde simülasyon sonuçlarına göre durdurma kuralı olarak .30 standart hata kullanıldığında, ortalama standart hatanın hiçbir alt ölçekte .30’un altında deęer almadığı görülmüştür. Durdurma kuralı .40 standart hata deęeri olarak alındığında, GKPM’nin her alt ölçekte .40’ın altına yaklaşık 4.9 sayıda madde ile ulaştığı ancak KTM’nin sayısal, mekanik, edebiyat ve güzel sanatlar alt ölçekleri dışında kalan 19 alt ölçekte .40’ın üzerinde bir ortalama standart hataya sahip olduęu bulunmuştur. Durdurma kuralı .50 standart hata deęeri olduęunda ise GKPM tüm alt ölçeklerde .50 standart hatanın altında bir ortalama standart hataya sahipken KTM için, İş birlięi dışındaki tüm alt ölçeklerde ortalama standart hata .50’nin altına düşmüştür. Post-hoc simülasyonlar GKPM ve KTM altında tekrarlandığında GKPM’nin daha az madde (orijinal madde sayısı olan 10 maddeden yaklaşık %50 daha az madde) ile .40’ın altında hata ile özellik kestirimi verebildiğı için MTK modeli olarak GKPM’nin seçilmesi uygun görülmüştür.

Tablo 2. KDE .30, .40 ve .50 Standart Hata Durdurma Kuralı İçin GKPM ve KTM Ortalama Madde, Standart Hata ve Tüm Test / Simülasyon θ Kestirimleri Arasındaki Korelasyon Değerleri

Alt Ölçek	ÖSH = .30						ÖSH = .40						ÖSH = .50					
	\bar{k}		$\overline{\text{ÖSH}}$		r		\bar{k}		$\overline{\text{ÖSH}}$		r		\bar{k}		$\overline{\text{ÖSH}}$		r	
	GKPM	KTM	GKPM	KTM	GKPM	KTM	GKPM	KTM	GKPM	KTM	GKPM	KTM	GKPM	KTM	GKPM	KTM	GKPM	KTM
Sözel	10.00	10.00	.333	.390	1.000	1.000	<u>5.366</u>	8.546	<u>.389</u>	.400	<u>.945</u>	.991	3.328	4.623	.459	.477	.883	.938
Sayısal	7.597	10.00	.300	.371	.980	1.000	<u>3.895</u>	7.185	<u>.385</u>	.396	<u>.915</u>	.981	3.091	3.723	.416	.481	.890	.926
Şekil-Uzay	10.00	10.00	.318	.423	1.000	1.000	<u>5.483</u>	9.714	<u>.384</u>	.426	<u>.943</u>	.999	3.225	5.993	.463	.485	.875	.958
Temel Bilimler	6.969	10.00	.305	.395	.980	1.000	<u>4.191</u>	8.136	<u>.371</u>	.410	<u>.925</u>	.993	3.339	4.524	.400	.479	.887	.944
Sosyal Bilimler	8.617	10.00	.309	.425	.984	1.000	<u>4.934</u>	9.730	<u>.383</u>	.428	<u>.903</u>	.997	3.224	6.215	.448	.486	.839	.955
Canlı Varlık	6.896	10.00	.303	.409	.977	1.000	<u>4.015</u>	8.837	<u>.372</u>	.418	<u>.914</u>	.994	3.246	4.941	.398	.483	.886	.955
Mekanik	5.203	8.483	.299	.319	.969	.997	<u>3.900</u>	4.792	<u>.333</u>	.385	<u>.947</u>	.967	3.450	3.586	.348	.421	.934	.945
İkna	8.719	10.00	.313	.429	.991	1.000	<u>4.706</u>	9.499	<u>.381</u>	.431	<u>.923</u>	.997	3.309	5.397	.431	.485	.878	.957
Ticaret	7.309	10.00	.329	.419	.980	1.000	<u>5.012</u>	8.556	<u>.384</u>	.429	<u>.933</u>	.993	3.935	5.298	.417	.487	.889	.947
İş Ayrıntıları	8.787	10.00	.302	.474	.989	1.000	<u>4.776</u>	10.00	<u>.375</u>	.474	<u>.924</u>	1.000	3.289	7.510	.430	.499	.886	.975
Edebiyat	6.067	9.653	.302	.342	.967	.999	<u>4.028</u>	5.756	<u>.351</u>	.394	<u>.932</u>	.969	3.239	3.821	.376	.445	.906	.939
Güzel Sanatlar	6.500	10.00	.305	.376	.973	1.000	<u>4.154</u>	10.00	<u>.364</u>	.376	<u>.927</u>	1.000	3.306	4.160	.390	.468	.895	.946
Müzik	6.503	10.00	.306	.377	.973	1.000	<u>4.156</u>	7.135	<u>.365</u>	.404	<u>.929</u>	.984	3.319	4.176	.390	.469	.898	.947
Sosyal Yardım	6.275	9.999	.303	.375	.974	1.000	<u>3.783</u>	7.130	<u>.366</u>	.403	<u>.928</u>	.985	3.333	4.119	.382	.469	.910	.942
Yeteneğini Kullanma	8.691	8.921	.323	.413	.990	.995	<u>5.619</u>	8.921	<u>.386</u>	.413	<u>.942</u>	.995	3.784	5.555	.451	.481	.878	.947
Yaratıcılık	10.00	10.00	.322	.454	1.000	1.000	<u>5.596</u>	9.983	<u>.388</u>	.454	<u>.932</u>	1.000	3.309	6.470	.463	.493	.865	.955
Yarışma	9.736	10.00	.316	.434	.998	1.000	<u>5.733</u>	9.906	<u>.386</u>	.435	<u>.932</u>	1.000	3.420	6.750	.471	.487	.850	.966
İşbirliği	10.00	10.00	.368	.512	1.000	1.000	<u>7.334</u>	10.00	<u>.395</u>	.512	<u>.960</u>	1.000	4.409	9.111	.469	.518	.878	.989
Değişiklik	10.00	10.00	.342	.467	1.000	1.000	<u>6.406</u>	10.00	<u>.396</u>	.467	<u>.949</u>	1.000	3.911	7.763	.474	.497	.879	.982
Düzenli Yaşam	10.00	10.00	.330	.460	1.000	1.000	<u>5.926</u>	10.00	<u>.391</u>	.460	<u>.938</u>	1.000	3.503	7.353	.470	.494	.865	.979
Liderlik	9.580	10.00	.312	.407	.998	1.000	<u>5.465</u>	9.307	<u>.381</u>	.414	<u>.929</u>	.998	3.257	5.612	.465	.483	.879	.955
Kazanç	7.544	10.00	.308	.405	.986	1.000	<u>4.111</u>	8.376	<u>.382</u>	.414	<u>.934</u>	.993	3.179	4.480	.413	.482	.909	.960
Ün Sahibi Olma	8.493	10.00	.312	.429	.989	1.000	<u>4.377</u>	9.513	<u>.384</u>	.431	<u>.931</u>	.998	3.290	5.191	.423	.486	.898	.964

Durdurma kuralı olarak .40 standart hata değeri, GKPM modeli temel alındığında FEYB, BEYB, BEDSV ve BEYSAB madde seçim yöntemleri ile elde edilen ortalama madde sayısı, standart hata ve toplam / simülasyon θ kestirimleri arasındaki korelasyonlar hesaplanarak en uygun madde seçim yöntemi bulunmaya çalışılmış ve elde edilen değerler Tablo 3'te verilmiştir.

Tablo 3. KDE.40 Standart Hata Durdurma Kuralı için FEYB, BEYB, BEDSV ve BEYSAB Madde Seçim Yöntemlerine Göre Ortalama Madde, Standart Hata ve Tüm Test / Simülasyon θ Kestirimleri Arasındaki Korelasyon Değerleri

Alt Ölçek	FEYB			BEYB			BEDSV			BEYSAB		
	\bar{k}	ÖSH	r	\bar{k}	ÖSH	r	\bar{k}	ÖSH	r	\bar{k}	ÖSH	r
Sözel	5366	389	.945	5371	390	.945	5372	390	.945	5364	390	.945
Sayısal	3895	385	.915	3888	385	.912	3895	385	.913	3894	385	.913
Şekil-Uzay	5483	384	.943	5474	384	.943	5474	384	.943	5475	384	.943
Temel Bilimler	4191	371	.925	4183	369	.921	4190	370	.922	4201	371	.925
Sosyal Bilimler	4934	383	.903	4934	383	.903	4934	383	.903	4934	383	.903
Canlı Varlık	4015	372	.914	3962	369	.905	3955	370	.903	4063	369	.907
Mekanik	3900	333	.947	3929	332	.948	3900	333	.947	3907	333	.946
İkna	4706	381	.923	4704	381	.923	4706	381	.923	4708	381	.922
Ticaret	5012	384	.933	5100	378	.936	5008	382	.934	5076	381	.935
İş Ayrıntıları	4776	375	.924	4776	375	.924	4776	375	.924	4776	375	.924
Edebiyat	4028	351	.932	4096	348	.937	4031	351	.932	4031	351	.932
Güzel Sanatlar	4154	364	.927	4100	362	.928	4161	362	.929	4166	362	.929
Müzik	4156	365	.929	3520	342	.878	3534	342	.878	3532	342	.878
Sosyal Yardım	3783	366	.928	3780	366	.929	3783	366	.929	3787	366	.929
Yeteneğini Kullanma	5619	386	.942	5642	385	.942	5618	386	.942	5626	386	.941
Yaratıcılık	5596	388	.932	5597	388	.933	5596	388	.932	5597	388	.932
Yarışma	5733	386	.932	5727	386	.933	5729	386	.932	5729	386	.932
İşbirliği	7334	395	.960	7316	395	.960	7332	395	.960	7330	396	.961
Değişiklik	6406	396	.949	6401	396	.950	6404	396	.949	6403	396	.950
Düzenli Yaşam	5926	391	.938	5901	390	.940	5923	391	.939	5927	391	.940
Liderlik	5465	381	.939	5465	381	.939	5464	381	.939	5466	381	.940
Kazanç	4111	382	.934	4073	382	.937	4080	382	.936	4124	382	.935
Ün Sahibi Olma	4377	384	.931	4344	384	.931	4352	384	.932	4397	384	.931

Tablo 3 incelendiğinde farklı madde seçim yöntemlerinin, uygulanan ortalama madde sayısı, standart hata ya da tüm maddelerle kestirilen θ düzeyleri (*tüm- θ*)

ile simülasyon ile kestirilen θ düzeyleri (*sim- θ*) arasındaki korelasyon değerlerinde büyük bir değişime neden olmadığı görülmektedir.

Canlı BOBUT Uygulaması Bulguları

BOBUT uygulamasının ortalama kaç madde ile sonlandığını belirlemek için KDE-KK ve BOBUT uygulamasında öğrencilere uygulanan toplam madde sayılarının alt ölçeklere göre ortanca değerleri hesaplanmıştır. KDE-KK formunda öğrenciler algılanan yetenekler boyutu için toplamda sabit sayıda (30) madde alırken, BOBUT uygulamasında karşılaştıkları toplam madde sayısı ortancası 15'tir. KDE-KK formunda öğrenciler ilgi boyutuna ait 110 maddeye yanıt vermeleri gerekirken, KDE-BOBUT formunda öğrencilerin ilgi boyutunda yanıtladıkları toplam madde sayısı ortancası 45'tir. KDE-KK formunda mesleki değerler boyutu için 90 madde yanıtlaması gereken öğrencilerin, KDE-BOBUT formunda yanıtladıkları toplam madde sayısı ortancası 50 olarak bulunmuştur.

Envanterin tamamı dikkate alındığında ise KDE-KK formunda 230 madde alan öğrencilerin, KDE-BOBUT formunda aldıkları madde sayısı ortancası 110'dur. Başka bir ifadeyle, KDE-BOBUT formu KDE-KK formuna göre yaklaşık olarak %52 daha az madde ile sonlanmıştır.

Öğrencilerin KDE'nin kağıt-kalem formundan aldıkları yüzdeler ile KDE-BOBUT formundan kestirilen θ düzeyleri arasındaki ilişkiyi araştırmak amacıyla her bir alt ölçek için Spearman sıra farkları korelasyon katsayısı hesaplanmıştır. Buna göre, tüm alt ölçekler için KDE-KK ve KDE-BOBUT puanları arasında $p = .05$ düzeyinde manidar korelasyon katsayıları elde edilmiştir. Ancak yeteneği kullanma, yarışma ve işbirliği alt ölçekleri için sırasıyla .50, .45 ve .55 katsayıları elde edilmiştir. Bu alt ölçeklerin kağıt-kalem formuna ait Cronbach alfa katsayıları da .73, .63 ve .68 olarak bulunmuştur. Post-hoc simülasyon sonuçlarına göre de bu boyutların beşten fazla madde ortalaması ile sonlandığı görülmektedir.

Öğrencilerin kağıt-kalem formu ve BOBUT uygulamasından elde edilen profilleri arasındaki uyumun belirlenmesinde χ^2 uyum istatistiğinden yararlanılmıştır. Buna göre her bir öğrencinin KDE'nin 23 alt ölçeğine ait KDE-KK ve KDE-BOBUT profilleri arasındaki uyumun nasıl olduğu belirlenmiş ve buna göre; 1, 5, 22, 23 ve 25 numaralı öğrencilere ait kağıt-kalem

ve BOBUT profillerinin uyum göstermediği ($p < .05$); kalan 20 öğrenci içinse profiller arasında uyumun olduğu ($p > .05$) bulunmuştur.

SONUÇ VE TARTIŞMA

Sonuç olarak, simülasyon bulgularına göre KTM'nin belirlenen standart hata ölçütünün altında θ kestirimi yapılabilmesi için GKPM'ye göre daha çok madde kullandığı, standart hata ölçütü .30 olduğunda hem KTM hem de GKPM'nin neredeyse tüm alt ölçeklerde tüm maddeleri kullandığı görülmüştür. Standart hata ölçütü .50 alındığında ise tüm maddeler kullanılarak kestirilen θ düzeyleri (*tiim- θ*) ile simülasyon sonuçlarına göre kestirilen θ düzeyleri (*sim- θ*) arasındaki korelasyonun düştüğü gözlemlenmiştir. Farklı madde seçim yöntemlerinin pratikte simülasyon sonuçlarını değiştirmede bulunmuş ve tüm bu bulgulardan yola çıkılarak KDE'ye en uygun MTK modelinin GKPM; madde seçim yönteminin FEYB, standart hata durdurma kuralının .40 olduğuna karar verilmiştir.

KDE-KK ve KDE-BOBUT profilleri arasındaki korelasyonlar yeteneği kullanma alt ölçeği için .50, yarışma alt ölçeği için .45, iş birliği alt ölçeği için .55 olarak bulunmuştur. Kalan 20 alt ölçek içinse korelasyon katsayıları .62 ile .88 arasında değişmektedir.

Hem simülasyon hem de canlı BOBUT uygulaması için ilk madde seçilirken, bireye ait herhangi bir bilgiye sahip olunmadığı için $\theta = 0$ olarak kabul edilmiş ve bu düzeyde en çok bilgi veren maddenin seçilerek bireye uygulanması sağlanmıştır. Ancak simülasyon ve canlı BOBUT uygulamasına ait madde kullanım sıklıkları incelendiğinde, simülasyonda tüm öğrencilere uygulanan bazı maddelerin, canlı BOBUT uygulamasında hiçbir öğrenci için kullanılmadığı; benzer şekilde canlı BOBUT uygulamasında tüm öğrencilere uygulanan bazı maddelerinse simülasyonda çok az öğrenciye uygulandığı görülmüştür.

Araştırma sonuçları KDE simülasyonunun yaklaşık %50; KDE-BOBUT'un ise %52 daha az madde kullandığını, ayrıca KDE-BOBUT'un yaklaşık olarak %60 daha kısa sürede sonlandığını göstermektedir. KDE-KK ve KDE-BOBUT profilleri karşılaştırıldığında ise çoğu öğrenci için profillerin uyum gösterdiği bulunmuştur.

Farklı madde seçim yöntemlerinin, uygulanan ortalama madde sayısı, standart hata ya da *tüm-θ* ile *sim-θ* arasındaki korelasyon değerlerinde büyük bir değişime neden olmadığı görülmektedir. Bu bulgu Choi ve Swartz (2009)'ın KTM altında yürüttüğü çalışmalarının, küçük madde havuzlarında madde seçim yönteminin uygulanan madde sayısı ve kestirilen θ düzeyinde farklılığa yol açmadığı yönündeki bulguları ile paralellik göstermektedir. Veldkamp (2003) ise GKPM altında farklı madde seçim yöntemlerini karşılaştırdığı simülasyon çalışmasında bu model altında farklı madde seçim yöntemleriyle %85 ile %100 arasında bir oranda aynı maddelerin seçildiğini ifade etmiştir.

GKPM altında, FEYB madde seçim yöntemi ve .40 standart hata durdurma kuralı ile simülatif BOBUT uygulaması ortalama 113 madde ile sonlanmıştır. KDE-KK formunun 230 maddeden oluştuğu düşünüldüğünde, envanterin yaklaşık %50 oranında daha az madde ile sonlanabileceği bulunmuştur. Bunun yanında, test uzunluğu %50 oranında azalırken, *tüm-θ* ile *sim-θ* düzeyleri arasında .90 ile .96 arasında değişen korelasyon katsayıları elde edilmiştir. Bu bulgular KDE gibi bir ilgi envanteri olan Strong İlgili Envanteri'nin BOBUT olarak uygulanabilirliğini araştıran ve BOBUT simülasyonu ile madde sayısının yaklaşık olarak %60 oranında azaldığına dikkat çeken ve *tüm-θ* ile *sim-θ* kestirimleri arasında .90 ile .98 aralığında korelasyon katsayıları elde eden Scullard (2007)'in bulguları ile paralellik göstermektedir. Benzer şekilde BOBUT simülasyonu ile daha az madde kullanarak *tüm-θ* ile *sim-θ* arasında yüksek korelasyon katsayıları elde eden Hol, Vorst ve Mellenbergh (2007), Chien ve diğerleri (2011), Smits, Cuijpers ve van Straten (2011), Betz ve Turner (2011), Gibbons ve diğerleri (2012)'nin çalışmaları desteklenmektedir.

Madde sayıları ile birlikte öğrencilerin KDE-BOBUT uygulamasına başlama ve bitirme zamanları kaydedilmiş ve iki zaman arasındaki fark hesaplanarak her bir öğrencinin KDE-BOBUT formunu toplam yanıtlama süresi elde edilmiştir. Buna göre öğrencilerin KDE-BOBUT formunu tamamlama süresi ortanca 18 dakika 51 saniye olarak bulunmuştur. KDE-KK formunun bir ders saati ve bir teneffüste, yaklaşık 50 dakikada sonlandığı düşünüldüğünde öğrenciler KDE-BOBUT'u, KDE-KK'e göre yaklaşık olarak %60 oranında daha kısa sürede tamamlamışlardır.

BOBUT uygulamasında test uzunluğunun kısalması Hol, Vorst ve Mellenbergh (2007)'in BOBUT uygulamasının tüm teste göre %62.5 oranında; Elhan ve diğerlerinin (2008), iki faktörlü bir ölçeğin BOBUT uygulaması ile sırasıyla %42

ve %71 oranında; Gibbons ve diğerleri (2008)'in 626 maddelik bir ölçekte test uzunluğunun %95 oranında; Kocalevent ve diğerlerinin (2009) 104 maddelik bir ölçeğin BOBUT uygulaması ile %85 oranında kısaltıldığına, başka bir deyişle ölçme aracının kullanılabilirliğinin arttığına yönelik bulguları desteklemektedir.

Araştırma bulguları KDE'nin BOBUT formunda uygulanabilir olduğunu gösterse de gerek bulgular gerekse araştırma sürecinde karşılaşılan zorluklardan yola çıkarak aşağıdaki önerilerin yapılması mümkündür:

1. Araştırmada MTK modeli olarak yalnızca GKPM ve KTM kullanılmıştır. Çok kategorili puanlanan maddeler için geliştirilmiş diğer MTK modelleri ile benzer çalışmalar yürütülebilir. KDE'nin her bir boyutu ve alt ölçeklerine ilişkin veriler ayrı ayrı simülasyon çalışmasına alınmıştır. Benzer şekilde canlı BOBUT uygulamasında da her bir alt ölçek ayrı ayrı uygulanmış ve öğrencinin bir alt ölçekte 0 düzeyi tamamen kestirildikten sonra, sonraki alt ölçeğe geçilmiştir. Yapılacak çalışmalarda bu türden çok boyutlu yapıları ölçen psikolojik testlerde post-hoc BOBUT simülasyonları ve uygulamaları içerik dengeleme yöntemleri kullanılarak yeni post-hoc simülasyonlar yapılabilir.
2. Concerto üzerinden uygulanan KDE-BOBUT için ilk maddenin, simülasyonda seçilen ilk madde olması sağlanarak yeni karşılaştırma çalışmaları yürütülebilir.
3. Araştırmada KDE-BOBUT formu Concerto platformunda geliştirilmiştir. Farklı BOBUT yazılımları ile KDE-BOBUT uygulamaları tekrarlanabilir. Araştırmada KDE-BOBUT formu catR paketi ve nextItem fonksiyonu kullanılarak geliştirilmiştir, farklı R paketleri ve fonksiyonları kullanılarak yeni çalışmalar yürütülebilir.
4. KDE'nin alt boyutlarında, geliştiricisinin de onayı alınarak, yeni maddeler eklenip parametreleri kestirilerek madde havuzu güncellenebilir. Madde havuzunun bu şekilde genişlemesi, özellikle yeteneği kullanma, yarışma ve iş birliği alt ölçekleri göz önüne alındığında, envanteri BOBUT uygulamaları için daha elverişli hale getirebilir.
5. KDE-BOBUT sonucunda elde edilen profilin yorumlanması için uzman psikolojik danışmanlar eşliğinde yeni bir yönergenin hazırlanması KDE-BOBUT'un yaygın bir şekilde kullanılabilmesi için faydalı olabilir.

6. KDE Kađıt-Kalem formu ve KDE-BOBUT'un uygulanması ve her iki uygulamadan alınan dönütlerin özellikleri bakımından öğrencilerin ve psikolojik danışmanların görüşlerinin incelenmesinin KDE-BOBUT'un daha kullanışlı hale getirilmesine katkı sağlayacağı düşünülmektedir.

Yazarlar Hakkında / About Authors

Eren Can Aybek. Lisans eğitimini Balıkesir Üniversitesi Fizik Öğretmenliği, yüksek lisans ve doktora eğitimlerini ise Ankara Üniversitesi Ölçme ve Değerlendirme programlarında tamamlamıştır. Halen Pamukkale Üniversitesi Eğitim Fakültesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı'nda Dr. Öğretim Üyesi olarak görevini sürdürmektedir. Bilgisayar Ortamında Bireye Uyarlanmış Testler ve Madde Tepki Kuramı üzerine çalışmalarını sürdürmektedir.

He completed his undergraduate studies at Balıkesir University Physics Teaching Department and his master's and doctoral studies at Ankara University Measurement and Evaluation programs. He is currently serving as a faculty member in the Department of Measurement and Evaluation in Education at Pamukkale University. He continues to work on Computerized Adaptive Tests and Item Response Theory.

R. Nükhet Çıkrıkçı. Lisans ve yüksek lisans eğitimini Hacettepe Üniversitesi Psikoloji, doktora eğitimini ise Ankara Üniversitesi Ölçme ve Değerlendirme programlarında tamamlamıştır. Halen İstanbul Aydın Üniversitesi'nde Ölçme ve Değerlendirme Uygulama ve Araştırma Merkezi Müdürlüğü görevini yürütmektedir. Test Geliştirme, Bilgisayar Ortamında Bireye Uyarlanmış Testler ve Madde Tepki Kuramı üzerine çalışmalarını sürdürmektedir.

She completed her undergraduate and graduate education at Hacettepe University Psychology and PhD at Ankara University Measurement and Evaluation programs. She is still working as the Director of Measurement and Evaluation Research & Application Center at Istanbul Aydın University. She continues to work on Test Development, Computerized Adaptive Testing, and Item Response Theory.

Yazar Katkıları / Author Contributions

Bu araştırma, RNÇ danışmanlığında ECA tarafından hazırlanmış olan *Kendini Değerlendirme Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) Olarak Uygulanabilirliğinin Araştırılması* başlıklı doktora tezinden üretilmiştir.

This research has been produced from the doctoral dissertation titled "An Investigation of Applicability of the Self-Assessment Inventory as a Computerized Adaptive Test (CAT)" which prepared by ECA under the advice of RNÇ.

Çıkar Çatışması / Conflict of Interest

Yazarlar tarafından çıkar çatışmasının olmadığı rapor edilmiştir.

It has been reported by the authors that there is no conflict of interest.

Fonlama / Funding

Herhangi bir fon desteđi alınmamıştır.


No funding was received.

Etik Bildirim / Ethical Standards

Araştırmanın yapılması için Milli Eğitim Bakanlığı'ndan gerekli izinler alınmış ve yalnızca araştırmaya katılım için gönüllü olan öğrenciler ile çalışılmıştır.

The necessary approvals were obtained from the Ministry of National Education for the purpose of conducting the research and only the students who volunteered to participate in the research were employed.

ORCID

Eren Can Aybek  <https://orcid.org/0000-0003-3040-2337>

R. Nüket Çıkrıkçı  <https://orcid.org/0000-0001-8853-4733>

KAYNAKÇA

- Babcock, B., ve Weiss, D. (2012). Termination criteria in computerized adaptive tests: Do variable - length CATs provide efficient and effective measurement? *Journal of Computerized Adaptive Testing*, 1(1), 1–18. doi:10.7333/1212-0101001
- Betz, N. E., ve Turner, B. M. (2011). Using item response theory and adaptive testing in online career assessment. *Journal of Career Assessment*, 19(3), 274–286. doi:10.1177/1069072710395534
- Cheng, Y. (2008). *Computerized adaptive testing – new developments and applications*. Doktora Tezi. Illionis Üniversitesi.
- Chien, T.W., Lai, W.P., Lu, C.W., Wang, W.C., Chen, S.C., Wang, H.Y., ve Su, S.B. (2011). Web-based computer adaptive assessment of individual perceptions of job satisfaction for hospital workplace employees. *BMC medical research methodology*, 11, 47.
- Chien, T.W., Wang, W.C., Wang, H.Y., ve Lin, H.J. (2009). Online assessment of patients' views on hospital performances using Rasch model's KIDMAP diagram. *BMC health services research*, 9, 135.
- Choi, S. W., Grady, M. W., ve Dodd, B. G. (2011). A new stopping rule for computerized adaptive testing. *Educational Psychological Measurement*, 70(6), 1–17. doi:10.1177/0013164410387338.A
- Choi, S. W. (2009). Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*, 33(8), 644–645. doi:10.1177/0146621608329892
- Choi, S. W., ve Swartz, R. J. (2009). Comparison of CAT item selection criteria for polytomous items. *Applied Psychological Measurement*, 33(6), 419–440. doi:10.1177/0146621608327801
- Davey, T. (2011). A guide to computer adaptive testing systems. *Council of Chief State School Officers*.
- Downing, S. M., ve Haladyna, T. M. (2006). *Handbook of test development*. New Jersey: Lawrence Erlbaum Assoc.
- Elhan, A. H., Öztuna, D., Kutlay, Ş., Küçükdeveci, A. A., & Tennant, A. (2008). An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. *BMC musculoskeletal disorders*, 9(1), 1.

- Gardner, W., Shear, K., Kelleher, K. J., Pajer, K. A., Mammen, O., Buysse, D., ve Frank, E. (2004). Computerized adaptive measurement of depression: a simulation study. *BMC Psychiatry*, 4(1).
- Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., ... & Immekus, J. C. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services*.
- Gibbons, R. D., Weiss, D. J., Pilkonis, P. a, Frank, E., Moore, T., Kim, J. B., ve Kupfer, D. J. (2012). Development of a computerized adaptive test for depression. *Archives of General Psychiatry*, 69(11), 1104–1112.
- Haley, S. M., Coster, W. J., Andres, P. L., Kosinski, M., & Ni, P. (2004). Score comparability of short forms and computerized adaptive testing: simulation study with the activity measure for post-acute care. *Archives of physical medicine and rehabilitation*, 85(4), 661-666.
- Haley, S. M., Ni, P., Jette, A. M., Tao, W., Moed, R., Meyers, D., & Ludlow, L. H. (2009). Replenishing a computerized adaptive test of patient-reported daily activity functioning. *Quality of Life Research*, 18(4), 461-471.
- Hart, D. L., Mioduski, J. E., ve Stratford, P. W. (2005). Simulated computerized adaptive tests for measuring functional status were efficient with good discriminant validity in patients with hip, knee, or foot/ankle impairments. *Journal of clinical epidemiology*, 58(6), 629-638.
- Hart, D. L., Cook, K. F., Mioduski, J. E., Teal, C. R., & Crane, P. K. (2006). Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. *Journal of Clinical Epidemiology*, 59(3), 290-298.
- Hol, a. M., Vorst, H. C. M., ve Mellenbergh, G. J. (2007). Computerized adaptive testing for polytomous motivation items: Administration mode effects and a comparison with short forms. *Applied Psychological Measurement*, 31(5), 412–429.
- Hooper, D., Coughlan, J., ve Mullen, M.: Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Hu, L., Bentler, P. M. (1995). Evaluating model fit in R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications*. London: Sage Publications.

- Kaya, A. (2011). *Kariyer danışmanlığında değerlendirme*. B. Yeşilyaprak (Ed.). *Mesleki rehberlik ve kariyer danışmanlığı: Kuramdan uygulamaya*. Ankara: Pegem Akademi.
- Kocalevent, R. D., Rose, M., Becker, J., Walter, O. B., Fliege, H., Bjorner, J. B., ... & Klapp, B. F. (2009). An evaluation of patient-reported outcomes found computerized adaptive testing was efficient in assessing stress perception. *Journal of Clinical Epidemiology*, 62(3), 278-287.
- Kline, R.B. (2005). *Principles and practices of structural equation modelling*. New York: The Guilford Press.
- Kuzgun, Y. (2004). *Meslek rehberliği ve danışmanlığına giriş*. Ankara: Nobel Yayıncılık.
- Kuzgun, Y. (2014). *Kendini değerlendirme envanteri el kitabı*. Ankara: Nobel Yayıncılık.
- Linacre, J.M. (2000). Computer-adaptive testing: a methodology whose time has come. MESA memorandum no. 69. S. Chae, U. Kang, E. Jeon ve J. M. Linacre (Ed.). *Development of computerised middle school achievement test*, Seoul, South Korea: Komesa Press.
- Linden, W., ve Glas, G. (2002). *Computerized adaptive testing: Theory and practice*. New York: Kluwer Academic Pub.
- Magis, D., Raiche, G., ve Barrada, J.R. (2016). *Package 'catR'*. 15 Mart 2016 tarihinde <https://cran.r-project.org/web/packages/catR/catR.pdf> adresinden erişilmiştir.
- Milli Eğitim bakanlığı (MEB) (2015). *Okullarda rehberlik ve psikolojik danışma hizmetleri kılavuzu*. Ankara: Milli Eğitim Bakanlığı.
- Oswald, F. L., Shaw, A., ve Farmer, W. L. (2015). Comparing simple scoring with IRT scoring of personality measures the navy computer adaptive personality scales. *Applied Psychological Measurement*, 39(2), 144-154.
- Özbaşı, D., ve Demirtaşlı, N. (2015). Bilgisayar okuryazarlığı testinin bilgisayar ortamında bireye uyarlanmış test olarak geliştirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(2), 218-237.
- Scalise, K., ve Allen, D. D. (2015). Use of open- source software for adaptive measurement: Concerto as an R- based computer adaptive development and delivery platform. *British Journal of Mathematical and Statistical Psychology*, 68(3), 478-496.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, 8(4), 350.

- Scullard, M. G. (2007). *Application of item response theory based computerized adaptive testing to the Strong Interest Inventory*. Doktora Tezi. Minnesota Üniversitesi.
- Simms, L. J., Goldberg, L. R., Roberts, J. E., Watson, D., Welte, J., ve Rotterman, J. H. (2011). Computerized adaptive assessment of personality disorder: Introducing the CAT-PD project. *Journal of Personality Assessment*, 93(4), 380–9. doi:10.1080/00223891.2011.577475
- Smits, N., Cuijpers, P., ve van Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry research*, 188(1), 147–55. doi:10.1016/j.psychres.2010.12.001
- Tabachnick, B. G., ve Fidell, L. S. (2007). *Using multivariate statistics*. USA: Pearson Education.
- Veldkamp, B.P. (2003). Item selection in polytomous CAT. H. Yanai, A. Okada, K. Shigemasu, Y. Kano, ve J.J. Meulman (Ed.). *New developments in psychometrics* (s.207-214). Tokyo, Japan: Springer Verlag
- Wainer, H. ve Dorans, N. (2000). *Computerized adaptive testing: A primer* New Jersey: Lawrence Erlbaum Assoc.
- Walter, O. B., Becker, J., Bjorner, J. B., Fliege, H., Klapp, B. F., & Rose, M. (2007). Development and evaluation of a computer adaptive test for 'Anxiety'(Anxiety-CAT). *Quality of Life Research*, 16(1), 143-155.
- Weiss, D. J., ve Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375.
- Wheaton, B., Muthen, B., Alwin, D. F., ve Summers, G. F. (1977). Assessing reliability and stability in panel models. D. R. Heise (Ed.), *Sociological Methodology* (s. 84–136). San Francisco: Jossey-Bass.

Extended Abstract

Introduction: The Self-Assessment Inventory, which is included in the research, is a tool that is frequently used in schools for vocational guidance services and is still being introduced to candidates for psychological counselors in training faculties. There are 230 items on this inventory and it can take a long time to reply to the items. It is also time-consuming to determine which subscale belongs to the inventory from the handbook after the application, to calculate the raw scores accordingly and then to mark the score on the profile sheet by converting the scores into percentiles using the norm table. In Turkey there is no computerized adaptive tool used in the field of vocational guidance developed to measure affective characteristics. Investigating the applicability of Self-Assessment Inventory as a Computerized Adaptive Test form and opening a new pathway to facilitate student diagnostic services in vocational guidance practices makes this work significant and important.

Method: Two separate study groups were included in the study. The first study group is consisting of 1367 students in 9th, 10th and 11th grade. Because of missing data in this group, it was worked with the net of 1144 students. This data was used for post-hoc simulation studies. In the second study group, both paper-pencil and developed live CAT application were carried out. The students in the first study group were not included in the second study group. The research process was announced to the students and worked with 25 volunteers to participate in the research.

Data Collection Tools: Self-Assessment Inventory Paper-Pencil Form (SAI-PP). Developed by Kuzgun (2014) for the first time in 1989, this inventory aims to measure the individual's perceived abilities, interests and vocational values. There are 23 subscales and 230 items in total in three dimensions in the inventory. Each subscale consists of 10 items. The items are scored in four grades. *Self-Assessment Inventory Computerized Adaptive Test Form (SAI-CAT).* As a result of the post-hoc simulations, the SAI-CAT developed on the Concerto platform according to the decided parameters. SAI-CAT was developed using R scripts, HTML and CSS codes, and MySQL database. *Data Analysis: Post-Hoc Simulations.* According to the simulation results, it has been tried to determine which stopping rule, item selection method and IRT model are most appropriate for the inventory. In addition, test information functions for different θ levels were established according to the determined stopping rule and IRT model.

Data Analysis for Live CAT Application. The Spearman's Rank Correlation Coefficient was used to calculate the correlation between the SAI-PP form scores and the θ levels estimated from the SAI-CAT form, due to the smallness of the study group. The

frequency of items usage was determined and frequency analysis was performed for the indicated items.

In order to compare SAI-PP and SAI-CAT profiles after calculation of correlation coefficients and item usages, the raw scores of the students in the study group from the SAI-PP form have been converted to standard z-scores.

As a result of SAI-CAT application, a θ estimation is done for each student. Since the θ estimation takes a value between -3 and +3, it is possible to make a comparison on the same profile with SAI-PP formulas converted to z-score. In order to determine whether the SAI-PP and SAI-CAT profiles are compatible, a χ^2 statistic was calculated from 23 subscales of each student and it was decided whether the value of χ^2 calculated according to the 22 degrees of freedom was significant.

Results: According to the simulation findings, when GRM uses the standard error criterion .30, which uses more items than GPCM to make the θ estimation below the standard error measure, it is seen that both GRM and GPCM use all the items in almost all subscales. When the standard error criterion .50 was taken, it was observed that the correlation between the estimated θ levels using all the items (all- θ) and the estimated θ levels from the simulation (sim- θ) decreased. It has been found that different methods of item selection do not change the simulation results in practice and it is found out that the most suitable IRT model for SAI is GPCM; it is decided that the item selection method is MFI and the standard error stop rule is .40.

Correlations between SAI-PP and SAI-CAT profiles were found as .50 for the using ability subscale, .45 for the competition subscale, and .55 for the cooperation subscale. For the remaining 20 subscales, the correlation coefficients ranged from .62 to .88. For both simulation and live CAT application, the first item is selected, $\theta = 0$ because no information about the individual is available, and the most informative item at this level is selected and applied to the individual. The research results are about 50% of the CAT simulation; SAI-CAT has 52% fewer items and SAI-CAT is about 60% shorter. Compared to the SAI-PP and SAI-CAT profiles, the profiles were found to be compatible for most students.

Discussion & Conclusion: The results of the study can be summarized below:

1. According to the results of SAI-CAT simulation, GPCM as IRT model, MFI as the item selection method and the .40 standard error as the stopping rule gives the optimum trait estimation.
2. SAI-CAT, stopped about 52% less item than the paper-pencil form; It ended in 60% shorter time.
3. Student profiles from SAI-PP and SAI-CAT are not different each other.