



Bireye Uyarlanmış Testlerde Madde Kullanım Sıklığı Kontrol Yöntemlerinin Farklı Koşullarda Ölçme Duyarlılığına ve Test Güvenliğine Etkisi *

Recep Gür ¹, H. Deniz Gülleroğlu ²

Öz

Bu araştırmada, bilgisayar ortamında bireye uyarlanmış test (BOBUT) uygulamalarında 1-0 şeklinde puanlanan maddelerde, madde kullanım sıklığı kontrol yöntemlerinin farklı örneklem büyüklüklerinde, test uzunluklarında ve yetenek dağılımlarında ölçme duyarlılığına ve test güvenliğine etkisinin incelenmesi amaçlanmıştır. Bu araştırma bir Monte Carlo simülasyon çalışmasıdır. Çalışma kapsamında uzun test 50, kısa test 25 madde; geniş örneklem 1000, küçük örneklem 250 birey olarak belirlenmiştir. Her bir örneklem büyüklüğü koşulu için, testi alan bireylerin yetenek parametresi (θ) değeri -3 ve +3 aralığında sola çarpık, tek biçimli, normal ve sağa çarpık dağılım gösterecek şekilde BOBUT örneklemi oluşturulmuştur. Çalışma kapsamında; Sympson Hetter stratejisi ve Aşamalı Düşürme madde kullanım sıklığı kontrol yöntemleri kullanılmış ve kullanım sıklığının kontrol edilmediği durum referans alınmıştır. Çalışmada 48 simülasyon koşulu belirlenmiş ve her bir koşulda 100 tekrar yapılarak 4800 veri dosyası oluşturulmuştur. Her bir koşul için test güvenliği ve ölçme duyarlılığı indeksleri hesaplanıp, karşılaştırılmalar yapılmıştır. Ölçme duyarlılıklarına ilişkin farklı madde kullanım sıklığı kontrolü koşullarında aralarında genel olarak büyük bir farklılık bulunmamıştır. İlgili koşullarda Aşamalı Düşürme yöntemi seçildiğinde, test güvenliği daha iyi sağlanmıştır. Dolayısıyla Aşamalı Düşürme yöntemi kullanıldığında, madde kullanım sıklığının dengeli dağılım göstererek, madde havuzu kullanımının daha verimli hale geldiği ve madde havuzu ifşâ olmadan uygulamalar için süreklilik sağlayabildiği sonucuna ulaşılmıştır. BOBUT uygulamasından önce, tekrarlı simülasyonlarla kullanım sıklığı kontrol parametreleri belirlenen Sympson Hetter stratejisi BOBUT uygulamalarında en yaygın kullanılan madde kullanım sıklığı kontrol yöntemidir. Oysaki BOBUT uygulamalarında farklı ölçme koşullarında, BOBUT uygulanma sürecinde kullanım sıklığı

Anahtar Kelimeler

Bilgisayar Ortamında Bireye Uyarlanmış Test
Madde Kullanım Sıklığı Kontrol Yöntemi
Sympson Hetter Yöntemi
Aşamalı Düşürme Stratejisi

Makale Hakkında

Gönderim Tarihi: 31.10.2018
Kabul Tarihi: 20.11.2019
Elektronik Yayın Tarihi: 04.04.2020

DOI: 10.15390/EB.2020.8256

* Bu makale Recep Gür'ün H. Deniz Gülleroğlu danışmanlığında yürüttüğü "Bireye uyarlanmış testlerde madde kullanım sıklığı kontrol yöntemlerinin farklı koşullarda ölçme duyarlılığına ve test güvenliğine etkisi" başlıklı doktora tezinden üretilmiştir.

¹ © Anadolu Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Türkiye, math.recepgur@gmail.com

² © Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Eğitim Bilimleri Bölümü, Türkiye, denizgulleroğlu@yahoo.com

kontrol parametrelerinin eş zamanlı olarak belirlendiği Aşamalı Düşürme yöntemi seçildiğinde, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak, test güvenliğinin daha iyi sağlandığı sonucuna ulaşılmıştır.

Giriş

Bireylerin psikolojik özelliklerini ve davranışlarını ölçmek amacıyla kullanılan veri toplama araçlarından elde edilen sonuçlar, onların yaşamında önemli bir yere sahiptir. Ölçme sonuçlarının, eğitim ve mesleki alanlara ilişkin önemli kararlarda kullanılabilmesi için, ölçme araçlarından elde edilen puanların geçerli ve güvenilir olması gerekmektedir. Eğitimde ve psikolojide birçok örtük özellik doğrudan gözlenemediğinden ilgili değişkenleri ölçme işleminin geçerli ve güvenilir olması için araştırmacılar tarafından iki ölçme kuramı geliştirilmiştir. Bunlar klasik test kuramı (KTK) ve madde tepki kuramı (MTK)'dır. Test geliştirme aşamasında, madde ve test istatistiklerini kestirirken bu kuramlardan yararlanılmaktadır (Lord, 1980).

MTK'nın uygulamada sağladığı avantajlardan biri bireye uyarlanmış (adaptive) test geliştirmeye imkân sunmasıdır. Binet ve Simon tarafından 1905 yılında geliştirilen Binet zekâ testi, bireye uyarlanmış testin tipik özelliklerini göstermekle birlikte ilk bireye uyarlanmış test uygulaması olarak kabul edilmektedir (Weiss, 1988). BOBUT uygulamaları, MTK'nın değişmezlik özelliğini kullanarak, her birey için nitelikli ölçme yapabilen maddelerin madde havuzundan seçilerek bireye sunulmasını sağlayan bir algoritma ile çalışmaktadır (Embretson ve Reise, 2000; Way, 2005).

BOBUT uygulamalarında, maddeler, bireyin yetenek (θ :theta) düzeyi ile eşleşecek şekilde seçilmektedir. Bu süreç çeşitli yöntemler kullanılarak farklı şekilde işleyebilmektedir. Genellikle ilk adımda, madde havuzundaki maddelerden orta derecede zorluğa sahip olan bir madde seçilerek bireyden bu maddeye yanıt vermesi beklenmektedir. Verilen yanıtta göre, bireyin θ düzeyine ilişkin kestirim elde edildikten sonra, madde havuzundaki hangi maddenin, bireyin θ kestirimi için en fazla bilgi sağlayacağına karar verilmektedir. BOBUT uygulamalarında madde seçiminin temel mantığı, bireyin doğru yanıtladığı maddeden sonra daha zor, yanlış yanıtladığı maddeden sonra ise daha kolay maddelerle karşılaşmasını sağlamaktır. Bireylerin her maddeye verdiği yanıttan sonra, yetenek düzeyleri tekrar hesaplanır ve yeni bir θ kestirimi yapılır. Farklı sonlandırma kuralları gerçekleştiğinde test sona ermektedir (Bulut ve Kan, 2012; Lord ve Stocking, 1988).

BOBUT uygulamalarından elde edilen puanlar bireyler hakkında verilen önemli kararlara yön verebilmektedir. Dolayısıyla BOBUT uygulamalarının geçerli ve güvenilir olarak gerçekleştirilmesi, alınan kararların doğruluğunu etkilemektedir. Bilgisayar, yazılım ve psikometri alanındaki gelişmelerle birlikte BOBUT uygulamalarının yaygınlaşmaya başlamasıyla BOBUT uygulamalarının geçerliliğini ve güvenilirliğini etkileyecek faktörlerden, test güvenliği (test security) ve ölçme duyarlılığı (measurement precision) gündeme gelmektedir (Boyd, Dodd ve Fitzpatrick, 2013; Weiss, 2004).

BOBUT uygulamalarında, test güvenliğini ve ölçme duyarlılığını olumsuz yönde etkileyecek durumlara ilişkin, madde kullanım sıklığı kontrolü (item exposure control) ile BOBUT algoritmasında kısıtlamalar yapılabilmektedir (Han, 2009). Madde havuzunun kalibrasyonu, başlatma ve sonlandırma kuralları, madde seçme ve yetenek kestirim yöntemleri, içerik dengelenmesinin (content balancing) yanı sıra madde kullanım sıklığı kontrolünün de BOBUT'un temel bileşenlerinden biri olduğu vurgulanmaktadır (Boyd, 2003; Segall, 2004; Magis ve Raîche, 2012).

BOBUT uygulamalarında, bireyler yetenek düzeylerine en uygun maddelerle karşılaştıkları için, özellikle geniş ölçekli testlerde, bir bireyin istediği sayıda ve istediği sıklıkta testi almasına izin verilebilmektedir. Bu tür uygulamalarda, bir sınava birden çok kez katılan bireye, aynı maddeler çok sık uygulandığında (over exposure) maddelerin ifşâ edilmesi maddelerin psikometrik özelliklerinin yapay olarak düşmesine neden olmaktadır (Segall, 2004; Revuelta ve Ponsoda, 1998). Bu durum ölçme sonuçlarının geçerliliğini ve test güvenliğini olumsuz yönde etkilemektedir (Lee ve Dodd, 2012).

Madde bilgi fonksiyonu ile yetenek kestiriminin standart hatası arasında ters yönde ilişki bulunmaktadır. Madde bilgi fonksiyonu, BOBUT uygulamalarında madde seçimi için ölçüt alındığından yüksek bilgi değerine sahip maddelerin seçilmesi, yetenek kestiriminin standart hatasını düşürerek ölçme duyarlılığını artırmaktadır (Kalender, 2009). Fakat ölçme duyarlılığını maksimum yapacak maddeler seçildiğinde, kullanım sıklığı oranları düzgün dağılmayacağından madde havuzundaki bazı maddeler çok sık uygulanırken, bazı maddeler ise hiçbir bireye uygulanmamaktadır (Pastor, Dodd ve Chang, 2002). Bu durum, madde havuzunun dengeli bir şekilde kullanılmayıp sadece sınırlı sayıda bazı maddelerin kullanılmasına sebep olmaktadır (Han, 2009). Bu nedenle, madde kullanım sıklığı kontrol yöntemleriyle, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin sağlanması hedeflenmektedir (Boyd, 2003; Davis ve Dodd, 2005; Pastord., 2002).

Bireysel farklılıkları gözetererek, bireye kendi yetenek düzeyine en uygun maddelerin seçilmesiyle gerçekleştirilen BOBUT uygulamaları için geniş madde havuzuna ihtiyaç duyulmaktadır (Embretson ve Reise, 2000; Magis ve Raîche, 2012). Çok sayıda yüksek derecede ayırt ediciliğe sahip, her bir yetenek düzeyine hitap edebilecek güçlük düzeyindeki maddelerden oluşan madde havuzlarıyla BOBUT uygulamaları daha iyi sonuçlar vermektedir (Veldkamp ve Van Der Linden, 2010; Weiss, 2004). Dolayısıyla madde havuzunun dengeli bir şekilde dağılım göstermeyip sadece sınırlı sayıda bazı maddelerin kullanılması, madde havuzu geliştirmek için harcanan zaman ve işgücünün iyi değerlendirilmemesine sebep olmaktadır (Aytuğ Koşan, 2013). Belirtilen bu nedenlerle test güvenliğini sağlayabilmek, madde havuzu kullanımını daha verimli hale getirebilmek ve madde havuzunun işlevsel olarak sürekliliğini sağlayabilmek için madde kullanım sıklığı kontrol yöntemleri geliştirilmiştir (Davis, 2002; Boyd, 2003; Revuelta ve Ponsoda, 1998).

Madde kullanım sıklığı kontrol problemine ilişkin geliştirilen ilk yöntemlerden biri, 5-4-3-2-1 stratejisidir. Tesadüfi seçme yöntemlerinden 5-4-3-2-1 stratejisi, BOBUT iterasyonundaki ilk adımda beş madde arasından, ikinci adımda dört madde arasından şeklinde, tesadüfi olarak seçilen bir maddenin uygulanma sıklığını kontrol etmeyi amaçlamaktadır. Kingsbury ve Zara (1989) ile Thomasson (1998) ise bütün maddelerin kullanım sıklık oranını azaltmak için farklı tesadüfi seçme yöntemleri geliştirmişlerdir (aktaran Veldkamp, Vershoor ve Eggen, 2010). Önceden belirlenmiş en çok bilgi veren on maddeden birinin rastgele seçildiği Randlesque yöntemi; hedeflenen madde güçlük düzeyinin .10 logit aralığı ranjından rastgele maddelerin seçildiği .10 logit aralığı (within .10 logit) yöntemi (Lunz ve Stahl, 1998), test ilerledikçe madde seçme üzerindeki tesadüfi bir bileşenin etkisini azaltarak madde bilgisinin öneminin belirgin bir şekilde artırılmasının amaçlandığı İlerlemeli yöntem [Progressive Method-(İY)] tesadüfi seçme yöntemlerindedir. Tesadüfi seçme yöntemlerinde ortak amaç, maksimum bilginin en ideal olduğu aralıktaki maddeler arasından rastgele seçilen bir maddenin uygulanma sıklığını kontrol etmektir (Georgiadou, Triantafillou ve Economide, 2007).

Koşullu seçme yöntemlerinde, BOBUT uygulamasından önce, tekrarlı simülasyonlarla belirlenen kullanım sıklığı kontrol parametresi kullanılarak, maddenin kullanım sıklık oranı kontrol edilmektedir. Davey ve Parshall yöntemi, Stocking ve Lewis çok terimli (mutinomial) yöntemi, hedeflenen (targeted) kullanım sıklığı kontrol yöntemi, Sınırlandırılmış Maksimum Bilgi (Restricted Maximum Information-[SMB]) yöntemi gibi koşullu seçme yöntemleri bulunmaktadır. Davey ve Parshall (1995) yönteminde tek tek maddelerin aşırı kullanımını kısıtlamakla kalmayıp aynı zamanda bireylerin aynı madde setleriyle karşılaşması engellenmeye çalışılırken; Stocking ve Lewis çok terimli yönteminde, her bir madde için her bir yetenek seviyesi sayısı (n) kadar madde kullanım sıklığı kontrol parametresi hesaplanmaktadır (Stocking ve Lewis, 1995). Hedeflenen kullanım sıklığı kontrol yöntemi, maddelerin aşırı kullanımını kontrol etmeye odaklanmak yerine kullanılmayan maddelerin uygulanma olasılıklarını artırmaya yöneliktir (Thompson, 2002). Koşullu seçme yöntemlerinden bir diğeri SMB yönteminde ise, hiçbir maddenin önceden belirlenmiş kullanım sıklık oranından daha fazlasının kullanılmasına izin verilmemektedir (Revuelta ve Ponsoda, 1998).

Dönüşümlü (rotating) madde havuzu yönteminde (Ariel, Veldkamp ve Van Der Linden, 2004; Way, 1998) maddelerin kullanım sıklık oranını azaltmaya yönelik madde havuzunu içerik ve madde

parametreleri bakımından benzer dağılım gösterecek şekilde, önsel dağılım kullanılarak maddeler farklı testlerde uygulanmaktadır. Bu yöntemlerin yanı sıra, madde havuzunun tabakalandırılarak kullanılmasının amaçlandığı tabakalı yöntemler (a-tabakalama stratejisi (Chang ve Ying, 1999), b parametresi blokeli a tabakalama (Chang, Qian ve Ying, 2001), içerik blokeli a tabakalamalı BOBUT deseni (Yi ve Chang, 2001), 0-1 tabakalamalı strateji (Chang ve Van der Linden, 2003) ve farklı yöntemler bir araya getirilerek birleştirilmiş yöntemler (İllemeli sınırlandırılmış strateji (Revuelta ve Ponsoda, 1998); a tabakalama ile SH stratejisinin birleşimi (Leung, Chang ve Hau, 2003); içerik blokeli (Yi, 2002) a tabakalamalı ile SH stratejisinin birleşimi geliştirilmiştir (aktaran Georgiadou vd., 2007). Farklı yöntemler bulunmasına rağmen, BOBUT uygulamalarında en yaygın kullanılan madde kullanım sıklığı kontrol yöntemi, koşullu seçme yöntemlerinden Sympson Hetter (SH) yöntemidir (Veldkamp vd., 2010).

BOBUT uygulaması öncesinde hedeflenen değer olarak maksimum kullanım sıklık oranı belirlenip madde seçildiğinde, bireyin o maddeyle karşılaşması, maddenin kullanım sıklığı kontrol parametresine bağlı olmaktadır (Davis ve Dodd, 2005). Örneğin, bütün maddeler için bu parametre .25 olarak tanımlanırsa, bir maddenin seçilmesi durumunda, yaklaşık olarak her dört bilgisayar ortamında bireye uyarlanmış testlerden birinde, ilgili madde uygulanabilmektedir (Weiss ve Guyer, 2012).

BOBUT uygulamalarında çok fazla seçilen i. maddenin kullanım sıklık oranının ($P_i(S)$), BOBUT uygulaması öncesinde belirlenen madde kullanım sıklığı kontrol parametresinden ($P_i(A|S)$) büyük olması ($P_i(A|S) < P_i(S)$), i. maddenin seçilmesi durumunda uygulanma olasılığını ($P_i(A)$) kısıtlarken; daha az kullanılan maddeler için madde kullanım sıklık oranı, madde kullanım sıklığı kontrol parametresinden daha küçük ($P_i(S) < P_i(A|S)$) olduğunda ilgili maddelerin uygulanma olasılığını artırmaktadır (Segall, 2004; Stocking ve Lewis, 2002). Böylelikle uygulanacak maddenin koşullu olasılığı $P_i(A|S)$ kontrol parametresi olarak tanımlanıp $P_i(A) = P_i(A|S) * P_i(S)$ eşitliğinde kullanılması, hedeflenen kullanım sıklığı oranına ulaşılabilme, seçilen maddenin uygulanıp uygulanmadığını kontrol etme ya da birden fazla BOBUT uygulamasına katılan bireyler için seçilen maddenin geçici olarak kullanılmamasına imkân vermektedir (Pastor vd., 2002; Veldkamp vd., 2010).

Han (2009) geliştirdiği yöntemle, kullanım sıklığı kontrol parametrelerinin belirlenmesinde, BOBUT uygulamasından önce, tekrarlı simülasyon yapılmasına gerek olmadığını, BOBUT uygulanma sürecinde eş zamanlı olarak kontrol parametrelerinin belirlenebileceğini vurgulamaktadır. Madde seçme sürecinde, havuzdan seçilmesi uygun olan her bir madde için madde seçme ölçütü, hedeflenen kullanım sıklık oranı (e) ve hali hazırdaki/gözlenen kullanılma sıklık oranı (r_i) arasındaki oran ile ters bir şekilde $I_i [\hat{\theta}_{m-1}] \frac{e-r_i}{e}$ ağırlıklandırılmaktadır. Böylelikle $e > r_i$ ise, madde seçme ölçütü Maksimum Fisher Bilgisi (MFB) $I_i [\hat{\theta}_{m-1}]$ pozitif iken; $r_i > e$ ise, $I_i [\hat{\theta}_{m-1}]$ negatif olmaktadır (Han, 2012). Bu durumda, az kullanılan maddeler daha sık kullanılırken, fazla kullanılan maddelerde ise, seçilme oranı aşamalı düşürüleceğinden, Han tarafından bu yöntem Aşamalı Düşürme (AD) Yöntemi (fade-away method) olarak adlandırılmaktadır.

BOBUT uygulamalarıyla ilgili yapılan çalışmalarda, genellikle geleneksel kâğıt kalem testi ile farklı stratejiler (farklı başlatma- sonlandırma kuralları, madde seçme ve yetenek kestirim yöntemleri) kullanılarak yapılan BOBUT uygulamalarıyla kestirilen yetenek düzeyleri arasında manidar bir ilişki bulunup bulunmadığı incelenmiştir (Bulut ve Kan, 2012; Cömert, 2008; Eroğlu, 2013; Gökçe, 2012; İşeri, 2002; Kalender, 2011; Kaptan, 1993; Kaskatı, 2011; Kezer, 2013; McDonald, 2002; Özbaşı, 2014; Öztuna, 2008; Scullard, 2007; Smits, Cuijper ve Straten, 2011; Sulak, 2013; Wang, Kuo, Tsai ve Laio, 2012; Zitny, Halama, Jelinek ve Kveton, 2012). İlgili alanyazında, BOBUT uygulaması sonucunda, kâğıt kalem testlerine benzer yetenek kestirimleri yapıldığı ve farklı BOBUT stratejileri kullanılarak yetenek düzeyleri kestirildiğinde benzer sonuçlar alınabileceği bulgusuna ulaşılmıştır. Ayrıca BOBUT uygulamalarının, kâğıt kalem testlerine göre ölçme duyarlılığını artırdığı ve yetenek kestirimi için gerekli madde sayıları dolayısıyla zaman bakımından da önemli ölçüde ekonomiklik sağladığı sonucuna ulaşılmıştır.

BOBUT'un temel bileşenlerinden, madde kullanım sıklığı kontrol yöntemleriyle ilgili yapılan çalışmalar incelendiğinde, çoklu puanlanan MTK modellerine (Burt, Kim, Davis ve Dodd, 2003; Davis, 2002, 2004; Davis ve Dodd, 2005); farklı madde seçme yöntemlerine (BoztunçÖztürk, 2014; Han, 2009, 2012); farklı madde ayırt edicilik dağılımlarına sahip madde havuzlarına (Revuelta ve Ponsoda, 1998), farklı ortalama güçlük düzeylerine sahip madde havuzlarına (BoztunçÖztürk, 2014; Lee ve Dodd, 2012), farklı madde havuzu büyüklüklerine (Chang ve Twu, 1998; Pastor vd., 2002; Revuelta ve Ponsoda, 1998); farklı test sonlandırma kurallarına (French ve Thompson, 2003; Revuelta ve Ponsoda, 1998) ve testlet (madde takımları) temelli BOBUT uygulamalarında farklı yetenek kestirimlerine (Boyd, 2003; Davis ve Dodd, 2003) göre madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı görülmüştür. İlgili araştırmalarda, madde kullanım sıklık yöntemlerinin farklı madde havuzu büyüklüklerinden etkilendiği, orta güçlükteki maddelerden oluşan madde havuzlarında kullanılmayan madde sayısının daha az ve ölçme duyarlılığının daha iyi olduğu, normal dağılıma sahip örneklemelerde madde havuzunun güçlük düzeyi arttıkça kullanılmayan madde sayısının arttığı, 1-0 şeklinde puanlanan maddelerde koşullu seçme yöntemlerinden Sympson Hetter stratejisinin, tesadüfi seçme yöntemlerine göre daha etkili olduğu, madde seçme yöntemlerinden a-tabakalama madde seçme yöntemi ve Aşamalı Düşürme madde kullanım sıklığı kontrol yöntemlerinin birlikte kullanıldığı durumlarda test güvenliğinin arttığı sonucuna ulaşılmıştır.

Rudman (1987), BOBUT uygulamalarının 21. yüzyılın ölçme yöntemi olarak nitelendirilebileceğini savunarak, BOBUT stratejilerine yönelik yapılacak çalışmaların önemine dikkat çekmektedir. Psikometrik yönden, BOBUT uygulamalarının iki önemli avantajı bulunmaktadır. Bunlardan biri ölçme duyarlılığını arttırması, diğeri ise güvenli bir test ortamının sağlanmasıdır (Weiss, 2004).

BOBUT uygulamalarında, bireyler yetenek düzeylerine en uygun maddelerle karşılaştıkları için, özellikle geniş ölçekli testlerde, bir bireyin istediği sayıda ve istediği sıklıkta testi almasına izin verilebilmektedir. Bu şekilde BOBUT uygulamalarının sürekliliğinin olması, her ne kadar sınav takvimini planlama, test uygulamaları için test merkezlerinde, okullarda vs. yeterli mekân ve bilgisayar sağlanması bakımından avantajlar sağlasa da dezavantajları da bulunmaktadır. Bu tür uygulamalarla, bir sınava birden çok kez katılan bireyin önceden karşılaştığı maddeleri hatırlaması, internet aracılığıyla sosyal medyada ya da arkadaşlarıyla çok sık uygulanan maddeleri paylaşması riskini ortaya çıkarmaktadır. Ayrıca BOBUT uygulamalarında, bireylerin ilgili yetenek düzeyleri ile yüksek bilgi değerine sahip maddelerin bilgisayar ortamında eşleştirilmesi, madde havuzunun dengeli bir şekilde kullanılmayıp sınırlı sayıda bazı maddelerin kullanılması, sık kullanılan maddelerin işâ edilmesiyle psikometrik özelliklerinin düşmesi ve madde havuzunun işlevselliğini kaybetmesi gibi sorunları da beraberinde getirmektedir. Bu tür etkenler, BOBUT uygulamalarının geçerliğini ve güvenilirliğini etkileyecek faktörlerden, test güvenliği ve ölçme duyarlılığına ilişkin tartışmalara yol açmıştır. Bu doğrultuda BOBUT'un temel bileşenleri arasında, madde kullanım sıklığı kontrolü yöntemleri de yer almaya başlamıştır.

BOBUT uygulamalarında bir sınava birden çok kez katılan bireyin aynı maddelerle karşılaşması durumunda, önceden pratik yapacak olması, Yen (1993)'e göre, madde yerel bağımlılığına neden olan faktörler arasında gösterilmektedir. MTK'ya dayalı ölçekleme yapılan BOBUT uygulamalarında, MTK'nın yerel bağımsızlık varsayımının ihlal edilmesi, güvenilirliği şişirmekte, yetenek ve madde parametrelerinin kestirimine ilişkin standart hataları etkilemekte ve buna bağlı olarak, yetenek ve madde parametrelerinin hatalı kestirilmesine sebep olmaktadır (Demars, 2006). Madde kullanım sıklığı kontrol yöntemlerine yönelik yapılacak çalışmaların bu tür sorunlara da çözüm getireceği düşünülmektedir.

MTK modellerinin uygulanmasında ve yorumlanmasında test uzunluğu, örneklem büyüklüğü ve yetenek dağılımı önemli bir etkiye sahiptir. Dolayısıyla farklı ölçme koşullarında (örneğin küçük örneklem - kısa test ve sola çarpık yetenek dağılımlarında) hangi madde kullanım sıklığı kontrol yönteminin daha işlevsel olduğunun belirlenmesi, hem test güvenliğinin artmasına hem de bireylerin yeteneklerinin daha az hata ile kestirilmesine katkı sağlamaktadır. Bu bağlamda, farklı örneklem büyüklüklerine, test uzunluklarına ve yetenek dağılımlarına göre madde kullanım sıklığı kontrol yöntemleri karşılaştırıldığında, ölçme duyarlılığı ve test güvenliği indekslerinin nasıl değiştiğinin incelenmesi bu araştırmanın problemini oluşturmaktadır.

Yöntem

Bu bölümde, araştırmanın modeli, verilerin üretilmesi, BOBUT koşulları ve verilerin analizi ile ilgili bilgiler yer almaktadır.

Araştırmanın Modeli

Ampirik veri bulmanın zor olduğu çalışmalarda, yöntemler arası bağıntıların anlaşılmasına yönelik simülasyon çalışmaları yapılmaktadır. Bu tür simülasyon çalışmaları, kuramın gelişmesine önemli katkılar sağlamaktadır (Davis, Eisenhardt ve Bingham, 2007). Bu çalışmada; monte-carlo simülatif BOBUT uygulamalarında 1-0 şeklinde puanlanan iki kategorili maddelerde madde kullanım sıklığı kontrol yöntemleri, farklı ölçme koşullarına göre karşılaştırıldığında ölçme duyarlılığı ve test güvenliği indekslerinin nasıl değiştiğinin incelenmesi amaçlandığı için temel araştırma olarak değerlendirilebilir.

Verilerin Üretilmesi

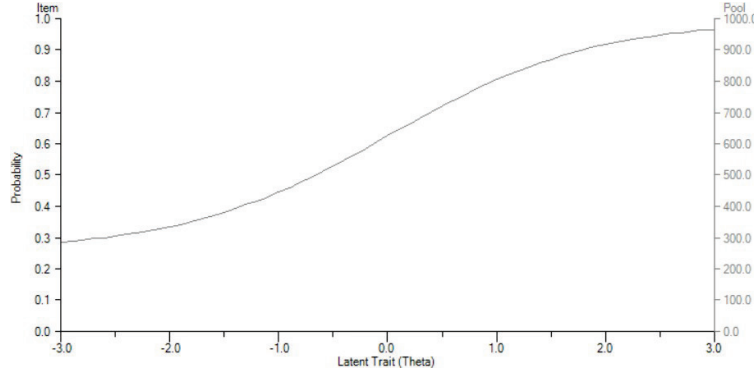
BOBUT uygulamasında MTK'ya göre ölçeklenmiş maddelerden oluşan geniş madde havuzunun yanı sıra BOBUT uygulamalarında en az hata ile kestirimler elde edilebilmesi için çok fazla sayıda katılımcının yanıt örüntülerine ihtiyaç duyulmaktadır. Bunun yanı sıra, araştırmanın amacı doğrultusunda farklı ölçme koşullarını sağlayabilecek veri setlerine ulaşmak pratik anlamda neredeyse imkânsız olduğundan simülatif veriler kullanılmıştır. Monte-carlo simülasyon çalışmaları, farklı stratejilerde veri setlerini çeşitlendirerek, BOBUT uygulamalarına ilişkin etkin ve hızlı bir şekilde karşılaştırma imkânı sunmaktadır (Weiss ve Guyer, 2012). Simülatif verilerin üretilmesi için SimulCAT (Han, 2011) yazılımından faydalanılmıştır.

MTK modellerinin uygulanmasında ve yorumlanmasında test uzunluğu ya da örneklem büyüklüğü önemli bir etkiye sahiptir. Şahin ve Anıl'a (2017) göre, tek boyutlu iki kategorili (1-0) puanlanan MTK modelleri yorumlanırken, test uzunluğunun ve örneklem büyüklüğünün birlikte ele alınıp incelenmesinin, madde parametreleri kestirimi açısından önemli bir etki oluşturduğu vurgulanmaktadır. Alanyazında test uzunluğunun ve örneklem büyüklüğünün dağılımına ilişkin farklılıklar bulunmakla birlikte genellikle kısa test için 25 madde (Demars, 2006; Guyer ve Thompson, 2011; Harwell, Stone, Hsu ve Kirişçi, 1996; Weiss ve Von Minden, 2012; Yoes, 1995) uzun test için 50 madde (Demars, 2006; Glas, 2002; Weiss ve Von Minden, 2012); küçük örneklemin 250 birey (Çetin, 2009; Goldman ve Raju, 1986; Harwell ve Janosky, 1991; Speron, 2009; Vaughn ve Wang, 2010; Yoes, 1995), geniş örneklemin ise 1000 birey (Çetin, 2009; Glas, 2002; Goldman ve Raju, 1986; Guyer ve Thompson, 2012; Hulin, Lissak ve Drasgow, 1982; Köse, 2010; Lord, 1968; Patsula ve Gessaroli, 1995; Speron, 2009; Tang, Way ve Carey, 1993; Thissen ve Wainer, 1982; Vaughn ve Wang, 2010; Yen, 1987; Yoes, 1995; Weiss ve Von Minden, 2012) olduğu vurgulanmaktadır. İlgili alanyazın bilgisi doğrultusunda, bu çalışmada, kısa test 25 madde, uzun test 50 madde; küçük örneklem 250 birey, geniş örneklem 1000 birey olarak belirlenmiştir.

Testi alan bireylerin yetenek parametreleri her bir örneklem büyüklüğü koşulu için (θ) değeri -3 ve +3 aralığında i) normal dağılım $\theta \sim N(0, 1)$; Weiss ve Guyer (2012)'e göre ii) sağa çarpık dağılım β (1, 4); iii) sola çarpık dağılım β (4, 1); iv) tek biçimli (uniform) dağılım β (1, 1) gösterecek şekilde üretilerek BOBUT örneklemi oluşturulmuştur. Bunun yanı sıra, tek biçimli dağılımda ortalama ve standart sapma değerleri için Agresti ve Coull'a (1998) göre, $U(-3,3)$ için $\bar{x} \cong .00$, $SS = 6/\sqrt{12} \cong 1.732$ olacak şekilde BOBUT örneklemi elde edilmiştir.

Madde kullanım sıklığı kontrol yöntemlerinden, SH stratejisinde BOBUT uygulamasından önce, tekrarlı simülasyonlarla kullanım sıklığı kontrol parametrelerinin belirlenmesi gerekmektedir (Veldkamp vd., 2010). Bunun için her bir BOBUT örnekleminde yer alan bireylerin yetenek dağılımları ile benzer dağılıma sahip olacak şekilde Sympson Hetter örneklemini için örneklem büyüklüğü 6000 ve replikasyon sayısı 5 alınarak madde kullanım sıklığı kontrol parametreleri hesaplanmıştır (Gu ve Reckase, 2007). Eğer her bir maddenin beş replikasyon sonrasında ortalama seçilme oranı ($P_i(S)$), maksimum hedeflenen kullanım sıklık oranından ($r=.20$) büyükse, madde kullanım sıklığı kontrol parametresi $k_i=r/P_i(S)$ olarak hesaplanmakta; ($P_i(S)$)< r ise $k_i=1$ alınarak ilgili maddelerin uygulanma olasılığı artırılmaktadır (Eggen, 2001; Veldkamp ve Van Der Linden, 2008).

BOBUT uygulamalarında maddeler her bir yetenek düzeyine hitap edecek şekilde yeterli sayıda dağılım gösteren ve ayırt ediciliği yüksek maddelerden oluşan havuzla çalışıldığı takdirde daha iyi sonuçlar vermektedir. Bu bilgi doğrultusunda, Weiss ve Von Minden (2012)'e göre a (.25 ile 1.75), b (-3 ile +3), c ise (.20 ile .30) ranj aralığında, ayırt edicilik ve şans parametreleri tek biçimli dağılım, güçlük parametresi ise normal dağılım gösteren 1000 maddeden oluşan simülatif madde havuzu elde edilmiştir. Ayrıca tek boyutlu 1-0 şeklinde puanlanan çoktan seçmeli maddelerde, şans başarısı parametresini de dikkate alan bir model olduğu için 3PL modele göre kestirimler yapılmıştır (Crocker ve Algina, 1986; Hambleton, Swaminathan ve Rogers, 1991). Madde havuzuna ilişkin yanıt fonksiyonuna aşağıda yer verilmiştir.



Şekil 1. Madde Havuzuna İlişkin Yanıt Fonksiyonu

BOBUT Koşulları

Araştırma kapsamında simülatif BOBUT uygulaması için, madde seçme yöntemlerinden en yaygın kullanılan Maksimum Fisher Bilgisi yöntemi ve başlatma kuralı stratejilerinden $-.50 < b < .50$ aralığı stratejisi kullanılmıştır (Weiss, 1988). Yetenek kestirim yöntemi olarak maddelere tümüyle doğru ya da tümüyle yanlış yanıt veren bireyler bulunabileceği için en yüksek olabilirlik kestirim yöntemi yerine önsel dağılım kullanarak kestirim yapan beklenen sonsal dağılım (EAP) yöntemi [Expected A Posteriori] tercih edilmiştir (Embretson ve Reise, 2000).

Madde kullanım sıklığı kontrol yöntemlerinden, BOBUT uygulamalarında en yaygın kullanılan Sympson Hetter stratejisi (Veldkamp vd., 2010) ve BOBUT uygulamasından önce, tekrarlı simülasyonlar gerektirmeyen, BOBUT uygulanma sürecinde kullanım sıklığı kontrol parametrelerinin eş zamanlı olarak belirlendiği Aşamalı Düşürme Stratejisi (Han, 2009) seçilmiştir. Sonlandırma kuralı olarak sabit test uzunluğu (kısa test:25 madde; uzun test:50 madde) tanımlanmıştır. Böylelikle ilgili değişkenler ve koşullar çaprazlandığında $2*4*3*2=48$ simülasyon koşulu belirlenmiştir. Ayrıca örneklem yanlılığı oluşmaması için (Harwell, 1996, aktaran Evans, 2010) her bir koşulda 100 tekrar (Han, 2009) kullanılarak 4800 veri dosyası oluşturulmuştur.

Verilerin Analizi

Araştırma kapsamında her bir koşulda ölçme duyarlılığı ve test güvenliği indeksleri hesaplanıp birbirleriyle karşılaştırılmıştır.

Farklı Ölçme Koşullarında Ölçme Duyarlılığına İlişkin Veri Çözümleme Yöntemleri

Ölçme duyarlılığı arttığında, hata değerleri azalacağı için, araştırmada ölçme duyarlılığının belirlenmesi amacıyla, her bir koşulda hata göstergeleri olan uyum (fidelity), yanlılık (bias) ve Hataların Ortalama Karekökü (Root Mean Squared Error [RMSE]) katsayıları hesaplanmıştır (French ve Thompson, 2003).

Uyum katsayısı, BOBUT uygulaması ile kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasındaki korelasyon katsayısıdır. Uyum katsayısını hesaplamak için Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılmıştır. BOBUT uygulamasında kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki ortalama anlamlı farklılığa ilişkin yanlılık ve mutlak farklılık

için RMSE değerleri hesaplanmıştır (Gu ve Reckase, 2007; Leroux, Lopez, Hembry ve Dodd, 2013; Wang ve Vispoel, 1998; Zheng ve Chang, 2014).

Farklı ölçme koşullarında ölçme duyarlılığına ilişkin uyum değerleri arasında manidar bir farklılık bulunup bulunmadığı, iki korelasyon katsayısı arasındaki farkın karşılaştırılmasına imkân veren Fisher'ın z testi ile incelenmiştir (Howell, 2010; Şencan, 2005). İki korelasyon katsayısı arasındaki farka ilişkin etki büyüklüğünü hesaplarken ise Cohen q (Cohen $q = r_1 - r_2$) katsayısı kullanılmıştır. Cohen q katsayısının .10'dan küçük olması etkinin olmadığı şeklinde yorumlanırken; .10 ile .30 arası küçük, .30 ile .50 arası orta düzeyde ve .50'den büyük olması ise geniş etki büyüklüğü şeklinde yorumlanmaktadır (Cohen, 1988).

Farklı değişkenlere (farklı ölçme koşullarında madde kullanım sıklığı yöntemlerinin yanlılık değerlerine) ait değişimlemelerin gruplandırılmasına vekümler arasındaki uzaklığın hesaplanmasına imkân sunduğundan, örüntüleri bilinen sonuçların karşılaştırmasında ve aşırı gözlemleri test etmede Mahalanobis uzaklıkları kullanılabilir (Hair, Anderson, Tatham ve Black, 1998; Pallant, 2010). Dolayısıyla farklı ölçme koşullarında ölçme duyarlılığına ilişkin yanlılık / RMSE değerleri arasında manidar bir fark bulunup bulunmadığını incelemek için Mahalanobis uzaklıkları hesaplanmıştır.

Hataların ortalama karekökü (RMSE), gerçek yetenek düzeyleri arasındaki kovaryans ile kestirilen yetenek düzeylerine ilişkin kovaryans matrisi arasındaki farkın bir başka ifadeyle hatanın derecesi doğrultusunda geliştirilmiş bir mutlak uyum indeksidir (Sümer, 2000). RMSE, modelden kestirilen yetenek parametrelerinin, evren kovaryansları (gerçek yetenek parametreleri) ile ne derece uyumlu olduğunu göstermekte olup, .00 ile 1.00 arasında değer almaktadır (Byrne, 1998). Modelden kestirilen yetenek parametrelerinin, gerçek yetenek parametrelerine ilişkin kovaryanslar arasındaki farklılık sifıra yaklaştıkça, modelin uyumlu olduğu söylenebilmektedir (Byrne, 1998; Kline, 2005). RMSE değerinin .05'den küçük olması mükemmel uyum (Brown, 2006; Jöreskog ve Sörbom, 1993); .05 ile .08 arasında olması kabul edilebilir uyum (Hooper, Coughlan ve Mullen, 2008; Howell, 2010; Tabachnick ve Fidell, 2007) ve .10'dan küçük olması zayıf uyum (Kelloway, 1998) olarak değerlendirilmektedir. Eğer örneklem küçükse ve modelde hesaplanan parametre sayısı fazla ise RMSE değeri .10'dan büyük çıkabilmektedir (Şimşek, 2007).

Farklı Ölçme Koşullarında Test Güvenliğine İlişkin Veri Çözümleme Yöntemleri

Her bir koşulda, kullanım sıklığı oranı, madde havuzundan yararlanmama (pool utilization) ve test örtüşme (test overlap) indeksi birlikte değerlendirilerek, test güvenliği hakkında yorum yapılabilmektedir (Davis ve Dodd, 2005). Madde havuzundan yararlanmama (pool utilization) ile BOBUT sürecinde uygulanmayan maddelerin yüzdesi kastedilmektedir (Leroux vd., 2013). Madde kullanım sıklığı oranlarının (r_i : maddelerin kullanılma sayısının örneklem büyüklüğüne oranlarının) dağılımı, madde kullanım sıklığı oranlarının standart sapması ve maksimum madde kullanım sıklığı oranı hesaplanarak her bir koşulda *kullanım sıklığı oranları* belirlenmektedir. Ayrıca uygulanmayan madde sayısı ve oranı hesaplanarak elde edilen *madde havuzundan yararlanmama (pool utilization) indeksi*, madde havuzunun kullanılma derecesinin belirlenmesine imkân sunmaktadır.

Test örtüşme (test overlap) indeksinde her bir koşulda, tesadüfi olarak seçilen iki birey için örtüşen maddelerin sayısı/karşılaştıkları aynı maddelerin sayısı dikkate alınmaktadır. Test örtüşme indeksi aşağıdaki eşitlikler yardımıyla hesaplanmıştır (Chen, Ankenmann ve Spray, 1999):

$$\sum_{i=1}^n \binom{m_i}{2}$$

BOBUT sürecinde m_i : uygulanan i. maddenin kullanılma sayısı n : madde havuzu genişliği olmak üzere, eğer $m_i < 2$ ise $\binom{m_i}{2} = 0$ olmaktadır. Başka bir ifadeyle, BOBUT uygulamalarında, herhangi bir madde, yalnızca bir kez uygulandıysa, test örtüşme indeksini etkilememektedir. Genel olarak testler arası örtüşme indeksinin ortalaması aşağıdaki formülle hesaplanmıştır.

$$\bar{T} = \frac{\sum_{i=1}^n \binom{m_i}{2}}{k \binom{N}{2}}$$

N :BOBUT uygulamasına katılan bireylerin sayısı ve k :BOBUT uygulamasındaki sabit test uzunluğunu göstermektedir. Test örtüşme indeksi ortalamasının büyük olması, test güvenliğinin zedelendiği anlamına gelmektedir (Huang, Chen ve Wang, 2012).

Araştırmanın sınırlılıkları doğrultusunda en ideal madde kullanım sıklık yöntemine karar verilmesi için F istatistiği sonucu kullanılmıştır (Chang vd., 2001; Grubbs, 1973). Chang ve Ying (1999, s. 215)'a göre, "Madde kullanım sıklık oranları karşılaştırılırken, χ^2 istatistikleri hesaplanarak, $\chi^2_{yöntem1}$ ve $\chi^2_{yöntem2}$ karşılaştırılması" gerektiğini belirtmektedir.

Karşılaştırma ölçüsü olarak $(F_{\chi^2_{yöntem1}, \chi^2_{yöntem2}} = \frac{\chi^2_{yöntem1}}{\chi^2_{yöntem2}})$ istatistiği tanımlanmıştır. Eğer $F_{\chi^2_{yöntem1}, \chi^2_{yöntem2}} < 1$ ise, madde kullanım sıklığı oranlarının genel olarak dengelenmesi bakımından birinci yöntemin, ikinci yönteme göre daha iyi bir yöntem olduğu şeklinde yorumlanmaktadır." χ^2 istatistiği aşağıdaki Eşitlik 1 yardımıyla hesaplanmıştır (Tay, 2015):

$$\chi^2 = \frac{\sum_{i=1}^n (r_i - uni(r_i))^2}{uni(r_i)} \dots \dots \dots \text{(Eşitlik 1)}$$

Madde kullanım sıklık oranlarının (r_i), bütün maddeler için istenilen tek biçimli madde kullanım sıklık oranına (test uzunluğunun madde havuzu büyüklüğüne oranına ($uni(r_i) = k/n$)) farkının kareleri toplamı, istenilen tek biçimli madde kullanım sıklık oranına bölünerek χ^2 değeri hesaplanmıştır.

Farklı ölçme koşullarında test güvenliğine ilişkin madde kullanım sıklığı oranlarının standart sapma değerleri arasında manidar bir fark bulunup bulunmadığı "*İki varyansa ilişkin hipotez testi*" ile incelenmiştir. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin maksimum madde kullanım sıklığı oranları; madde havuzundan yararlanmama oranları; test örtüşme indeksleri arasında manidar bir fark bulunup bulunmadığı ise "*İki oran arasındaki farkın testi*" ile incelenmiştir.

Her bir koşulda 100 tekrar sonucunda elde edilen değerlerin ortalamaları (Leroux vd., 2013; Ross, 2013; Zheng ve Chang, 2014) dikkate alınarak ölçme duyarlılığı için uyum, yanlılık ve RMSE katsayıları; test güvenliği için, maksimum madde kullanım sıklığı oranı, madde kullanım sıklığı oranının standart sapması, madde havuzundan yararlanmama ve test örtüşme indeksi; madde kullanım sıklık oranlarını karşılaştırmak için χ^2 istatistikleri; en ideal madde kullanım sıklık yöntemine karar verilmesi için ise F istatistiğine ilişkin değerler Excel'de ayrı ayrı hesaplanmıştır. Yanlılık ve RMSE değerlerinin Mahalanobis uzaklıklarının hesaplanması için R yazılım programında yer alan {stats} paketindeki "mahalanobis" komutu kullanılmıştır (R Core Team, 2017).

Bulgular

Bu bölümde, araştırmanın amacı çerçevesinde cevap aranan araştırma sorularına ilişkin bulgular ve yorumlar yer almaktadır.

BOBUT Uygulamalarında Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Ölçme Duyarlılığına Etkisi

BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin ölçme duyarlılığına etkisini incelemek amacıyla aşağıda sırasıyla ölçme duyarlılığı indekslerinden uyum katsayılarına, yanlılık değerlerine ve RMSE değerlerine ilişkin bulgular ve yorumlara yer verilmiştir. Bu doğrultuda, BOBUT uygulamalarında madde kullanım sıklığı kontrol yöntemlerinin farklı ölçme koşullarında ölçme duyarlılığına etkisi Tablo 1'de sunulmuştur.

Tablo 1. BOBUT Uygulamalarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Farklı Ölçme Koşullarında Ölçme Duyarlılığına Etkisi

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Madde Kullanım Sıklığı Kontrol Yöntemleri								
			YOK			SH			AD		
			<i>Uyum</i>	<i>Yanlılık</i>	<i>RMSE</i>	<i>Uyum</i>	<i>Yanlılık</i>	<i>RMSE</i>	<i>Uyum</i>	<i>Yanlılık</i>	<i>RMSE</i>
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.9990	.1236	.1523	.9988	.1249	.1584	.9954	.2111	.2715
		Normal	.9997	-.0025	.0428	.9997	-.0035	.0434	.9997	-.0044	.0497
		Tek Biçimli	.9997	.0111	.1014	.9998	.0094	.0987	.9997	.1277	.1112
		Sola Çarpık	.9992	-.0922	.1155	.9991	-.0963	.1201	.9977	-.1420	.1822
	Geniş (N=1000)	Sağa Çarpık	.9990	.1113	.0045	.9989	.1135	.0046	.9962	.1844	.0076
		Normal	.9997	.0009	.0015	.9997	-.0013	.0015	.9997	-.0016	.0017
		Tek Biçimli	.9998	.0095	.0032	.9998	.0097	.0032	.9998	.0125	.0036
		Sola Çarpık	.9993	-.0881	.0036	.9993	-.0895	.0036	.9985	-.1263	.0051
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.9994	.0668	.0863	.9994	.0689	.089	.9964	.1526	.2043
		Normal	.9999	-.0004	.0241	.9999	-.0003	.0242	.9999	-.0025	.0285
		Tek Biçimli	.9999	.0056	.0546	.9999	.0044	.0544	.9999	.0076	.0638
		Sola Çarpık	.9996	-.0522	.0681	.9996	-.0524	.0677	.9985	-.0985	.1321
	Geniş (N=1000)	Sağa Çarpık	.9995	.0587	.0025	.9996	.0597	.0025	.9973	.1275	.0055
		Normal	.9998	-.0005	.0008	.9998	-.0001	.0009	.9998	-.0002	.0009
		Tek Biçimli	.9998	.0059	.0018	.9998	.0051	.0018	.9998	.0074	.0020
		Sola Çarpık	.9997	-.0465	.0020	.9997	-.0467	.0020	.9989	-.0816	.0035

BOBUT uygulamalarında, farklı ölçme koşullarındaki uyum katsayılarının .9954 ile .9999 arasında değiştiği görülmektedir. Bu durumda, farklı ölçme koşullarında kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasında yüksek düzeyde uyum bulunduğu söylenebilir. BOBUT uygulamalarında, farklı ölçme koşullarındaki uyum katsayılarına ilişkin Cohen q değerlerinin .00 ile .90 arasında değişmektedir., Farklı ölçme koşullarında uyum değerleri arasında manidar bir farklılık bulunup bulunmadığına ilişkin Fisher'ın z testi sonuçları incelendiğinde, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığı (YOK-SH) durumlarda uyum katsayıları arasında manidar bir farklılık bulunmamıştır ($p>.05$). Ayrıca normal ve tek biçimli yetenek dağılımlarında, kullanılan madde kullanım sıklık yöntemlerine ilişkin hesaplanan uyum katsayıları arasında manidar farklılık bulunmadığı sonucuna ulaşılmıştır ($p>.05$).

Etki büyüklüğü bakımından, sağa çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayılarına göre, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları lehine geniş düzeyde manidar bir farklılık bulunmuştur ($p<.05$). Sola çarpık dağılımlarda ise, Aşamalı Düşürme Stratejisi ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları arasında kısa test olduğunda orta düzeyde manidar farklılık bulunmakta iken, uzun testlerde geniş düzeyde manidar bir farklılık bulunmuştur ($p<.05$). Bunun yanı sıra, sola çarpık dağılımlarda kısa test- geniş örneklem koşulu haricinde, Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayıları ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında hesaplanan uyum katsayıları arasında geniş düzeyde manidar bir farklılık bulunmakta iken, kısa test- geniş örneklem koşulunda (*Cohen $q=.38$*) orta düzeyde manidar farklılık bulunduğu sonucuna ulaşılmıştır.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin uyum katsayıları genel olarak değerlendirildiğinde, her bir koşulda kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasında yüksek düzeyde uyum bulunduğu sonucuna ulaşılmıştır. Bunun yanı sıra diğer ölçme koşulları sabit tutulduğunda, yetenek dağılımlarının çarpıklık katsayısı sifira yaklaştıkça madde kullanım sıklığı kontrol yöntemlerine ilişkin uyum katsayıları her ne kadar artsa da aralarında manidar bir farklılık bulunmamıştır. Sağa ve sola çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında, hesaplanan uyum katsayılarına göre, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları lehine manidar bir farklılık bulunduğu sonucuna ulaşılmıştır. Diğer ölçme koşullarında ise madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan uyum katsayıları arasında manidar farklılık bulunmamıştır.

Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayıları ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları arasında kısa test - sola çarpık dağılımlarda orta düzeyde ya da orta düzeye yakın (*Cohen $q=.53$*) etki büyüklüğü bulunmakta iken; kısa test – sağa çarpık dağılımlarda ise, geniş düzeyde etki büyüklüğü bulunmuştur. Ayrıca uzun test- hem sağa çarpık dağılımlarda hem de sola çarpık dağılımlarda, Aşamalı Düşürme Stratejisi kullanıldığında, hesaplanan uyum katsayıları ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları arasında geniş düzeyde etki büyüklüğü elde edilmiştir. Dolayısıyla çarpık dağılımlarda madde kullanım sıklık yöntemlerinden Aşamalı Düşürme Stratejisi seçilmesi durumunda diğer madde kullanım sıklık yöntemlerine göre kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki uyum katsayısının önemli derecede azaldığı belirtilebilir. Bir başka ifadeyle, ilgili koşullarda kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki uyum katsayısı büyük oranda azalacağı için, hatanın artacağı söylenebilir.

Yanlılık değerleri incelendiğinde; bu değerler $-.1420$ ile $.2111$ aralığında değiştiği görülmektedir. Uzun test, geniş örneklem, normal yetenek dağılımı ve Sympson Hetter Stratejisi koşulunda yanlılık değeri ($-.0001$) sifira en yakın değer olduğu için kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki ortalama farklılık en düşüktür. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin yanlılık değerleri arasında manidar bir fark bulunup bulunmadığı Mahalanobis uzaklıkları hesaplanarak incelenmiştir.

Madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan yanlılık katsayılarının Mahalanobis uzaklık değerleri .19 ile 14.04 arasında değişmektedir ($p>.001$). Bu doğrultuda, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin kestirilen yetenek puanları ile gerçek yetenek puanları arasındaki ortalama farklılıklar bakımından manidar bir değişim bulunmadığı yorumu yapılabilir.

Uzun test-geniş örneklem ve normal yetenek dağılımında, madde kullanım sıklığı kontrolü yapılmadığında hesaplanan yanlılık değeri, Sympson Hetter Stratejisi ile Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan yanlılık değerlerine göre sifra daha uzaktır. Bu koşul haricindeki tüm koşullarda, madde kullanım sıklığı kontrol yöntemleri bakımından Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan yanlılık değerlerinin, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan yanlılık değerlerine göre, sifra daha uzak olduğu görülmektedir. Dolayısıyla diğer ölçme koşullarının sabit tutulması şartıyla, madde kullanım sıklığı kontrolü yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında kestirilen yetenek düzeyi ile gerçek yetenek düzeyi arasındaki ortalama farklılık diğer yöntemlere göre daha yüksektir.

Yanlılık değerleri her koşulda sifra yakın değerler olduğu için, her bir koşulda kestirilen yetenek parametreleri ve gerçek yetenek parametreleri arasındaki ortalama farklılığın düşük olduğu belirtilebilir. Bunun yanı sıra, her bir ölçme koşulunda yanlılık değerlerine ilişkin elde edilen bulgularda manidar farklılıklar bulunmamakla birlikte, diğer ölçme koşulları sabit tutulduğunda, genel olarak, örneklem büyüklüğü ve test uzunluğu arttıkça, yetenek dağılımlarının çarpıklık katsayısı sifra yaklaştıkça, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanılmadığında kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki ortalama farklılığın düştüğü bulgusuna ulaşılmıştır.

RMSE değerleri incelendiğinde, uzun test-geniş örneklem-normal yetenek dağılımı ve madde kullanım sıklığı kontrol yöntemi uygulanmadığı koşulda en düşük RMSE değeri (.0008); kısa test-küçük örneklem-sağa çarpık dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında en yüksek RMSE değeri (.2715) bulunmuştur.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin RMSE değerleri arasında manidar bir fark bulunup bulunmadığı Mahalanobis uzaklıkları hesaplanarak incelenmiştir. Madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan RMSE katsayılarının Mahalanobis uzaklık değerleri .61 ile 7.32 arasında değişmektedir ($p>.001$). Bu doğrultuda, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin kestirilen yetenek düzeyleri ile gerçek yetenek düzeyleri arasındaki mutlak farklılıklar bakımından manidar bir değişim bulunmadığı sonucuna ulaşılmıştır.

Uzun test-geniş örneklemelerde RMSE değerleri .05'ten küçük olduğundan mükemmel düzeyde uyum (Brown, 2006; Jöreskog ve Sörbom, 1993) bulunduğu görülmüştür. Ayrıca farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin RMSE değerlerine göre, en yüksekten en düşüğe doğru Aşamalı Düşürme Stratejisi, Sympson Hetter Stratejisi, madde kullanım sıklığı kontrol yöntemi uygulanmadığında hesaplanan RMSE değerleri olarak sıralanmaktadır. Buna göre, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemi uygulanmadığında, madde kullanım sıklığı kontrolünün yapıldığı koşullara göre, kestirilen yetenek parametrelerine ilişkin kovaryans ile gerçek yetenek parametrelerine ilişkin kovaryans arasındaki farklılığın sifra daha çok yaklaştığı ortaya çıkmıştır. Dolayısıyla, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemi uygulanmadığında hesaplanan RMSE değerleri, Sympson Hetter Stratejisi ve Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan RMSE değerlerine göre daha küçük olduğundan, madde kullanım sıklığı kontrol yöntemi uygulanmadığında kestirilen yetenek düzeyleri ile gerçek yetenek düzeyleri arasındaki mutlak farklılıkların azalabileceği söylenebilir.

Farklı ölçme koşullarında RMSE değerlerine ilişkin elde edilen bulgularda manidar farklılıklar bulunmamakla birlikte, diğer ölçme koşulları sabit tutulduğunda, genel olarak, örneklem büyüklüğü ve test uzunluğu arttıkça, yetenek dağılımlarının çarpıklık katsayısı sifra yaklaştıkça ve madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanılmadığında kestirilen yetenek puanı ile gerçek yetenek puanı arasındaki mutlak farklılık azalmıştır.

BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin ölçme duyarlılığına etkisini inceleyebilmek için, ölçme duyarlılığı arttığında, hata değerleri azalacağı için, her bir koşulda hata göstergeleri olan uyum, yanlılık ve RMSE katsayıları birlikte ele alınarak değerlendirilmelidir (Gu ve Reckase, 2007; Wang ve Vispoel, 1998; Zheng ve Chang, 2014). Bu bilgi doğrultusunda, ölçme duyarlılığına ilişkin elde edilen bulgular genel olarak değerlendirildiğinde, örneklem büyüklüğü ve test uzunluğu arttıkça yetenek dağılımlarının çarpıklık katsayısı sifira yaklaştıkça genelde ölçme duyarlılığının arttığı sonucuna ulaşılmıştır. Ayrıca madde kullanım sıklığı kontrolü yapılmadığında, sifira en yakın yanlılık ve en düşük RMSE değeri bulgusuna ulaşılsa da yanlılık ve RMSE değerleri bakımından madde kullanım sıklığı kontrolü yöntemleri arasında manidar farklılıklar bulunmadığı sonucuna ulaşılmıştır.

Uyum katsayıları bakımından ise, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığı durumlara ilişkin uyum katsayıları arasında manidar bir farklılık bulunmamakta iken; sağa ve sola çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayılarına göre, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları lehine manidar bir farklılık bulunmuştur.

Madde kullanım sıklığı kontrolü yapılmadığında, madde kullanım sıklığı kontrolünün yapıldığı koşullara göre ölçme duyarlılığının genelde daha yüksek olduğu bulgusuna ulaşılmıştır. Elde edilen bu bulgu, kullanım sıklığının kontrol edilmediği durumun referans olarak alındığı Boyd (2003)'un ve Davis (2004)'in çalışmasındaki bulgularla paralellik göstermektedir. BOBUT uygulamalarında madde seçiminde madde bilgi fonksiyonu ölçüt alındığından, madde kullanım sıklığı kontrolü yapılmadığında madde seçimi kısıtlanmayacağı için, yüksek bilgi değerine sahip maddeler seçilerek yetenek kestiriminin standart hatasının düşmesi bu bulguya gerekçe gösterilebilir. Bir başka ifadeyle, yüksek bilgi değerine sahip maddeler seçildiğinde yetenek kestiriminin standart hatasının düşmesinin, madde kullanım sıklığı kontrolü yapılmadığında, madde kullanım sıklığı kontrolünün yapıldığı koşullara göre ölçme duyarlılığının genelde daha yüksek bulunmasına neden olduğu söylenebilir.

Ölçme duyarlılığını maksimum yapacak maddeler seçildiğinde, kullanım sıklığı oranları düzgün dağılmayacağından (Pastor vd., 2002) madde havuzunun dengeli bir şekilde kullanılmayıp, sadece sınırlı sayıda bazı maddelerin kullanılmasına sebep olmamak için (Hulin, Drasgow ve Parson, 1983, aktaran Revuelta ve Ponsoda, 1998) madde kullanım sıklığı kontrol yöntemleriyle, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin sağlanması amaçlanmaktadır (Pastor vd., 2002). Bu durumda, ilgili değişkenler birlikte ele alındığında, farklı madde kullanım sıklığı kontrolü koşullarında ölçme duyarlılıkları arasında genel olarak büyük farklılıklar bulunmamıştır. Bu bulgu sonucunda, ilgili madde kullanım sıklığı kontrol yöntemleri kullanıldığında, ölçme duyarlılığının düşürülmeyeceği ifade edilebilir.

BOBUT Uygulamalarında Farklı Ölçme Koşullarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Test Güvenliğine Etkisi

BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin test güvenliğine etkisini incelemek amacıyla aşağıda sırasıyla madde kullanım sıklığı oranlarının standart sapma değerleri, maksimum kullanım sıklık oranı dağılımları, madde havuzundan yararlanmama indeksleri, madde kullanım sıklık oranlarının çarpıklık katsayıları (χ^2 değerleri) ve test örtüşme indekslerine ilişkin bulgular ve yorumlara yer verilmektedir. BOBUT uygulamalarında madde kullanım sıklığı kontrol yöntemlerinin farklı ölçme koşullarında test güvenliğine etkisine ilişkin bulgular Tablo 2'de yer almaktadır.

Tablo 2. BOBUT Uygulamalarında Madde Kullanım Sıklığı Kontrol Yöntemlerinin Farklı Ölçme Koşullarında Test Güvenliğine Etkisi

Test Uzunluğu	Örneklem	Yetenek Dağılımı	Madde Kullanım Sıklığı Kontrol Yöntemleri														
			YOK					SH					AD				
			SS	Max	Yarar	χ^2	Örtüşme	SS	Max	Yarar	χ^2	Örtüşme	SS	Max	Yarar	χ^2	Örtüşme
Kısa (25 madde)	Küçük (n=250)	Sağa Çarpık	.097	.834	.1259	375.06	.396	.091	.622	.1256	330.99	.353	.048	.193	.0980	90.78	.112
		Normal	.074	.750	.1254	218.66	.239	.071	.486	.1253	201.84	.223	.037	.117	.0939	54.50	.076
		Tek Biçimli	.068	.677	.1242	183.84	.205	.064	.448	.1242	165.06	.187	.038	.152	.0952	58.49	.079
		Sola Çarpık	.101	.922	.1272	406.71	.429	.092	.642	.1264	341.64	.363	.044	.184	.0928	77.97	.099
	Geniş (N=1000)	Sağa Çarpık	.095	.830	.0308	362.35	.386	.090	.594	.0308	321.22	.346	.047	.189	.0243	89.39	.114
		Normal	.075	.739	.0309	224.52	.249	.072	.492	.0309	208.19	.232	.037	.117	.0232	54.17	.078
		Tek Biçimli	.068	.683	.0310	182.74	.207	.064	.454	.0309	164.07	.188	.038	.151	.0238	58.23	.082
		Sola Çarpık	.099	.915	.0312	387.70	.412	.091	.623	.0308	328.23	.352	.044	.181	.0239	76.78	.101
Uzun (50 madde)	Küçük (n=250)	Sağa Çarpık	.142	.840	.0534	403.50	.452	.138	.727	.0528	381.86	.429	.063	.200	.0312	79.54	.126
		Normal	.098	.761	.053	193.19	.239	.096	.508	.0536	185.85	.232	.054	.151	.0322	58.51	.105
		Tek Biçimli	.089	.686	.051	158.61	.205	.087	.463	.0504	150.22	.197	.055	.183	.0330	60.25	.107
		Sola Çarpık	.143	.9294	.0533	406.23	.454	.137	.733	.0530	375.25	.422	.060	.196	.0288	72.14	.119
	Geniş (N=1000)	Sağa Çarpık	.139	.836	.0127	389.23	.439	.136	.721	.0126	369.19	.419	.063	.197	.0075	78.38	.128
		Normal	.099	.749	.0127	197.60	.247	.098	.523	.0126	191.56	.241	.054	.151	.0078	57.01	.106
		Tek Biçimli	.089	.689	.0126	157.27	.206	.086	.461	.0126	148.79	.198	.055	.181	.0082	59.82	.109
		Sola Çarpık	.139	.921	.0128	387.87	.437	.134	.714	.0127	359.85	.409	.060	.192	.0069	71.50	.121

Tablo 2 incelendiğinde, madde kullanım sıklığı oranlarının standart sapma değerlerine ilişkin, en yüksek değer uzun test-küçük örneklem-sola çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında (.143); en düşük değer kısa test-küçük örneklem-normal dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında (.037) bulunmuştur. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin madde kullanım sıklığı oranlarının standart sapma değerleri arasında manidar bir fark bulunup bulunmadığına ilişkin hipotez testi sonuçları incelendiğinde, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında (kısa test-geniş örneklem-sola çarpık yetenek dağılımı koşulu haricinde) madde kullanım sıklığı oranlarının standart sapma değerleri arasında manidar bir farklılık bulunmamıştır ($p>.05$). Ayrıca Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan madde kullanım sıklığı oranlarının standart sapma değerleri ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında (F_{YOK-AD}) ve Sympson Hetter Stratejisi kullanıldığında hesaplanan (F_{SH-AD}) madde kullanım sıklığı oranlarının standart sapma değerleri arasında manidar bir farklılık bulunduğu sonucuna ulaşılmıştır ($p<.05$).

Kısa test-geniş örneklem-sola çarpık yetenek dağılımı koşulunda madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında madde kullanım sıklığı oranlarının standart sapma değerleri arasında her ne kadar manidar farklılık bulunsa da F değerinin kritik değere çok yakın olduğu saptanmıştır ($F_{YOK-SH} = 1.18 > F_{999,999,.025} = 1.13$). Bu bulguya gerekçe olarak, sola çarpık yetenek dağılımındaki (üstün yetenekli) bireylerin (geniş örneklem) yetenek düzeylerine hitap edebilecek madde havuzundaki maddeleri seçmede yaşanan güçlüklerin yanı sıra test sonlandırma kuralının kısa test olması, madde seçme yönteminde de önsel dağılım kullanılarak bir kısıtlamanın yapılıp yapılmamasına göre madde kullanım sıklığı oranları arasındaki çarpıklığın artmasına neden olabileceği söylenebilir.

Farklı ölçme koşullarında madde kullanım sıklığı oranlarının standart sapma değerlerine ilişkin elde edilen bulgular genel olarak değerlendirildiğinde, SH ile YOK arasında genelde manidar farklılık bulunmamakta iken; SH ile AD arasında ve YOK ile AD arasında manidar farklılık bulunmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan madde kullanım sıklığı oranlarının standart sapma değerleri, diğer iki yönteme göre manidar bir şekilde daha küçüktür. Dolayısıyla Aşamalı Düşürme Stratejisi kullanıldığında, madde kullanım sıklığı oranlarının diğer stratejilere göre manidar bir şekilde daha homojen dağılım gösterdiği söylenebilir.

Diğer ölçme koşulları sabit tutulduğunda, genel olarak, test uzunluğu azaldıkça, yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça, örneklem büyüklüğü genelde arttıkça ve madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan madde kullanım sıklığı oranlarının standart sapma değerleri azaldığından madde kullanım sıklığı oranlarının daha homojen dağılım gösterdiği ortaya çıkmaktadır. Bir başka ifadeyle, ilgili koşullarda madde kullanım sıklığı oranları daha homojen dağılım gösterdiğinden, madde kullanım sıklığı dengeli dağılım göstermekte ve madde havuzu daha verimli kullanılmaktadır.

Maksimum madde kullanım sıklığı oranlarına ilişkin, en yüksek değer uzun test-küçük örneklem-sola çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında (.929); en düşük değer kısa test-hem küçük hem de geniş örneklem-normal dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında (.117) bulunmuştur. Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin maksimum madde kullanım sıklığı oranları arasında manidar bir farklılık bulunup bulunmadığına ilişkin iki oran arasındaki farkın testi sonuçları incelendiğinde, maksimum madde kullanım sıklığı oranları arasındaki farklara ilişkin z değerleri 5.19 ile 48.82 arasında değiştiğinden ilgili iki oran arasındaki farklılıkların manidar olduğu sonucuna ulaşılmıştır ($p<.05$).

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin hesaplanan maksimum madde kullanım sıklığı oranlarına göre, en düşükten en yükseğe doğru Aşamalı Düşürme Stratejisi, Sympson Hetter Stratejisi, madde kullanım sıklığı kontrolünün yapılmadığı koşul olarak sıralanmaktadır. Maddelerin kullanım sıklığı oranı arttıkça maddelerin ifşâ olma ihtimalleri arttığı için

ilgili koşullarda Aşamalı Düşürme Stratejisi kullanıldığında, maddelerin ifşâ olarak psikometrik özelliklerini kaybetme riskinin diğer koşullara göre manidar bir şekilde azaldığı görülmektedir.

Her bir koşulda madde havuzunun kullanılma derecesini belirleyebilmek için, madde havuzundan yararlanmama oranları (pool utilization index) incelendiğinde, en yüksek değer kısa test-küçük örneklem-sağa çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında (.1259); en düşük değer uzun test-geniş örneklem-sola çarpık dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında (.0069) bulunmuştur. Dolayısıyla farklı ölçme koşullarında madde havuzunun kullanılma derecesi ($1-.1259=87.41$) ile ($1-.0069=99.31$) arasında değişim göstermektedir.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin madde havuzundan yararlanmama oranları arasında manidar bir fark bulunup bulunmadığına ilişkin iki oran arasındaki farkın testi sonuçları incelendiğinde, madde havuzundan yararlanmama oranları arasındaki farklara ilişkin z değerleri -.03 ile 1.38 arasında değiştiğinden ilgili iki oran arasındaki farklılıkların manidar olmadıkları sonucuna ulaşılmıştır ($p>.05$).

Farklı ölçme koşullarında madde havuzundan yararlanmama oranları arasında manidar farklılıklar bulunmamakla birlikte, diğer ölçme koşulları sabit tutulduğunda, örneklem büyüklüğü ve test uzunluğu arttıkça, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde havuzundan yararlanmama oranları azalmaktadır. Dolayısıyla örneklem büyüklüğü ve test uzunluğu arttıkça, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde havuzunun kullanılma derecesinin arttığı söylenebilir.

Madde kullanım sıklık yöntemlerine ilişkin madde kullanım sıklık oranlarının istenilen kullanım sıklık oranına uzaklığının belirlenmesi, bir başka ifadeyle madde havuzu kullanımının ne kadar dengeli dağılım gösterdiğinin saptanabilmesi için farklı ölçme koşullarında madde kullanım sıklık oranlarının çarpıklık katsayısı (χ^2 değerleri) incelendiğinde, en yüksek χ^2 değeri kısa test-küçük örneklem-sola çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında ($\chi^2 = 406.71$) elde edilirken; en düşük χ^2 değeri kısa test-uzun örneklem-normal dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında en düşük ($\chi^2 = 54.17$) elde edilmiştir.

Farklı ölçme koşullarında hesaplanan χ^2 değerlerine göre, en düşükten en yükseğe doğru Aşamalı Düşürme Stratejisi, Sympson Hetter Stratejisi, madde kullanım sıklığı kontrol yöntemi uygulanmadığında hesaplanan χ^2 değerleri olarak sıralanmaktadır. Dolayısıyla madde kullanım sıklığı kontrolü yapılmadığında ya da madde kullanım sıklığı kontrol yöntemlerinden Sympson Hetter stratejisi ve Aşamalı Düşürme stratejisi kullanıldığında hangisinin en ideal madde kullanım sıklık yöntemi olduğuna karar verilebilmesi için, karşılaştırma ölçüsü olan F istatistiklerine ilişkin her bir koşulda $F_{\chi^2_{AD}, \chi^2_{SH}} < 1$ ve $F_{\chi^2_{YOK}, \chi^2_{SH}} > 1$ sonuçlarına ulaşılmıştır. Bu durumda her bir koşulda, madde kullanım sıklığı oranlarının genel olarak dengelenmesi bakımından, madde kullanım sıklığı kontrolü yapılmamasının, Sympson Hetter Stratejisine göre daha kullanışsız ($F_{\chi^2_{YOK}, \chi^2_{SH}} > 1$) bir yöntem olduğu; Aşamalı Düşürme Stratejisinin ise Sympson Hetter Stratejisine göre daha iyi bir yöntem ($F_{\chi^2_{AD}, \chi^2_{SH}} < 1$) olduğu görülmektedir. Elde edilen bulgular ve ilgili koşullar doğrultusunda madde kullanım sıklığı kontrolü yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde havuzunun daha dengeli dağılım gösterdiği sonucuna ulaşılmıştır.

Diğer ölçme koşulları sabit tutulduğunda, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında *test uzunluğu arttığında* χ^2 değeri, sağa çarpık dağılımda ve sola çarpık dağılımda (küçük örneklem- kontrol yöntemi uygulanmadığı koşul dışında) artmakta iken normal ve tek biçimli dağılımlarda χ^2 değerleri azalmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında ise diğer iki strateji koşullarında elde edilen bulguların aksine, *test uzunluğu arttığında*, χ^2 değeri sağa çarpık dağılımda ve sola çarpık dağılımda azalmakta iken normal ve tek biçimli dağılımlarda χ^2 değerleri artmaktadır.

Örneklem büyüklüğü arttığında sağa çarpık, sola çarpık ve tek biçimli dağılımlarda, χ^2 değerleri azalmaktadır. Normal yetenek dağılımlarında ise, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında χ^2 değeri artmakta iken; Aşamalı Düşürme Stratejisi kullanıldığında χ^2 değeri azalmaktadır.

Yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça, χ^2 değeri azalmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında sırasıyla *sağa çarpık, sola çarpık, tek biçimli ve normal dağılıma doğru* χ^2 değeri azalmaktadır. Ayrıca, Aşamalı Düşürme Stratejisi kullanıldığında, *normal dağılımlarda* χ^2 değeri en küçük iken; madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında *tek biçimli dağılımlarda* χ^2 değeri en küçüktür.

Farklı ölçme koşullarında madde kullanım sıklık oranlarının çarpıklık katsayılarına (χ^2 değerlerine) ilişkin elde edilen bulgular genel olarak incelendiğinde, Aşamalı Düşürme Stratejisi kullanıldığında en düşük χ^2 değeri, kısa test-geniş örneklem-normal dağılım koşulunda elde edilmiştir. Madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında, en düşük χ^2 değeri, uzun test-geniş örneklem- tek biçimli dağılım koşulunda elde edilmiştir. Fakat tüm koşullarda da Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan χ^2 değeri diğer stratejilere göre daha küçük bulunmuştur. Dolayısıyla madde kullanım sıklık yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında madde kullanım sıklık oranının istenilen kullanım sıklık oranına (kısa testlerde .025; uzun testlerde .05) uzaklığının azaldığı söylenebilir. Başka bir deyişle Aşamalı Düşürme Stratejisi kullanıldığında, madde kullanım sıklık oranının çarpıklık katsayısı sıfıra yaklaştığı için madde havuzu kullanımı daha dengeli dağılım göstermektedir.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin test örtüşme indeksleri incelendiğinde, farklı ölçme koşullarında test örtüşme indekslerine ilişkin, en yüksek değer uzun test-küçük örneklem-sola çarpık dağılım ve madde kullanım sıklığı kontrol yöntemi uygulanmadığında (.454); en düşük değer kısa test-küçük örneklem-normal dağılım ve Aşamalı Düşürme Stratejisi kullanıldığında (.076) bulunmuştur.

Farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerine ilişkin test örtüşme indeksleri arasında manidar bir fark bulunup bulunmadığına ilişkin iki oran arasındaki farkın testi sonuçları incelendiğinde, Z_{YOK-AD} değerleri 3.05 ile 17.06 arasında değişmekte iken; Z_{SH-AD} değerleri 2.82 ile 15.47 arasında değişmektedir. Dolayısıyla ilgili ölçme koşullarında, Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan test örtüşme indeksleri ile madde kullanım sıklığı kontrol yöntemi uygulanmadığında (Z_{YOK-AD}) ve Sympson Hetter Stratejisi kullanıldığında hesaplanan (Z_{SH-AD}) test örtüşme indeksleri arasında manidar farklılık olduğu sonucuna ulaşılmıştır ($p < .05$).

Madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında (kısa test-geniş örneklem-sola çarpık yetenek dağılımı koşulu $Z_{YOK-SH} = 2.78$ haricinde) hesaplanan test örtüşme indeksleri arasında manidar bir farklılık bulunmamıştır ($p > .05$).

Farklı ölçme koşullarında hesaplanan test örtüşme indekslerine ilişkin elde edilen bulgular genel olarak değerlendirildiğinde, SH ile YOK arasında genelde manidar farklılık bulunmamakta iken; SH ile AD arasında ve YOK ile AD arasında manidar farklılık bulunmuştur. Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan test örtüşme indekslerinin diğer iki yönteme göre manidar bir şekilde daha küçük olduğu ortaya çıkmıştır. Test örtüşme indeksi büyüdükçe, test güvenliği zedelenmektedir (Huangvd., 2012). Dolayısıyla Aşamalı Düşürme stratejisi kullanıldığında, madde kullanım sıklığı kontrolü yapılmadığı ve Sympson Hetter stratejisi kullanıldığı koşullara göre test güvenliğinin daha az zedelendiği ifade edilebilir.

Test uzunluğu ve yetenek dağılımı sabit tutulup, madde kullanım sıklığı kontrol yöntemi uygulanmadığı ve Sympson Hetter Stratejisi kullanıldığı koşullarda *örneklem büyüklüğü arttığında* sağa çarpık ve sola çarpık dağılımlarda, test örtüşme indeksleri azalmakta iken; normal ve tek biçimli yetenek dağılımlarında ise, test örtüşme indeksleri artmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında *örneklem büyüklüğü arttıkça* test örtüşme indeksleri de artmaktadır.

Test uzunluğu ve yetenek dağılımı sabit tutulup, örneklem büyüklüğü arttığında; madde kullanım sıklığı kontrol yöntemi uygulanmadığı ve Sympson Hetter Stratejisi kullanıldığı koşullarda, sağa çarpık ve sola çarpık dağılımlarda, test örtüşme indekslerinde azalma olsa da normal ve tek biçimli yetenek dağılımlarına göre test örtüşme indeksleri daha yüksek olduğu için test güvenliğinin daha olumsuz yönde etkilendiği yorumu yapılabilir. Ayrıca Aşamalı Düşürme Stratejisi kullanıldığında *örneklem büyüklüğü arttıkça* test örtüşme indeksleri artsa da diğer madde kullanım sıklığı kontrol yöntemlerinde hesaplanan test örtüşme indekslerine göre, her bir koşulda manidar bir şekilde test örtüşme indeksleri daha küçük olduğundan Aşamalı Düşürme Stratejisi kullanıldığında test güvenliğinin daha olumlu yönde etkileneceği ifade edilebilir.

Diğer ölçme koşulları sabit tutulduğunda, madde kullanım sıklığı kontrol yöntemi uygulanmadığı ve Sympson Hetter Stratejisi kullanıldığı koşullarda *test uzunluğu arttığında*, test örtüşme indeksleri sağa ve sola çarpık dağılımda artmakta iken normal ve tek biçimli dağılımda büyük farklılık bulunmamıştır. O halde test uzunluğu arttıkça; sağa ve sola çarpık yetenek dağılımlarında hesaplanan test örtüşme indeksleri ile normal ve tek biçimli yetenek dağılımlarında hesaplanan test örtüşme indeksleri arasındaki genişlik daha da arttığından, sağa ve sola çarpık yetenek dağılımlarında test güvenliğinin daha olumsuz yönde etkilendiği çıkarımında bulunulabilir.

Aşamalı Düşürme Stratejisi kullanıldığında *test uzunluğu arttıkça* test örtüşme indeksleri artsa da diğer madde kullanım sıklığı kontrol yöntemlerinde hesaplanan test örtüşme indekslerine göre, her bir koşulda manidar bir şekilde daha küçük olduğu için Aşamalı Düşürme Stratejisi kullanıldığında test güvenliğinin daha olumlu yönde etkilendiği söylenebilir.

Yetenek dağılımlarının çarpıklık katsayısı sıfıra yaklaştıkça, test örtüşme indeksinin azaldığı görülmüştür. Aşamalı Düşürme Stratejisi kullanıldığında sırasıyla *sağa çarpık, sola çarpık, tek biçimli ve normal dağılıma doğru* test örtüşme indeksi azalmaktadır. Aşamalı Düşürme Stratejisi kullanıldığında, *normal dağılımlarda* test örtüşme indeksi en küçük iken; madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında *tek biçimli dağılımlarda* test örtüşme indeksi en küçüktür.

Farklı ölçme koşullarında test örtüşme indekslerine ilişkin elde edilen bulgular genel olarak incelendiğinde, Aşamalı Düşürme Stratejisi kullanıldığında en düşük test örtüşme indeksi, kısa test - küçük örneklem - normal dağılım koşulunda elde edilmiştir. Madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında, kısa test - küçük örneklem - tek biçimli dağılım koşulunda en düşük test örtüşme indeksi bulgusuna ulaşılmıştır. Fakat tüm koşullarda da Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan test örtüşme indeksleri diğer stratejilere göre manidar şekilde daha küçük bulunmuştur. Dolayısıyla BOBUT uygulamalarında, madde kullanım sıklığı kontrol yöntemlerinden Aşamalı Düşürme Stratejisi kullanıldığında, herhangi iki bireyin karşılaştıkları aynı maddelerin sayısı azaldığından hem test güvenliği daha az zedelendiği hem de madde havuzunun yıllarca kullanılabilir olmasına imkân sunduğu belirtilebilir.

Madde kullanım sıklığı oranının $r_1 = .20$ olması, ilgili maddenin yaklaşık olarak her beş BOBUT uygulamasından birinde uygulandığı anlamına gelmektedir (Weiss ve Guyer, 2012). Maddenin kullanım sıklığı oranı arttıkça, ilgili maddenin ifşâ edilerek psikometrik özelliklerinin yapay olarak düşmesine neden olmaktadır (Segall, 2004; Revuelta ve Ponsoda, 1998). Bu nedenle, madde havuzundaki maddelerin ne çok sık kullanılması ne de hiç kullanılmaması istenen durumdur. Bu bilgiler doğrultusunda, madde havuzunun etkili kullanılma/işlevsellik derecesini belirleyebilmek için farklı ölçme koşullarında madde havuzunun kullanımına ilişkin dağılıma Tablo 3'te yer verilmiştir.

Tablo 3. Farklı Ölçme Koşullarında Madde Havuzunun Kullanımına İlişkin Dağılım

Test Uzunluğu	Örneklem Büyüklüğü	Yetenek Dağılımı	Kullanım Sıklığı Yöntemi	$r_i=0$	$0 < r_i \leq .2$	$.2 < r_i \leq .4$	$.4 < r_i \leq .6$	$.6 < r_i \leq .8$	$.8 < r_i \leq 1$
Kısa (25 madde)	Küçük (n=250)	Sağa çarpık	YOK	787	168	18	20	6	1
			SH	785	165	28	20	2	0
			AD	613	387	0	0	0	0
		Normal	YOK	784	177	33	3	3	0
			SH	783	173	36	8	0	0
			AD	587	413	0	0	0	0
		Tek Biçimli	YOK	776	191	28	4	1	0
			SH	776	181	42	1	0	0
			AD	595	405	0	0	0	0
		Sola Çarpık	YOK	795	168	13	15	6	3
			SH	790	166	19	23	2	0
			AD	608	392	0	0	0	0
	Geniş (N=1000)	Sağa Çarpık	YOK	771	186	19	19	4	1
			SH	770	180	28	22	0	0
			AD	608	392	0	0	0	0
		Normal	YOK	773	186	32	6	3	0
			SH	772	184	35	9	0	0
			AD	581	419	0	0	0	0
		Tek Biçimli	YOK	774	195	26	4	1	0
			SH	773	185	41	1	0	0
			AD	596	404	0	0	0	0
		Sola Çarpık	YOK	780	182	16	15	5	2
			SH	771	184	21	23	1	0
			AD	598	402	0	0	0	0
Uzun (50 madde)	Küçük (n=250)	Sağa çarpık	YOK	668	251	28	26	24	3
			SH	660	253	37	26	24	0
			AD	390	610	0	0	0	0
		Normal	YOK	667	226	95	9	3	0
			SH	670	222	98	10	0	0
			AD	403	597	0	0	0	0
		Tek Biçimli	YOK	634	279	82	4	1	0
			SH	630	274	95	1	0	0
			AD	412	588	0	0	0	0
		Sola Çarpık	YOK	666	261	26	16	28	3
			SH	662	259	29	26	24	0
			AD	360	640	0	0	0	0
	Geniş (N=1000)	Sağa Çarpık	YOK	635	280	33	31	18	3
			SH	633	279	36	32	20	0
			AD	374	626	0	0	0	0
		Normal	YOK	633	270	85	9	3	0
			SH	631	272	81	16	0	0
			AD	390	610	0	0	0	0
		Tek Biçimli	YOK	631	283	81	4	1	0
			SH	630	274	95	1	0	0
			AD	412	588	0	0	0	0
		Sola Çarpık	YOK	639	287	28	20	23	3
			SH	633	286	31	32	18	0
			AD	344	656	0	0	0	0

Tablo 3 incelendiğinde, 100 replikasyon sonucunda her bir BOBUT uygulamasının hiçbirinde kullanılmayan ($r_i = 0$) madde sayısı, madde kullanım sıklığı kontrol yöntemi uygulanmadığında 631 ile 795; Sympson Hetter Stratejisi'nde 630 ile 790; Aşamalı Düşürme Stratejisi'nde 344 ile 613 aralığında değişmektedir. İlgili koşullarda madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında her bir BOBUT uygulamasının hiçbirinde kullanılmayan ($r_i = 0$) madde sayısı bakımından büyük farklılıklar bulunmamıştır. Aşamalı Düşürme Stratejisi kullanıldığında ise, diğer stratejilere göre tüm koşullarda kullanılmayan ($r_i = 0$) madde sayısı bakımından daha küçük olduğu sonucuna ulaşılmıştır.

Her bir BOBUT uygulamasının hiçbirinde kullanılmayan ($r_i = 0$) madde sayısının bir başka ifadeyle, madde havuzundan yararlanmama bakımından uzun test ve geniş örneklerde diğer ölçme koşullarına göre daha düşük değerlerde olduğu belirtilebilir. Benzer olarak madde havuzunun kullanılma derecesi bakımından uzun test ve geniş örneklerde madde havuzunun kullanılma derecesinin daha yüksek olduğu görülmüştür.

Bütün maddeler için istenilen tek biçimli kullanım sıklık oranları (Tay, 2015) kısa test için .025, uzun test için .05'dir. Bunun yanı sıra ilgili literatürde BOBUT uygulamalarında hedeflenen kullanım sıklık oranı .20 (Eggen, 2001; Eignor, Stocking, Way ve Steffen, 1993; Veldkamp ve Van Der Linden, 2008) olarak tanımlanmaktadır. Bu bilgiler doğrultusunda madde havuzunun verimli/işlevsel kullanımına ilişkin, $.00 < r_i \leq .20$ madde kullanım sıklığı aralığındaki değerler incelendiğinde, test uzunluğu ve örneklem büyüklüğü arttıkça tek biçimli dağılım koşullarında büyük farklılıklar bulunmamakta iken diğer tüm ölçme koşullarında kullanılan madde sayısının arttığı bulgusuna ulaşılmıştır.

Aşamalı Düşürme Stratejisi kullanıldığında, hedeflenen kullanım sıklık oranından daha yüksek ($r_i > .20$) madde kullanım sıklığı aralığında, kullanılan madde sayısı bulunmamaktadır. Sympson Hetter Stratejisi kullanıldığında hedeflenen kullanım sıklık oranından daha yüksek madde kullanım sıklık oranlarının dağılımı incelendiğinde, uzun test-küçük örneklem-sola çarpık dağılımda hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı fazla olduğu için ilgili maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin en fazla (*Skewness*=-1.86) olduğu görülmüştür. Sympson Hetter Stratejisi kullanıldığı - tek biçimli ve uzun testlerde hem küçük örneklem hem de geniş örneklem koşullarında hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı az olduğu için ilgili maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin en az (*Skewness* =2.00) olduğu belirtilebilir.

Madde kullanım sıklığı kontrol yöntemi uygulanmadığında, hedeflenen kullanım sıklık oranından ($r_i = .20$) daha yüksek madde kullanım sıklık oranlarının dağılımları incelendiğinde, uzun test-küçük örneklem-sağa çarpık dağılımda hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı fazla olduğu için ilgili maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin en fazla (*Skewness*=-1.09) olduğu görülmüştür. Bunun yanı sıra, uzun test tek biçimli dağılımda hedeflenen kullanım sıklık oranının üzerinde maddelerin sayısı az olduğu için ilgili maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin en az (*Skewness* =1.15) olduğu sonucuna ulaşılmıştır.

Madde kullanım sıklığı kontrol yöntemi uygulanmadığında, sağa ve sola çarpık dağılımlarda ve Sympson Hetter Stratejisi-uzun test sağa ve sola çarpık dağılım koşullarında hedeflenen kullanım sıklık oranından ($r_i = .20$) yüksek madde kullanım sıklıklarına ilişkin dağılımın çarpıklık katsayıları negatiftir. Dolayısıyla sağa ve sola çarpık dağılımlarda, madde kullanım sıklığı kontrol yöntemi uygulanmadığında hem kısa testlerde hem de uzun testlerde; Sympson Hetter Stratejisi kullanıldığında ise sadece uzun testlerde maddelerin ifşâ edilerek psikometrik özelliklerinin düşme riskinin arttığı ifade edilebilir.

Tartışma, Sonuç ve Öneriler

BOBUT uygulama sürecinde farklı ölçme koşullarında ölçme duyarlılığı ve test güvenliğine ilişkin bulgular genel olarak değerlendirildiğinde, madde kullanım sıklığı kontrolü yapılmadığında en düşük RMSE ve sifira en yakın yanlılık değeri bulgusuna ulaşılsa da RMSE ve yanlılık değerleri bakımından madde kullanım sıklığı kontrolü yöntemleri arasında manidar farklılıklar bulunmadığı sonucuna ulaşılmıştır. Uyum katsayıları bakımından ise, farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığı durumlara ilişkin uyum katsayıları arasında manidar bir farklılık bulunmamakta iken; sağa ve sola çarpık dağılımlarda Aşamalı Düşürme Stratejisi kullanıldığında hesaplanan uyum katsayılarına göre, madde kullanım sıklığı kontrol yöntemi uygulanmadığında ve Sympson Hetter Stratejisi kullanıldığında hesaplanan uyum katsayıları lehine manidar bir farklılık bulunmuştur.

Ölçme duyarlılığına ilişkin ilgili değişkenler birlikte ele alınıp değerlendirildiğinde, farklı madde kullanım sıklığı kontrolü koşullarında ölçme duyarlılıkları arasında genel olarak büyük farklılıklar bulunmamıştır. Bu bulgu doğrultusunda, ilgili madde kullanım sıklığı kontrol yöntemleri kullanıldığında, ölçme duyarlılığının düşmeyeceği belirtilebilir.

BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinin test güvenliğine etkisini inceleyebilmek için, madde kullanım sıklığı oranı, madde havuzundan yararlanmama, test örtüşme indeksi birlikte ele alınarak değerlendirilmektedir (Davis ve Dodd, 2005). Bu durumda, test güvenliğine ilişkin ilgili değişkenler birlikte ele alınıp değerlendirildiğinde, ilgili koşullarda Aşamalı Düşürme yöntemi seçildiğinde, test güvenliğinin daha iyi sağlanarak madde kullanım sıklığının dengeli dağılım gösterdiği ve madde havuzu kullanımının daha verimli hale geldiği görülmüştür (Boyd, 2003; Davis, 2002). Bunların yanı sıra, Aşamalı Düşürme yöntemi seçildiğinde, bireyler için örtüşen maddelerin oranı (Chen vd., 1999) ve maksimum kullanım sıklık oranı düşük olduğundan madde havuzu ifşâ olmadan birçok uygulama için süreklilik sağlayacağı (Revueleta ve Ponsoda, 1998) ortaya çıkmıştır.

Madde kullanım sıklığı kontrol yöntemleriyle, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin sağlanması amaçlanmaktadır (Boyd, 2003; Boydv., 2013; Davis ve Dodd, 2005; Pastor vd., 2002). Bu bilgi doğrultusunda, yapılan bu araştırmanın sonucunda, BOBUT uygulamalarında farklı ölçme koşullarında madde kullanım sıklığı kontrol yöntemlerinden en yaygın kullanılan (Veldkampvd., 2010) Sympson Hetter stratejisi yerine Aşamalı Düşürme Stratejisi seçilmesi durumunda, ölçme duyarlılığını düşürmeden madde havuzunu daha dengeli kullanarak test güvenliğinin sağlandığı çıkarımında bulunulabilir. Bu bulgu, Davis (2002)'in belirttiği, Sympson Hetter stratejisinde her ne kadar BOBUT uygulamasından önce, hedeflenen kullanım sıklığı oranına ulaşılabilmesi için tekrarlı simülasyonlarla kullanım sıklığı kontrol parametreleri belirlense de madde örtüşme indeksi ve madde havuzunun kullanılma derecesi bakımından hedeflenen düzeyde madde kullanım sıklığı sonucuna ulaşılmadığına ilişkin ifadeyle benzerlik göstermektedir. Ayrıca madde seçme yöntemlerine ve farklı güçlük düzeyindeki madde havuzlarına göre, çeşitli madde kullanım sıklığı kontrol yöntemlerinin karşılaştırıldığı BoztunçÖztürk (2014)'ün çalışmasında Sympson Hetter Stratejisi yerine Aşamalı Düşürme madde kullanım sıklığı kontrol yönteminin kullanılması durumunda test güvenliğinin arttığına ilişkin bulguyla paralellik göstermektedir.

Bu araştırmada, BOBUT uygulamalarında farklı ölçme koşullarında, 1-0 şeklinde puanlanan çoktan seçmeli maddelerde madde kullanım sıklığı kontrol yöntemleri karşılaştırılmıştır. Çok kategorili puanlanan maddelerde madde kullanım sıklığı kontrol yöntemlerinin farklı yetenek dağılımlarında, test uzunluklarında ve örneklem büyüklüklerinde ölçme duyarlılığına ve test güvenliğine etkisi üzerine bir araştırma önerilebilir. Ayrıca, örneklem yanlılığı oluşmaması için araştırma kapsamında, her bir koşulda 100 tekrar kullanılmıştır. Tekrar sayılarının madde kullanım sıklığı kontrol yöntemlerinin ölçme duyarlılığına ve test güvenliğine etkisini incelemek amacıyla yetenek dağılımları, test uzunlukları ve örneklem büyüklükleri sabit tutulup tekrar sayıları üzerinde farklı koşullar oluşturarak bir araştırma yapılabilir.

Kaynakça

- Agresti, A. ve Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2), 119-126.
- Ariel, A., Veldkamp, B. P. ve Van Der Linden, W. J. (2004). Constructing rotating item pools for constrained adaptive testing. *Journal of Educational Measurement*, 41(4), 345-359.
- Aytuğ Koşan, A. M. (2013). *Tıp eğitiminde gelişim sınavı soru bankası oluşturulması ve benzetim verileri ile bilgisayar uyarlamalı test uygulaması* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Boyd, A. M., Dodd, B ve Fitzpatrick, S. (2013). A comparison of exposure control procedures in CAT systems based on different measurement models for testlets. *Applied Measurement in Education*, 26(2), 113-135.
- Boyd, M. A. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems* (Yayımlanmamış doktora tezi). The University of Texas, Austin.
- Boztunç Öztürk, N. (2014). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı kontrol yöntemlerinin incelenmesi* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, N.J.: Guilford Press.
- Bulut, O. ve Kan, A. (2012). Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eurasian Journal of Educational Research*, 49, 61-80.
- Burt, W. M., Kim S. J., Davis, L. Lve Dodd, B. G. (2003, April). *A comparison of item exposure control procedures using a CAT system based on the generalized partial credit model*. Annual meeting of the American Educational Research Association sunulmuş bildiri, Chicago, IL.
- Byrne, B. M. (1998). *Structural equation modeling with lirscl, prelis, and simplis: Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum Associates.
- Chang, H. H. ve Ying, Z. (1999). A-stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, 23, 211-222.
- Chang, H. H., Qian, J. ve Ying, Z. (2001). A-stratified multistage computerized adaptive testing with blocking. *Applied Psychological Measurement*, 25, 333-341.
- Chang, S. W. ve Twu, B. Y. (1998). *A comparative study of item exposure control methods in CAT* (Research Report 98-3). Iowa City, IA: ACT.
- Chen, S. Y., Ankenmann, R. D. ve Spay, J. A. (1999). *Exploring the relationship between item exposure rate and test overlap rate in computerized adaptive testing* (ACT Research Report Series. 99-5). http://www.act.org/content/dam/act/unsecured/documents/ACT_RR99-05.pdf adresinden erişildi.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.
- Cömert, M. (2008). *Bireye uyarlanmış bilgisayar destekli ölçme ve değerlendirme yazılımı geliştirilmesi* (Yayımlanmamış yüksek lisans tezi). Bahçeşehir Üniversitesi, İstanbul.
- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- Çetin, E. (2009). *Dikey ölçekte klasik test ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Davey, T. ve Parshall, C. G. (1995, April). *New algorithms for item selection and exposure control with computerized adaptive testing*. Annual meeting of the American Educational Research Association (AERA) sunulmuş bildiri, San Francisco, CA.
- Davis, J. P., Eisenhardt, K. M. ve Bingham, C. B. (2007) Developing theory through simulation methods. *Academy of Management Review*, 32(2), 480-499.
- Davis, L. L. (2002). *Strategies for controlling item exposure in computerized adaptive testing with polytomously scored items* (Yayımlanmamış doktora tezi). The University of Texas, Austin.

- 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.
- Davis, L. ve Dodd, B. (2003). Item exposure constraints for testlets in the verbal reasoning section of the MCAT. *Applied Psychological Measurement*, 27(3), 335-356.
- Davis, L. L. ve Dodd, B. G. (2005). *Strategies for controlling item exposure in computerized adaptive testing with partial credit model* (PEM Research Report No. 05-01). Austin, TX: Pearson Educational Measurement.
- DeMars, C. E. (2006). Application of the Bi-Factor multidimensional item response theory model to Testlet-Based tests. *Journal of Educational Measurement*, 43(2), 145-168.
- Eggen, T. J. H. M. (2001). *Overexposure and underexposure of items in computerized adaptive testing* (Measurement and Research Department Reports. 2001-1). Arnhem, The Netherlands: CITO.
- Eignor, D. R., Stocking, M. L., Way, W. D. ve Steffen, M. (1993). *Case studies in computer adaptive test design through simulation* (ETS Research Report. RR-93-56). Princeton, NJ: Educational Testing Service.
- Embretson, S. E. ve Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Eroğlu, M. G. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Evans, J. J. (2010). *Comparability of examinee proficiency scores on computer adaptive tests using real and simulated data* (Yayımlanmamış doktora tezi). The State University of New Jersey, New Jersey.
- French, B. ve Thompson, T. (2003). *The evaluation of exposure control procedures for an operational CAT*. Annual meeting of the American Educational Research Association (AERA) sunulmuş bildiri, Chicago, IL.
- Georgiadou, E., Triantafillou, E. ve Economide, 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.
- Glas, G. A. W. (2002). Item calibration and parameter drift. W. J. Van Der Linden ve G. A. W. Glas (Ed.), *Computerized adaptive testing: Theory and practice* içinde (s. 183-199). New York: Kluwer Academic Publishers.
- Goldman, S. H. ve Raju, N. S. (1986). Recover of one and two parameter logistic item parameters: An empirical study. *Educational and Psychological Measurement*, 24, 11-21.
- Gökçe, S. (2012). *Comparison of linear and adaptive versions of the Turkish pupil monitoring system (pms) mathematics assessment* (Yayımlanmamış doktora tezi). Orta Doğu Teknik Üniversitesi, Ankara.
- Grubbs, F. E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics*, 15, 53-66.
- Gu, L. ve Reckase, M. D. (2007). Designing optimal item pools for computerized adaptive tests with Sympton-Hetter exposure control. D. J. Weiss (Ed.), *Proceedings of the 2007 GMAC Conference on Computerize Adaptive Testing* içinde. <http://www.psych.umn.edu/psylabs/CATCentral/> adresinden erişildi.
- Guyer, R. ve Thompson, N. (2011). *Item response theory parameter recovery using Xcalibre 4.1*. St. Paul MN: Assessment Systems Corporation.
- Guyer, R. ve Thompson, N.A. (2012). *User's manual for Xcalibre item response theory calibration software, version 4.1.6*. St. Paul MN: Assessment Systems Corporation.
- Hair, J., Anderson, R., Tatham, R. ve Black, W. (1998). *Multivariate data analysis*. New Jersey: Prentice Hall, Upper Saddle River.

- Hambleton, R. K., Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications Inc.
- Han, K. T. (2009). A gradual maximum information ratio approach to item selection in computerized adaptive testing. D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing* içinde. <http://www.psych.umn.edu/psylabs/CATCentral/> adresinden erişildi.
- Han, K. T. (2011). *User's manual: SimulCAT: Windows software for simulating computerized adaptive test administration*. ABD: Graduate Management Admission Council.
- Han, K. T. (2012). An efficiency balanced information criterion for item selection in computerized adaptive testing. *Journal of Educational Measurement*, 49(3), 225-246.
- Harwell, M. R. ve Janosky, J. E. (1991). An empirical study of the effects of small datasets and varying prior variances on item parameter estimation in BILOG. *Applied Psychological Measurement*, 15(3), 279-291.
- Harwell, M., Stone, C. A., Hsu, T. ve Kirişçi, L. (1996). Monte carlo studies in item response theory. *Applied Psychological Measurement*, 20(2), 101-125.
- Hooper, D., Coughlan, J. ve Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronical Journal of Business Research Methods*, 6(1), 53-60.
- Howell, C. D. (2010). *Statistical methods for psychology*. USA: Cengage Wadsworth.
- Huang, H. Y., Chen, P. H. ve Wang, W. C. (2012). Computerized adaptive testing using a class of high-order item response theory models. *Applied Psychological Measurement*, 36(8), 689-706.
- Hulin, C. L., Lissak, R. L. ve Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A monte carlo study. *Applied Psychological Measurement*, 6(3), 249-260.
- İşeri, A. I. (2002). *Assessment of students' mathematics achievement through computer adaptive testing procedures* (Yayımlanmamış doktora tezi). Orta Doğu Teknik Üniversitesi, Ankara.
- Jöreskog, K. ve Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software.
- Kalender, İ. (2009). Başarı ve yetenek kestirimlerinde yeni bir yaklaşım: Bilgisayar ortamında bireyselleştirilmiş testler (Computerized adaptive tests-CAT). *CITO Eğitim Kuram ve Uygulama*, 5, 39-48.
- Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability* (Yayımlanmamış doktora tezi). Orta Doğu Teknik Üniversitesi, Ankara.
- Kaptan, F. (1993). *Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Kaskatı, O. T. (2011). *Rasch modelleri kullanarak romatoid artirit hastaları özürüllük değerlendirimi için bilgisayar uyarlamalı test yöntemi geliştirilmesi* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. United States of America: Sage Publications.
- Kezer, F. (2013). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Publications.
- Köse, İ. A. (2010). *Madde tepki kuramına dayalı tek boyutlu ve çok boyutlu modellerin test uzunluğu ve örneklem büyüklüğü açısından karşılaştırılması* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.

- Lee, H. Y. ve Dodd, B. G. (2012). Comparison of exposure controls, item pool characteristic, and population distributions for CAT using the partial credit model. *Educational and Psychological Measurement*, 72(1), 159-175.
- Leroux, A. J., Lopez, M., Hembry, I. ve Dodd, B. G. (2013). A comparison of exposure control procedures in CATs using the 3PL model. *Educational and Psychological Measurement*, 73, 857-874.
- Lord, F. ve Stocking, M. (1988). Item response theory. J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* içinde. New York: Pergamon Press.
- Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28(4), 989-1020.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lunz, M. E. ve Stahl, J. A. (1998). *Patterns of item exposure using a randomized CAT algorithm*. Annual meeting of the National Council on Measurement in Education sunuşmuş bildiri, San Diego, CA.
- Magis, D. ve 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.
- McDonald, P. L. (2002). *Computer adaptive test for measuring personality factors using item response theory* (Yayımlanmamış doktora tezi). The University Western of Ontario, London.
- Özbaşı, D. (2014). *Bilgisayar okuryazarlığı testinin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliğine ilişkin bir araştırma* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Öztuna, D. (2008). *Kas-iskelet sistemi sorunlarının özürüllük değerlendiriminde bilgisayar uyarlamalı test yönteminin uygulanması* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Maidenhead: Open University Press/McGraw-Hill.
- Pastor, D. A., Dodd, B. G. ve 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. ve Gessaroli, M. E. (1995, April). *A comparison of item parameter estimates and ICCs produced with TESTGRAF and BILOG under different test lengths and sample sizes*. Annual meeting of the National Council on Measurement in Education sunulmuş bildiri, San Francisco.
- R Core Team. (2017). *R: A language and environment for statistical computing*. <https://www.R-project.org/>. adresinden erişildi.
- Revue, J. ve Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(4), 311-327.
- Ross, S. M. (2013). *Simulation*. USA: Elsevier.
- Rudman, H. C. (1987). The future of testing is now. *Educational Measurement: Issues and Practice*, 4, 5-11.
- Scullard, M. G. (2007). *Application of item response theory based computerized adaptive testing to the strong interest inventory* (Yayımlanmamış doktora tezi). University of Minnesota, USA.
- Segall, D. O. (2004). Computerized adaptive testing. K. Kempf-Leanard (Ed.), *The encyclopedia of social measurement* içinde (s. 429-438). San Diego, CA: Academic Press.
- Smits, N., Cuijpers, P. ve Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry Research*, 188, 145-155.
- Speron, E. (2009). *A comparison of metric linking procedures in item response theory* (Yayımlanmamış doktora tezi). Chicago, Illinois: Illinois Institute of Technology.
- Stocking, M. L. ve Lewis, C. (1995). *A new method of controlling item exposure in computerized adaptive testing* (Research Report. 95-25). Princeton, NJ: Educational Testing Service.

- Stocking, M. L. ve Lewis C. (2002). Methods of controlling the exposure of items in CAT. W. J. Van Der Linden ve G. A. W. Glas (Ed.), *Computerized adaptive testing: Theory and practice* içinde (s. 163-182). New York: Kluwer Academic Publishers.
- Sulak, S. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında kullanılan madde seçme yöntemlerinin karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Sümer, N. (2000). Yapısal eşitlik modelleri: Temel kavramlar ve örnek uygulamalar. *Türk Psikoloji Yazıları*, 3(6) 49-74.
- Şahin, A. ve Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice*, 17, 321-335.
- Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlik*. Ankara: Seçkin Yayıncılık.
- Şimşek, Ö. F. (2007). *Yapısal eşitlik modellemesine giriş: Temel ilkeler ve LISREL uygulamaları*. Ankara: Ekinoks Yayınları.
- Tabachnick, B. G ve Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Allyn and Bacon.
- Tang, K. L., Way, W. D. ve Carey, P. A. (1993). *The effect of small calibration sample sizes on TEOFL IRT-based equating* (TOEFL technical report. TR-7). Princeton, NJ: Educational Testing Service.
- Tay, P. H. (2015). *On-the-fly assembled multistage adaptive testing* (Yayımlanmamış doktora tezi). The University of Texas, Urbana.
- Thissen, D. ve Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika*, 47(4), 397-412. doi:10.1007/BF02293705
- Thompson, T. (2002, April). *Employing new ideas in CAT to a simulated reading test*. Annual Meeting of the National Council on Measurement in Education sunulmuş sözlü bildiri, New Orleans, LA.
- Vaughn, B. K. ve Wang, Q. (2010). DIF trees: using classifications trees to detect differential item functioning. *Educational and Psychological Measurement*, 70(6), 941-952.
- Veldkamp, B. P. ve Van Der Linden, W. J. (2008). Implementing Sympson-Hetter item-exposure control in a shadow-test approach to constrained adaptive testing. *International Journal of Testing*, 8(3), 272-289.
- Veldkamp, B. P. ve Van Der Linden W. J. (2010). Designing item pools for adaptive testing. W. J. Van Der Linden ve C. A. W. Glas (Ed.), *Elements of adaptive testing* içinde. New York: Springer.
- Veldkamp, B. P., Verschoor, A. J. ve Eggen, T. J. H. M. (2010). A multiple objective test assembly approach for exposure control problems in computerized adaptive testing. *Psicológica*, 31(2), 335-355.
- Wang, H. P., Kuo, B. C., Tsai, Y. H. ve Liao, C. H. (2012). A cefr-based computerized testing system for chinese proficiency. *The Turkish Journal of Educational Technology*, 11(4), 1-12.
- Wang, T. ve Vispoel, W. P. (1998). Properties of ability estimation methods computerized adaptive testing. *Journal of Educational Measurement*, 35(2), 109-135.
- Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice*, 17, 17-27.
- Way, W. D. (2005). *Practical questions in introducing computerized adaptive testing for k-12 assessments* (PEM Research Report. 05-03). Londra: Pearson Educational Measurement.
- Weiss, D. J. (1988). Adaptive testing. J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* içinde (s. 881-884). New York: Pergamon Press.
- Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37(2), 70-84.
- Weiss, D. J. ve Guyer, R. (2012). *Manual for CATSim: Comprehensive simulation of computerized adaptive testing*. St. Paul MN: Assessment Systems Corporation.

- Weiss, D. J. ve Von Minden, S. (2012). *A comparison of item parameter estimates from Xcalibre 4.1 and Bilog-MG*. St. Paul MN: Assessment Systems Corporation.
- Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*, 52(2), 275-291. doi:10.1007/BF02294241
- Yen, W. M. (1993). Scaling performance assessment: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187-213.
- Yoes, M. (1995). *An updated comparison of microcomputer-based item parameter estimation procedures used with the 3-parameter IRT model*. Saint Paul, MN: Assessment Systems Corporation.
- Zheng, Y. ve Chang, H. H. (2014). On-the-fly assembled multistage adaptive testing. *Applied Psychological Measurement*, 39(2), 105-118.
- Zitny, P., Halama, P., Jelinek, M. ve Kveton, P. (2012). Validity of cognitive ability tests-comparison of computerized adaptive testing with paper and pencil and computer-based forms of administrations. *Studia Psychologica*, 54(3), 181-194.

