

BİLGİSAYAR ORTAMINDA BİREYE UYARLANMIŞ TEST YÖNTEMİ

Dr. Fatih Kezer
Kocaeli Üniversitesi

Özet

Ölçme ve değerlendirme süreci, eğitim ve öğretimle kazandırılması amaçlanan özelliklerin gerçekleşip gerçekleşmediğinin ortaya konmasında kritik rol oynamaktadır. Ölçme ve değerlendirme amacıyla özellikle Türkiye’de ağırlıklı olarak Klasik Test Kuramı’na dayalı geleneksel kâğıt kalem testleri kullanılmaktadır. Klasik Test Kuramı’nda testten kestirilen madde ve yetenek parametrelerinin yine testin geliştirildiği gruba bağımlı olması, yapılan testlerin geçerliği ve güvenilirliğinin uygulandıkları gruba bağlı olması sınırlılığını doğurmaktadır. Bu sınırlılığı ortadan kaldırarak değişmezlik özelliği ile madde ve yetenek parametrelerinin gruptan bağımsız olması ve testten elde edilen puanların daha karşılaştırılabilir olmasını sağlayan Madde Tepki Kuramı’na dayanan modeller dünyada giderek yaygınlaşmaya başlamıştır. Bilgi ve teknolojinin hızlı gelişimiyle birlikte özellikle Amerika Birleşik Devletleri, Hollanda ve Uzakdoğu Ülkeleri gibi bazı ülkelerde Madde Tepki Kuramı’na dayanan testlerin kullanımı hızla artmıştır. Özellikle geniş ölçekli testlerde bu modellerin kullanımının arttığı gözlenmektedir. Dünya’da ölçme yaklaşımlarının değişimine göz atıldığında gelişen ölçme ve değerlendirme kuramlarına ayak uydurmanın yanında gelişen teknolojiye de ayak uydurulduğu görülmektedir. Dünya genelinde kullanılan Madde Tepki Kuramı’nın bir uygulaması olan Bilgisayar Ortamında Bireye Uyarlanmış Test uygulamaları buna en güzel örnek olarak değerlendirilebilir. Bu test yöntemi ile bireylerin yetenek düzeylerine uygun, daha kısa, geçerli ve güvenilir test geliştirmek mümkün olabilmektedir. Teknolojik gelişmeler ışığında Türkiye’de geniş ölçekli testlerin gözden geçirilmesi, Madde Tepki Kuramı modelleri doğrultusunda geçerli ve güvenilir, bireylerin yetenek düzeylerine hitap eden az sayıda madde ile oluşturulmuş bireye uyarlanmış testlerin kullanımının yaygınlaştırılması beklenmektedir. Bu çalışmanın amacı, Bilgisayar Ortamında Bireye Uyarlanmış Test Yöntemi’nin daha iyi anlaşılması adına bu test yapısını birçok yönüyle ortaya koymaktır.

Anahtar Kelimeler

Madde tepki kuramı, Bilgisayar ortamında bireye uyarlanmış test, Test etme.

COMPUTERIZED ADAPTIVE TESTING METHOD

Dr. Fatih Kezer
Kocaeli University

Abstract

Measurement and evaluation process has a critical role in revealing whether aims of education are achieved or not. Mostly, traditional paper and pencil tests basing on Classical Test Theory are used for measurement and evaluation, especially in Turkey. In Classical Test Theory, since item and ability parameters estimated from the tests are also dependent to the target group, it causes limitation on reliability and validity of tests related to the target group. In order to get over this limitation, models which base on Item Response Theory; make item and ability parameters be independent from the group invariance property and provide points got from the tests to become more comparable become more widespread all over the world. With the rapid development of information and technology, especially in some countries like United States of America, Netherlands and Far East Countries, usage of these tests basing on Item Response Theory increased rapidly. More particularly, increase in usage of these models at large scale tests is observed. When looked at the changes in measurement approaches in the world, it is seen that in addition to keeping up with developing measurement and evaluation theories, they also keep up with developing technology. Therefore, Computerized Adaptive Testing which is a kind of application of Item Response Theory used all over the world can be a very good example for this. With the help of this test method, it can be possible to develop shorter, more reliable and valid tests that are also appropriate for individuals' ability levels. In the light of technological developments, it is expected that large scale tests in Turkey will be reviewed and usage of adaptive tests that are valid and reliable according to Item Response Theory; appropriate for individuals' ability levels; composed of few items will become widespread. For this reason, this study aims to reveal different aspects of this kind of tests in order to better understand Computerized Adaptive Testing.

Keywords

Item response theory, Computerized adaptive test, Testing

GİRİŞ

Eğitimle kazandırılmak istenen yeni davranışların beklenen düzeyde kazandırılmış olup olmadığına karar verebilmek için, öncelikle bu davranışların, geçerliği ve güvenirliliği yeterli derecede yüksek olan ölçme araçlarıyla ölçülmesine ihtiyaç vardır. Bir kişinin belli davranışlarda erişmiş olduğu yetkinlik ve kararlılık nesnel bir biçimde belirlenmedikçe, bu kişinin ilgili davranışlarda beklenen düzeye erişmiş olup olmadığına karar verilemez (Özçelik, 2010). Uygulanan bir eğitimin başarılı olup olmadığına, başarılı ise ne derecede ve hangi öğrenciler için başarılı olduğunun bilinmesi istenir. Eğitim işlemi devam ederken, başarısızlığın ve başarısız bireylerin erken tanınması, önlem alınmasını kolaylaştırır. Başarı derecesinin bilinmesi ve başarısızlık durumlarının ortaya çıkarılması, ileride girilecek benzer eğitim etkinliklerinin daha gerçekçi esaslarla planlanmasına yardım eder (Turgut,1992; Özçelik, 2010; Turgut ve Baykul, 2010). Bu bağlamda ölçme ve değerlendirme süreci eğitimle ilgili alınacak kararlarla ilgili daha sistematik ve nesnel kanıtlar elde edilmesini sağlar (Linn ve Gronlund, 1995).

Eğitim ve psikoloji alanlarında ölçme sonuçlarından yola çıkarak doğru yargılara varılmasında ve ölçme eylemine karışan hatanın en aza indirilmesinde önemli role sahip olan nitelikli ölçme araçları geliştirmek amacıyla çeşitli kuramlar geliştirilmiştir (Özçelik, 2011; Wilson, 2005). Eğitim ve psikoloji alanında en çok kullanılan kuramlar “Klasik Test Kuramı (KTK)” ve “Madde Tepki Kuramı (MTK)” dır.

1905 yılında Spearman tarafından temelleri atılan KTK, gerçek puanı, ölçme sonuçlarından elde edilen puanla tahmin etmeye çalışan bir kuramdır. KTK'nın en büyük avantajı, varsayımlarının test edilebilmesinin kolay olmasıdır. Böylelikle kuramın varsayımları çoğu durumda kolaylıkla test edilebilmekte ve uygulamalara uyarlanabilmektedir (Hambleton ve Jones, 1993). Varsayımların kolay karşılanması, madde parametrelerinin kolay kestirilmesinden dolayı günümüzde hala kuram test geliştirme sürecinde MTK'ya göre daha çok tercih edilmektedir (Kelecioğlu, 2001). KTK'da madde güçlük parametresi ve madde ayırt edicilik parametresi verilerin toplandığı gruba (örnekleme) bağlı olarak değişmektedir (Lord ve Novick, 1968). Maddelerin güçlüğü grubun yetenek düzeyine bağlı, grubun yetenek düzeyi ise maddelerin güçlüğüne bağlı olarak yorumlanmaktadır. Bir birey, aynı özelliği ölçen farklı zorluktaki ölçme araçlarından farklı puanlar alır. Eğer bir test zor ise kişi düşük yeteneğe sahipmiş gibi, eğer test kolay bir test ise testi alan kişi yüksek yetenek düzeyine sahipmiş gibi görünebilir. Bu nedenle bireylerin karşılaştırılması sadece aynı ya da birbirine yakın/benzer ölçme araçları kullanıldığında mümkün olabilmektedir. Böyle bir durumda bile testi alanların yetenek düzeyleri birbirinden farklı olacağı için test puanları farklı büyüklüklerde hataya sahip olabilmektedir (Hambleton,

Swaminathan ve Rogers, 1991). KTK ile geliştirilen ölçme araçları, genellikle orta yetenek düzeyindeki bireylere hitap eder. Bu durumda, yeterli sayıda uygun madde olmadığı için alt ve üst yetenek düzeyindeki bireylerin, yetenekleri güvenilir bir şekilde kestirilemez. KTK'da, ölçme hatasına ait varyansın tüm bireyler için eşit olduğu varsayılmaktadır. Dolayısıyla bir örneklemedeki yetenek düzeyi maddenin güçlüğünü, bununla birlikte maddenin ayırtediciliğini doğrudan etkilemektedir. Bu bağlamda bir test, geliştirildiği grubun yetenek düzeyine paralel bir grupta daha iyi işleyecek ve böyle bir grupta daha kullanışlı olacaktır. KTK'da testin güvenilirliği de yine testin geliştirildiği gruba bağlıdır. Bir grupta geçerli ve güvenilir olan bir test başka bir grupta geçerli ve güvenilir olmayabilmektedir (Crocker ve Algina, 1986). KTK'nın bir başka dezavantajı teste yönelik oluşudur. Kuram kişinin ya da bir grubun verilen bir madde üzerinde nasıl bir performans göstereceğine ilişkin kestirim olanağı sağlamaz (Hambleton ve Swaminathan, 1989). KTK'nın bu gibi sınırlılıklarını giderdiği düşünülen bir model olan MTK 1930'lu yıllarda ortaya konmuştur. MTK güçlü varsayımları ve dayandığı matematiksel modeller nedeniyle KTK'nın kuramsal yetersizliklerini giderdiği iddiasındadır. MTK, bireyin ölçülen özelliğe ilişkin sahip olduğu yetenek düzeyi ve verdiği cevaplar arasındaki ilişkiyi matematiksel bir fonksiyon ile açıklayan bir modeldir (Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1989). Kuram, bireyin testle ölçüldüğü düşünülen gözlenemeyen örtük özellikleri ile bireyin testte gösterdiği tepki örüntüsü arasındaki doğrusal ilişkiyi ortaya koymaya çalışmaktadır. KTK'ya göre daha güç karşılanabilen varsayımlara sahiptir. Bir ölçme aracında ölçülmek istenen özelliğin tek olması anlamına gelen tek boyutluluk ve test maddelerinin birbirinden bağımsız olmasını ifade eden yerel bağımsızlık varsayımları kuramın önemli iki varsayımdır.

MTK'nın değişmezlik özelliği ve tanımlanan madde karakteristik eğrisi, matematiksel altyapısıyla beraber birçok avantaja sahiptir. Kuram, test ve madde parametrelerinin, uygulandığı gruptan bağımsız olması; bireylerin yetenek düzeylerinin (0) ise uygulanan ölçme aracından bağımsız olması iddiasıyla KTK'nın önemli bir sınırlılığını ortadan kaldırmaktadır. Böylelikle farklı gruplara ait ölçme sonuçlarının karşılaştırılabilmesi mümkün olmaktadır. MTK'nın her birey için ayrı ayrı hatayı kestiriminde bulunabilmesi ve her madde/yetenek düzeyi için ayrı ayrı güvenilirlik -test ve madde bilgi fonksiyonu aracılığı ile sunulabilmesi kuramın önemli avantajlarından (Adams, 2005; Çelen, 2008; Yurdugül, 2006). MTK'nın test geliştirme, madde haritalama, test eşitleme, test ve madde yanlılığı gibi birçok uygulamasından bir tanesi de bilgisayar ortamında bireye uyarlanmış test (BOBUT) uygulamasıdır. Teknoloji ve ölçme alanındaki gelişmelerle birlikte BOBUT önem kazanmış ve getirdiği avantajlar ile birlikte eğitim bilimlerinde yeni uygulama alanları bulmuştur.

BİREYE UYARLANMIŞ TEST

MTK pek çok açık avantaja sahip olmasına rağmen, asıl gücünü maddelerin bağımsız ele alınmasından almaktadır. Kuram, içinde maddelerin en kolaydan en zora doğru dizildiği temel bir özellik ortaya koymaktadır. Test etmenin amacı aynen maddeler gibi katılımcıları da yetenek düzeyi düşük olandan yüksek olana doğru bir düzlem üzerinde sıralamaktır. Diğer bir ifade ile tüm maddelerin tüm bireylere sunulması yerine, sadece katılımcının bu düzlem üzerinde yerlerinin doğru bir biçimde belirlenmesine yetecek kadar madde yeterlidir. Tüm katılımcıları aynı doğru üzerinde sıralamak her bir test katılımcısı için bir testin özel olarak uyarlanabilmesi olasılığını ortaya çıkarmıştır. Bu türden bir teste *bireye uyarlanmış/bireyselleştirilmiş* denmektedir. *Bireye uyarlanmış test (BUT)*, bireyin ölçülen özellik düzeyine bağlı olarak, her bir bireye farklı test soru (madde) takımlarının uygulandığı testlerdir. BUT, testi alan herkese aynı soru kümesinin uygulandığı *geleneksel testlerden* (örneğin tipik kalem kağıt testi) tamamen farklıdır (Weiss, 1985; Wainer ve diğ., 2000; Ware ve diğ., 2003).

Geleneksel test yapısında, yetenekli test katılımcıları, gerçek yeteneklerini göstermelerini sağlayacak zor maddelere gelene kadar basit pek çok şeyle uğraşmak zorunda kalmaktadırlar. Böyle bir durumda ise zamanın ve çabanın boşa harcanması, sıkılmanın yol açacağı dikkatsizlik gibi birçok faktör ölçme sürecine dahil olmaktadır. Ters durumda yine benzer şekilde, daha az yetenekli katılımcılara hitap eden maddeler nispeten kolay maddelerdir. Zor maddeler ise bu bireylerle ilgili çok az bilgi sağlamaktadır. Bununla birlikte zor maddeler, kafa karışıklığı, umutsuzluk, motivasyonda azalma vb. birçok duruma yol açabilir. Şansla doğru cevabı bulma da diğerleri ile birlikte bir dışsal değişken olarak devreye girmektedir (Sands, Waters ve McBride, 1997; Wainer ve diğ., 2000).

BUT'ta, testi alan bireyin cevaplarından yola çıkılarak, güçlük ve ayırt edicilik parametreleri bilinen maddeleri içeren madde havuzundan bireyin yetenek düzeyine uygun maddeler seçilmektedir. Yetenek düzeyi yüksek olan bir bireyin testinde, daha az yetenek düzeyine sahip olan başka bir bireyin aldığı test ile ya çok az ortak madde bulunacak yada hiçbir ortak madde bulunmayacaktır. Geleneksel bir testte doğru yanıt sayısı gibi bir ölçüt kullanışlı iken BUT'ta bu kullanışlı olmamaktadır. Yetenek düzeyi yüksek olan bir birey ile yetenek düzeyi düşük olan bir birey aynı testte aynı doğru yanıt sayısına sahip olsa bile yetenek düzeyleri çok farklı çıkabilmektedir.

BUT'un genel prensipleri ilk olarak Alfred Binet ve meslektaşları tarafından 1900'lerin başında Fransa'da geliştirilen ve daha sonra İngilizce konuşulan ülkelerde Stanford-Binet Zekâ Testi olarak kullanıma sunulan Binet zekâ testinde uygulanmıştır (Weiss, 1985; Turgut ve Baykul, 2010). 1900'lerin başındaki Binet'nin BUT'unu takiben, 1950'lerin sonuna kadar hiç başka BUT ortaya çıkmamıştır (Weiss, 1988). Bu testlerin tamamı deneme amaçlı olarak ve yalnızca bireye uyarlanmış testlere çeşitli yaklaşımların özelliklerini çalışmak için

kullanılmışlardır (Weiss, 1985). 1960'larda Lord çalışmalarıyla, uyarlanmış testlere önemli katkılar sağlamıştır. Bilgisayar ortamında uygulanmayan bir bireye uyarlanmış test olan *kendi kendine puanlanabilen esnek düzeyli test (self-scoringflexilevel test)* Lord tarafından 1971 yılında tanımlanmıştır.

BİLGİSAYAR ORTAMINDA BİREYE UYARLANMIŞ TEST

Bireye uyarlanmış testlere ilişkin bazı yaklaşımlar, test maddelerinin interaktif bir bilgisayar ortamında belli algoritmalarla uygulanmasını şart koşmaktadır. Bireye uyarlanmış test yapmanın imkânlarını sonuna kadar kullanan bu son yaklaşım "*Bilgisayar Ortamında Bireye Uyarlanmış Test Yöntemi (BOBUT)*" olarak adlandırılmaktadır (Weiss, 1985; Wainer ve diğ., 2000). Bilgisayar ortamında uygulanabilen bireye uyarlanmış test, klasik testlerden farklı olarak katılımcıların bilgisayarlarla yönetildiği ölçme aracıdır (Embretson ve Reise, 2000). Bilgisayar kullanımının yaygınlaşması ile birlikte BOBUT kavramı yaygınlaşmış ve uygulama alanları genişlemiştir.

Bilgisayar Ortamında Bireye Uyarlanmış Test Yöntemleri

BOBUT, maddelerin geniş bir havuzdan seçilerek bireye uygulanmasında farklı yöntemler kullanmakta olup, bunlar aşağıda özetlenmiştir (Thorndike, 1982; Hambleton ve Swaminathan, 1989; Kaptan, 1993; akt: Demirtaşlı, 1999).

1. İki Aşamalı Uygulama (Twostagetesting)
2. Kendinden Seçimli Uygulama (Self-selectingtesting)
3. Piramitsel Çok Aşamalı Uygulama (PyramidalMultistageTesting)
4. Değişimlemeli Uygulama (AlternatingTesting)
5. Tabakalı Uygulama (StradaptiveTesting)
6. Çok Düzeyli Uygulama (Multilevel Format)

Bu bireye uyarlanmış testlerin ilki kâğıt kalemle uygulanmak üzere geliştirilmiş ve değişken başlangıç ve sonlandırmanın avantajlarını hiçe sayarak görece olarak basit dallanma kurallarını kullanmıştır.

İki-aşamalı bireye uyarlanmış test (two-stageadaptive test), en düşük uyarlama kapasitesi olan mekanik bir dallandırma yöntemidir. Teste tabi tutulan herkes yönlendirme (routing) testi denilen kısa bir test alırlar ki bu genellikle ortalama güçlükte bir testtir. Yönlendirme testindeki puanlarına dayanarak, teste tabi tutulanlar, ikinci-aşama bir ölçme (measurement) testine yönlendirilirler (Glas ve Linden, 2003). Bu test, yetenek düzeylerine kabaca uyarlanmıştır (Weiss, 1985; Linacre, 2000).

İki aşamalı uygulamada, ilk aşamada alınan testin puanlamasında yaşanan mantıksal problemler ve testi alan bireylerin ikinci aşamada alacakları teste

yönlendirilmesindeki sıkıntılar nedeniyle *kendinden seçimli bireye uyarlanmış test* yöntemi önerilmiştir. Bu yöntemde göre testte yer alan maddeler güçlüklerine göre sıralanmaktadır. Testi alan birey kendi yaş ve sınıfı düzeyindeki ortalama güçlükteki bir madde ile teste başlamaktadır. Birey maddeyi doğru yanıtlarsa daha zor bir madde, yanlış yanıtlarsa daha kolay bir maddeye geçilir. Test uygulaması, birey hakkında yeterli bilgiye ulaşıncaya kadar devam etmektedir (Thorndike, 1982; Hambleton ve Swaminathan, 1989; akt: Demirtaşlı, 1999).

Piramitsel bireye uyarlanmış test, piramiti andıran bir yapıda güçlük düzeyine göre önceden yapılandırılmış maddeler toplamını içermektedir. Piramidin tepesinde ortalama güçlükte bir madde vardır. Bir sonraki aşamada iki madde bulunmaktadır, bunlardan biri ilk maddeden biraz daha zor bir maddedir, diğeri ise ilk maddeden biraz daha kolaydır. Takip eden aşamalarda, her madde iki ek maddeye yönlendirmektedir. Bu şekilde, maddelerin tümünü ya da çoğunu doğru cevaplayan bireye artan güçlük düzeyinde maddeler uygulanmaya devam edecekken, çoğu maddeyi yanlış cevaplayan bireye azalan güçlük düzeyinde maddeler uygulanmaya devam edecektir. Piramitsel testin daha az sayıda madde kullanan bir varyasyonu flexilevel (esnek-düzeyle) testtir. Bu test yapma yöntemi, çok kolaydan çok zora değişen, eşit mesafeli güçlüklerde madde içeren bir maddeler serisidir. Her bir güçlük seviyesinde sadece bir madde olduğu için, her güçlük seviyesinde birçok madde olan piramitsel testin aksine, yüksek yetenekli biri havuzdaki en zor maddeleri alacaktır ve düşük yetenekli biri havuzdaki daha düşük zorluktaki maddelerin %50'sini alacaktır. Esnek-düzeyle, piramitsel test ve de iki-aşamalı test ile onun varyasyonları sabit başlangıç ve sabit sonlandırma kuralı kullanmaktadır (Weiss, 1985).

Bir başka bireye uyarlanmış test yöntemi olan *değişimlemeli test* uygulamasında testte yer alan maddeler iki kol halinde düzenlenmiştir. Kollardan birinde kolay maddeler, diğerinde zor maddeler sıralanmıştır. Test orta güçlükte bir madde ile başlamakta ve madde doğru yanıtlanırsa zor madde, yanlış yanıtlanırsa kolay madde koluna geçilmektedir. Bireyin karşısına çıkan maddeler doğru yanıtlandığı sürece aynı kolda test devam ettirilmektedir. Doğru yanıtlar devam ederken yanlış yanıt verilirse kolay madde koluna, yanlış yanıtlar devam ederken doğru yanıt verilirse zor madde koluna geçilmektedir (Thorndike, 1982; Hambleton ve Swaminathan, 1989; akt: Demirtaşlı, 1999).

Binet testlerine benzer olarak, değişken başlangıç ve sonlandırmayı içeren, *sabit tabakalı bireye uyarlanmış test* olarak adlandırılan sabit dallanmalı bir yöntem geliştirilmiştir (Weiss, 1988). Binet testleri gibi bu test, farklı güçlük düzeylerindeki maddeler alt kümelerine tabakalandırılmış maddeler havuzuyla hareket etmektedir. Yine Binet testlerine benzer olarak, test herhangi bir zorluk seviyesinde başlayabilmekte; ancak Binet testlerinden farklı olarak, madde seçimi her bir madde uygulandıktan sonra yapılmaktadır. Bu test de Binet testleri gibi bir tavan katmanına erişilene kadar, yani bireyin hiçbir maddeyi doğru

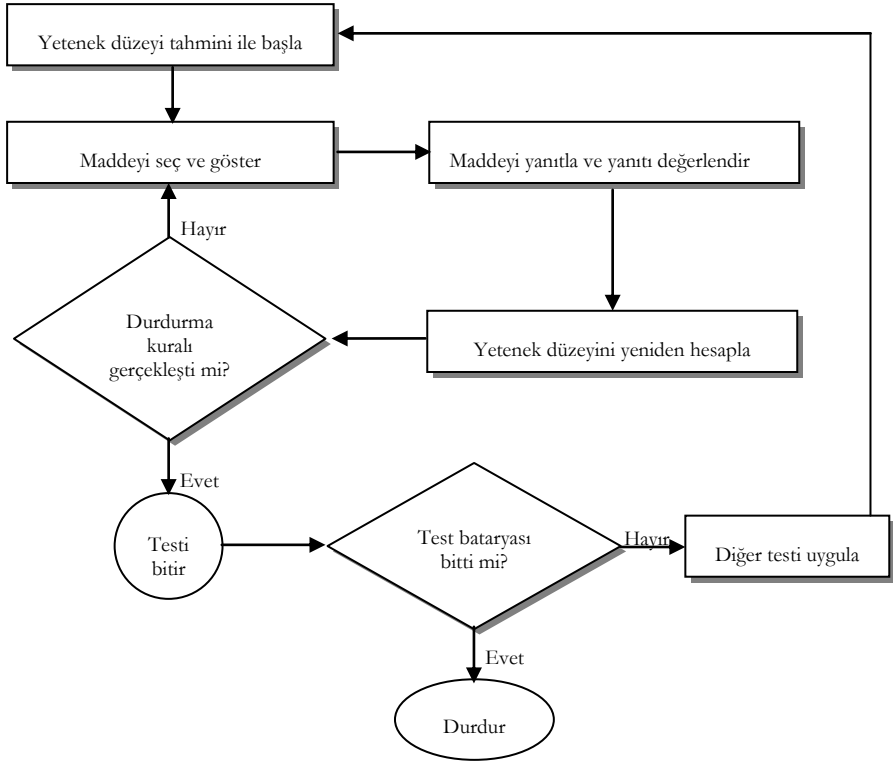
yanıtlamadığı bir zorluk düzeyine erişilene kadar madde madde dallanma kuralını kullanmaktadır.

Tabakalı bireye uyarlanmış testler, etkili sonlandırma kurallarının kullanımına olanak sağlamaktadır. Bireyin tavan seviyesini belirlemek için herhangi bir seviyede var olan tüm maddelerin uygulanmasını istemektense, daha dinamik bir sonlandırma kuralı kullanılmaktadır. Bu yaklaşımda, doğru yanıtlanmış maddelerin oranı her zorluk seviyesi için her bir madde uygulandıktan sonra güncellenmektedir. Testin sonlandırılması, bir zorluk seviyesinde belli sayıda madde uygulandıktan sonra doğru cevap oranları belirlenmiş bir orandan daha az ya da ona eşit olduktan sonra gerçekleşmektedir. Bu sonlandırma kuralını kullanan çoğu katmanlı bireye uyarlanmış testler, test uygulama süresini ortalama %50 oranında azaltmaktadır (Vale ve Weiss, 1975; Thompson ve Weiss, 1980).

Bir başka test yöntemi olan *çok düzeyli bireye uyarlanmış test yönteminde* başlangıç düzeyinin seçimi var olan bilgiye dayandırılmaktadır. Başlangıç düzeyi, sınıf, yaş veya önceki teste göre ya da sınıf içinde gösterdiği performansa göre seçilmektedir. Bu stratejide alt testler kolaydan zora doğru dizilmiştir. Böylelikle her düzeyden giderek daha zorlaşan 8–10 test birey tarafından yanıtlanmış olmaktadır. Her bir test tamamıyla yanıtlandığında bir sonraki teste geçilmektedir (Thorndike, 1982; Hambleton ve Swaminathan, 1989; akt: Demirtaşlı, 1999).

Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulaması

Bireye uyarlanmış testte daha önce belirtildiği gibi amaç, bir testte bireyin karşısına en uygun maddeyi çıkarmak ve yetenek düzeyini (θ) bu doğrultuda belirlemektir. Bir yetenek düzlemi boyunca her hangi bir noktada bireyler hakkında bir maddenin sağladığı bilgi sadece bu maddenin parametrelerine bağlıdır (Bejar 1983; Folk ve Smith, 2002; Hambleton, Swaminathan ve Rogers 1991; Öztuna, 2008). Bir bireyin yetenek düzeyi bilinerek bu birey için madde parametreleri kullanılarak en fazla bilgi veren madde belirlenebilir. Dolayısıyla bireye ilişkin bir θ tahmininden sonra, bu θ için en çok bilgiyi sağlayan madde belirlenip bireyin karşısına çıkarılmaktadır. Yeni θ kestiriminde sonra tekrar yeni bir madde belirlenmekte ve testin sonlandırılması kararına kadar döngü devam etmektedir. BOBUT yönteminin algoritması ana hatlarıyla Şekil 1'de gösterilmiştir (Wainer ve diğ., 2000).



Şekil 1. Bilgisayar ortamında bireye uyarlanmış test yönteminin algoritması

BOBUT'ta öncelikle bireyin yetenek düzeyine ilişkin bir tahminde bulunulur. Bu öncül test ya da bireyin daha önceki performanslarına dayalı olabilmektedir (Segall, 2003). Birey bazında daha önceki performanslara göre yetenek tahmini yapılabileceği gibi, grup ortalaması alınarak her birey için aynı başlangıç yetenek tahmini de zaman zaman kullanılmaktadır. Parametreleri daha önceden belirlenmiş madde havuzundan bireyin yetenek düzeyine uygun madde seçilir ve teste başlanır. İlk madde seçimi kritiktir. Bireyin yetenek düzeyinden daha zor bir madde bireyin motivasyonunu kırabilecekken, daha kolay bir madde testi ciddi almamasına ve basit hatalara neden olabilecektir (Linacre, 2000; Tian, Miao, Zhu ve Gong, 2007). Bu nedenle orta güçlükte bir madde seçimi iyi bir yaklaşımdır. Seçilen maddeye verilen yanıtla göre birey puanı belirlenir ve yetenek düzeyi yeniden kestirilir. Eğer kestirime ilişkin standart hata yeterince küçük bir değere ulaşmışsa ya da her kestirimden elde edilen hatalar arasında fark önemsiz hale gelmişse ya da testi sonlandırmak için karar verilen soru sayısına ulaşılmışsa test sonlandırılır. Eğer test sonlandırılmıyorsa, tekrar havuzdan madde seçme basamağına dönülür ve bireyin yeteneğine ve son verdiği cevaba uygun olarak belli nitelikte soru havuzdan seçilerek bireyin

cevaplanması istenir. Testin güvenliğini sağlamak için, seçilen maddelerin havuzdan en yüksek bilgiyi verecek şekilde rastgele seçilmesi gereklidir (Green, Bock, Humphreys, Linn ve Reckase, 1984).Yeni madde seçiminde bir yöntem bireyin yetenek düzeyine uygun en yüksek bilgi veren maddenin seçilmesi iken bir başka bir yöntem ise madde güçlükleri kullanılarak yeni madde seçimidir (Weiss 1988; Hambleton, Swaminathan ve Rogers, 1991; Eggen ve Straetmans, 1996; Folk ve Smith, 2002; Kalender, 2004;).

Tüm yetenek ranjında iyi ölçme yapma amacına ulaşmak için, madde havuzunun yetenek ranjına geniş olarak yayılmış madde güçlüğü parametreleri ile ayırıcılığı yüksek çok sayıda madde içermesi gerekmektedir. Bu madde sayısının ne kadar olması gerektiği kesin olarak belirtilmemiş olmasına rağmen kabaca bir kestirimle ikili puanlanan veriler için 100 madde olduğu belirtilmektedir (Embretson ve Reise, 2000).BOBUT'un tatmin edici uygulamaları, bu özellikteki 100 madde ile sağlansa da 150-200 madde aralığında hazırlanmış iyi bir biçimde yapılandırılmış madde havuzları daha iyi sonuçlar sağlamaktadırlar. İkili puanlanan MTK modellerinin aksine çoklu puanlanan MTK modellerinde havuz için 25-30 maddenin yeterli olduğu belirtilmektedir (Dodd, 1990; Dodd ve DeAyala, 1994). Genel olarak, özellik seviyelerinde maddeler eşit olarak dağıtılınca, BOBUT, yüksek derecede ayırt edici çok sayıda madde içeren havuzlarla daha iyi sonuçlar vermektedir (Weiss, 1985). Araştırmalar kâğıt kalem testlerinde yer alan maddelerin çoğunun BOBUT'a aktarıldıktan sonra da zorluk seviyelerini koruduklarını göstermektedir (Tian, Miao, Zhu ve Gong, 2007).

BOBUT'un yetenek kestirimi algoritmasında testin ne zaman sonlandırılacağı yetenek düzeyinin doğru belirlenebilmesi için önemli unsurlardan biridir. Bireyin yetenek düzeyinin belirlenmesinde, yetenek düzeyine uygun soruların seçilmesi kadar ne kadar soru sorulacağıının da iyi belirlenmesi gerekmektedir. Testin olması gerekenden daha kısa olması durumunda birey yeterli sayıda soru ile karşılaşmayacağından dolayı yetenek düzeyi yanlış belirlenebilir. Kullanılan sonlandırma kurallarından bazıları; madde havuzunda soru kalmaması, sabit uzunluk tercih ediliyorsa testin maksimum uzunluğa ulaşması, bireyin verdiği her yanıt doğrultusunda kestirilen yetenek parametresine ilişkin standart hatanın belli bir keskinliğe ulaşması (kestirilen son iki standart hata arasındaki farkın 0.01'den düşük olması ya da standart hatanın kabul edilebilir bir değerden düşük olması -0.25, 0.20 gibi-) ve adayın test dışı davranışlar sergilemesi olarak sıralanabilir (Hambleton ve diğ., 1991; Segall, 2003; Way, 2006; Tian ve diğ., 2007; Öztuna, 2008; Choi, Grady ve Dodd, 2011; Weiss, 2011).

Bilgisayar Ortamında Bireye Uyarlanmış Test Yönteminin Avantajları

BOBUT yönteminin iki önemli avantajından biri ölçmenin doğruluğunu arttırması, diğeri ise kontrollü ve güvenilir test ortamı sağlamasıdır (Weiss, 1988; Hambleton ve diğ., 1991). Bireye uyarlanmış testler her bir bireye uygulanan çok daha az maddeyle geleneksel testlerin ortaya koyduğu nitelikten üstün ya da en azından kıyaslanabilir ölçme sunmaktadır. Test yapma verimliliğindeki bu artışlar, test uygulama süresindeki kazanımlara yansımıştır, bu da geleneksel testler kullanılarak tek bir özelliğin ölçülebildiği sürede, iki ya da daha fazla özelliğin ölçülebilmesini mümkün hale getirmiştir (Weiss, 1985). BOBUT uygulamalarında, en az geleneksel testler kadar geçerli ve güvenilir bir şekilde testin uzunluğunda %50 ile %80 arasında değişen bir kısalma sağlanabilmektedir (Olsen, Maynes, Slavvson ve Ho, 1989; Kaptan, 1993). BOBUT yönteminin diğeri avantajları; uygulama hızının bireye göre ayarlanması, bireyin istendiği anda test edilebilmesi, testten sıkılma faktörünün en aza indirilmesi, farklı madde türlerinde uygulanabilmesi, test sonuçlarının anında değerlendirilmesi, test izleme süresinin azalması ve test gizliliğinin artması; madde havuzuna kolaylıkla madde eklenebilmesi/çıkarılabilmesi ve karşılaştırılabilir puanlar ile geniş bir test standardizasyonu sağlamasıdır (Lord ve Stocking, 1988; Hambleton ve Swaminathan, 1989; Sands, Waters ve McBride, 1997; Rudner, 1998; Demirtaşlı, 1999; Linacre, 2000; Sukamolson, 2002).

Bilgisayar Ortamında Bireye Uyarlanmış Test Yönteminin Sınırlılıkları

Bilgisayar ortamında bireye uyarlanmış test yönteminin avantajları yanında belli sınırlılıkları vardır. Yöntemin güçlü bir algoritmaya sahip olması ve bilgisayara ihtiyaç duyması bazı sınırlılıkları da beraberinde getirmektedir. Madde havuzunun oluşturulabilmesi için geniş gruptan elde edilmiş çok sayıda maddeye ihtiyaç duyulmaktadır. Bu anlamda, yeterince büyük madde havuzuna sahip olmayan testlerin BOBUT olarak uygulanması mümkün değildir. Her konu için kullanılmayacağı gibi, küçük ölçekli testlerde kullanılması da çok kullanışlı değildir. Geniş grupta yapılacak uygulama için bilgisayar donanımı sınırlılığı ortaya çıkabilir. Bilgisayar kullanımı konusunda olumsuz tutuma, kaygıya sahip bireyler için uygulama sıkıntısı ortaya çıkabilmektedir (Lord ve Stocking, 1988; Hambleton ve Swaminathan, 1989; Rudner, 1998; Demirtaşlı, 1999; Tian ve diğ., 2007).

SONUÇ

BOBUT günümüzde dünya genelinde giderek artan uygulamalara sahiptir. Bunlardan bazı örnekler; ASVAB (Armed Services Vocational Aptitude Test Battery), DYNHA (Dynamic Health Assessments), GMAT (Graduate Management Admission Test), GRE (Graduate Record Examination), J-CAT (Japanese Computerized Adaptive Test), NCLEX (The National Council Licensure Examinations), TOEFL (Test of English as a Foreign Language) olarak sıralanabilir. Bu sınavlardan biri olan TOEFL, Türkiye’de de kâğıt-kalem ve internet temelli olarak sıklıkla kullanılan bir İngilizce sınavıdır. TOEFL dışında Türkiye’de geniş ölçekli bir BOBUT kullanılmamaktadır.

Türkiye’de geniş ölçekli sınavlar merkezi olarak Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) ve Milli Eğitim Bakanlığı (MEB) sorumluluğunda yapılmaktadır. Yapılmakta olan sınavlar 2013 yılı itibarıyla hala geleneksel kâğıt kalem testleri olarak uygulanmaktadır. Uygulanmakta olan geleneksel testler, sabit uzunlukta ve testi alan tüm bireylerin aynı maddelerle karşılaşacağı şekilde hazırlanmaktadır. Sabit uzunlukta hazırlanan testler, bireylerin sahip oldukları farklı yetenek düzeyleri dikkate alındığından farklı güçlük düzeylerinden fazla sayıda maddeye sahiptirler. Bu durum, testi alan bireylerin kendi yetenek düzeyine uygun olmayan maddeleri de yanıtlamak zorunda kalmalarını gerektirmektedir. KTK çerçevesinde hazırlanmakta olan bu sınavlar geniş katılımcı sayısına sahip olmakla birlikte katılan bireyler için -ülke şartlarından dolayı- oldukça öneme sahiptirler. Testlere ait katılımcı sayısının bir hayli fazla olduğu ve bu sayının giderek arttığı göz önüne alındığında, seçme ve yerleştirme sorumluluğu ve bu amaçla yapılan testlerin geçerliğini ve güvenilirliğini koruması daha da önem kazanmaktadır. Bu anlamda kurumların, geleneksel kâğıt kalem testlerinin bu sınırlılıkları göz önüne alınarak her bireyin kendi yetenek düzeyine hitap eden maddelerle karşılaştığı yeni ölçme yaklaşımlar doğrultusunda sınavlarını yenileme ve geliştirme zorunluluğu kaçınılmazdır.

Dünya’da MTK’nın yaygın olarak kullanılma sebeplerinin başında madde ve yetenek parametrelerinin gruptan bağımsız olması ve testten elde edilen puanların daha karşılaştırılabilir olmasıdır. Uygulama kolaylıklarını yanında bu avantajları ile MTK’ya dayanan BOBUT uygulamalarının yaygınlaşması gerekmektedir. Türkiye’de yapılan başta YDS (Yabancı Dil Sınavı) olmak üzere, ALES (Akademik Lisansüstü Eğitimi Giriş Sınavı), AÖK (Açık Öğretim Kurumları Sınavları), DGS (Dikey Geçiş Sınavı), DUS (Dış Hekimliği Uzmanlık Eğitimi Giriş Sınavı), MTSAS (Motorlu Taşıt Sürücü Adayı Sınavı), TUS (Tıpta Uzmanlık Eğitimi Giriş Sınavı) gibi merkezi sınavların; daha kullanışlı, daha geçerli ve güvenilir olması için mevcut teknolojik ve alana ait gelişmeler ışığında dünya normlarının gerisinde kalmayacak şekilde bireye uyarlanmış olarak değerlendirilmesi ölçme ve değerlendirme sürecine önemli katkı getirecektir.

KAYNAKLAR

- Adams, R. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31, 162–172.
- Bejar, I.I. (1983). *Achievement testing: Recent advances*. London: Sage Publications Ltd.
- Choi, S.W., Grady, M.W., & Dodd, B.G. (2011). A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurement*, 71(1), 37–53.
- Crocker, L., & Algina, J. (1986). *Introduction classical and modern test theory*. New York: Harcourt Brace Javonovich College Publishers.
- Çelen, Ü. (2008). Klasik test kuramı ve madde tepki kuramı yöntemleriyle geliştirilen iki testin geçerlik ve güvenilirliğinin karşılaştırılması. *İlköğretim Online*, 7(3), 758-768.
- Demirtaşlı, N. (1999). Psikometride yeni ufuklar: Bilgisayar ortamında bireye uyarlanmış test. *Türk Psikologlar Derneği*.
http://www.psikolog.org.tr/articles_detail.asp?cat=2&id=25 adresinden 25.12.2010 tarihinde erişilmiştir.
- Dodd, B.G. (1990). The effect of item selection procedure and stepsize on computerized adaptive attitude measurement using the rating scale model. *Applied Psychological Measurement*, 14, 355-366.
- Dodd, B.G., & De Ayala, R. J. (1994). Item information as a function of threshold values in the rating scale model. In Wilson, M. (Ed.). *Objective measurement: Theory into practice*. Norwood NJ: Ablex.
- Eggen, T.J.H.M., & Straetmans, G. J. J. M. (1996). Computerized adaptive testing for classifying examinees into three categories. *Measurement and Department Reports*. Cito, Arnhem.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Folk, W.G., & Smith, R.L. (2002). Models for delivery of CBTs. Mills, C.N., Potenza, M.T., Fremer, J.J., & Ward, W.C. (Ed.). *Computer-based testing: Building the foundation for future assessments*. New Jersey: Lawrence Erlbaum Associates.
- Glas, C.A., & Linden, W. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement*, 27(4), 247–261.
- Green, B.F., Bock, R.D., Humphreys, L.G., Linn, R.L., & Reckase, M.D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21, 347-360.
- Hambleton, R.K., & Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(2), 38–47.
- Hambleton, R.K., & Swaminathan, H. (1989). *Item response theory: Principles and applications*. USA: Kluwer Nijhoff Publishing.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. California: Sage Publications Inc.
- Kalender, İ. (2004). Bilgisayar ortamında bireyselleştirilmiş testlerin eğitimde kullanımı. XIII. *Ulusal Eğitim Bilimleri Kurultayı*. İnönü Üniversitesi, Eğitim Fakültesi, Malatya.

- Kaptan, F. (1993). *Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması* (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Kelecioğlu, H. (2001). Örtük özellikler teorisindeki b ve a parametreleri ile klasik test teorisindeki p ve r istatistikleri arasındaki ilişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20, 104-110.
- Linacre, J.M. (2000). Computer adaptive testing: A methodology whose time has come. Chae, S., Kang, U., Jeon, E., & Linacre, J.M. (Ed.). *Development of computerized middle school achievement test*. Seoul: Komesa Press.
- Linn, R., & Gronlund, N.E. (1995). *Measurement assessment in teaching*. New Jersey: Prentice – Hall Inc.
- Lord, F.M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison- Wesley.
- Lord, F.M., & Stocking, M.L. (1988). *Item response theory*. J.P. Keeves (Ed.). *Educational research, methodology, and measurement: An international handbook*. New York: Pergamon Press.
- Olsen, J.B., Maynes, D.D., Slavvson, D., & Ho, K. (1989). Comparison of paper administered, computer administered and computerized adaptive achievement tests. *Journal of Educational Computing Research*, 5(31), 311-326.
- Özçelik, D.A. (2010). *Eğitim programları ve öğretim*. Ankara: Pegem Akademi Yayıncılık.
- Özçelik, D.A. (2011). *Ölçme ve değerlendirme*. Ankara: Pegem Akademi Yayıncılık.
- Öztuna, D. (2008). *Kas-iskelet sistemi sorunlarının özgülülük değerlendirmesinde bilgisayar uyarlamalı test yönteminin uygulanması* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
- Rudner, L.M. (1998). *An on-line, Interactive computer adaptive testing minitutorial*. <http://edres.org/scripts/cat/catdemo.htm> adresinden 02.01.2011 tarihinde erişilmiştir.
- Sands, W.A., Waters, B.K., & McBride, J.R. (1997). *Computerized adaptive testing: From inquiry to operation*. USA: American Psychological Association.
- Segall, D.O. (2003). *Computerized adaptive testing*. *Encyclopedia of social measurement*. www.academicpress.com/tefer/measure/ adresinden 02.12.2011 tarihinde erişilmiştir.
- Sukamolson, S. (2002). Computerized test/item banking and computerized adaptive testing for teachers and lecturers. http://www.stc.arts.chula.ac.th/ITUA/Papers_for_ITUA_Proceedings/Suphat2.pdf adresinden 15.11.2011 tarihinde erişilmiştir.
- Thompson, J.G., & Weiss, D.J. (1980). *Criterion-related validity of adaptive testing strategies* (Research Rep. No. 80-3). Minneapolis: University of Minnesota, Department of Psychology, Computerized Adaptive Testing Laboratory.
- Tian, J., Miao, D., Zhu, X., & Gong, J. (2007). An introduction to the computerized adaptive testing. *US-China Education Review*, 4(1), 72-81.
- Turgut, F. (1992). *Eğitimde ölçme ve değerlendirme metotları*. Ankara: Yargı Matbaası.
- Turgut, F. ve Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi Yayıncılık.

Vale, C.D., & Weiss, D.J. (1975). *A study of computer administered strataptive ability testing* (Research Rep. No. 75-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J. Steinberg, L., & Thissen, D. (2000). *Computerized adaptive testing: A primer*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Ware, J.E., Kosinski, M., Bjorner, J.B., Bayliss, M.S., Batenhorst, A., Dahlöf, C.G. H., Tepper, S., & Dowson, A. (2003). Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Quality of Life Research*, 12, 935-952.

Way, W.D. (2006). *Practical questions in introducing computerized adaptive testing for K-12 assessments* (Research Report). Pearson Educational Measurement.
http://education.pearsonassessments.com/NR/rdonlyres/EC965AB8-EE70-46E5-B1A5036BE41AB899/0/RR_05_03.pdf?WT.mc_id=TMRSPractical_Questions_in_Introducing_Computerized adresinden 08.06.2012 tarihinde erişilmiştir.

Weiss, D.J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, 53(6), 774-789.

Weiss, D.J. (1988). Adaptive testing. Keeves J. P. (Ed.). *Educational research, methodology, and measurement: An international handbook*. New York: Pergamon Press.

Weiss, D.J. (2011). Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences*, 2(1), 1-27.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New Jersey: Lawrence Erlbaum Associates.

Yurdugül, H. (2006). Çoktan seçmeli testlerde güvenilirlik terimlerinin madde-yanıt kuramından temin edilmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 39(2), 27-44.

EXTENDED ABSTRACT

Measurement and evaluation process plays a critical role in revealing whether aims of education are achieved or not. In Turkey, mostly the traditional paper and pencil tests based on Classical Test Theory are used for measurement and evaluation. In Classical Test Theory, since item and ability parameters estimated from the tests are also dependent to the target group, it causes limitation on reliability and validity of tests related to the target group. In order to overcome this limitation, models based on Item Response Theory have become more widespread all over the world in which item and ability parameters are made independent from the group invariance feature and scores from the tests are transformed to be more comparable.

Along with the advancements in technology use for various educational purposes, measurement and evaluation processes have been successfully integrating technology. Computerized Adaptive Test (CAT) that has been utilized globally can be considered as an outstanding example. This kind of method of testing enables developing shorter, valid and reliable tests that are appropriate to the individual's ability levels. This study attempts to reveal the details of CAT method for a better understanding.

Contrary to the individualized adaptive tests and traditional paper and pencil tests; instead of posing all items to all individuals, only some items are used to accurately determine the ability levels of individuals. By analyzing the responses of the individual taking the individualized adaptive test, items appropriate to the ability levels are chosen from the item pool identified with difficulty and discrimination parameters. Numerous methods are available about Computerized Adaptive Testing. These are; Two-stage Testing, Self-selecting Testing, Pyramidal Multistage Testing, Alternating Testing, Stradaptive Testing and Multilevel Format. The practice of CAT generally starts with an item of medium difficulty. The initial score of the individual is considered as an already existing ability or an average score. Ability estimation score for each item is performed according to the responses from the individual and the process continues until the stopping rule is achieved; the process goes on along with item selection procedures/rules. Timing for the termination of the test is an essential element to properly define the ability level for the CAT ability estimation algorithm. The number of questions to be posed and selecting the ability appropriate questions need to be well defined to identify the ability level of the individual. Some of the stopping rules can be listed as; shortage of questions in the item pool, reaching the maximum length of test if fixed-length is preferred, strength of standard error calculated for each ability estimation, and individual's behaviors irrelevant of the test. CAT offers plentiful advantages by applying the ability estimation principle for each item after selecting individual-appropriate items. One of the major advantages of the

method is providing a controllable test environment by increasing its reliability and validity. Contrary to the traditional tests, the method can estimate the comparable ability scores with little number of items. Considering the large-scaled tests, ability estimation with fewer items leads to shortening of the test time and ease of test implementation. Test speed can be adjustable according to the pacing of the individuals with differing response speed. One of the advantages of the method is the convenience of adding to and removing items from the pool by using various item types. Its complex algorithm and need for computers are among the limitations of the method. Numerous items from large groups are needed to form an item pool. To this end, tests without proper number of items cannot be implemented as CAT. It is not possible to use the method for every topic and in the small-scale tests. Computer availability can be a limitation with the large number of groups. In addition, individuals with negative attitudes or anxiety towards computer use can cause problems.

CAT has been widely used worldwide due to its numerous advantages. One implementation is the English proficiency test TOEFL that can be applied paper and pencil or computer-based in Turkey. There is no other large scale CAT exam than TOEFL in Turkey. The tests applied are paper and pencil based as of 2013. Considering the increasing number of test-takers, responsibility for selection and placement as well as reliability and validity issues gain more importance. Therefore, it is inevitable for the institutions to revise and update the tests regarding the limitations of paper and pencil test and advantages of new evaluation approaches with adapted items for each individual. The centralized tests of YDS (Foreign Language Test), ALES (Academic Graduate Education Entrance Test), AÖK (Open Education Tests), MTSAS (Motor Vehicle Driver Examination) need to be reliable and valid; hence, their adaption in line with the technological innovations will highly contribute to the measurement and evaluation processes.

YAZAR HAKKINDA

Arş. Gör. Dr. Fatih Kezer Kocaeli Üniversitesi Eğitim Fakültesi Ölçme ve Değerlendirme Anabilim Dalı'nda görev yapmaktadır. Çalışma konuları; bilgisayar ortamında bireye uyarlanmış test, eğitimde istatistik ve araştırma yöntemleridir. / İletişim Adresi: Kocaeli Üniversitesi Eğitim Fakültesi Umuttepe Yerleşkesi 41380 Kocaeli/Türkiye Tel: (0262) 303 24 76 / Eposta: fatibkezer@yahoo.com

ABOUT THE AUTHOR

Research Assistant Fatih Kezer, PhD. He works at the Department of Measurement and Evaluation at Faculty of Education, Kocaeli University. His researches include computerized adaptive testing, educational statistics and research methods. / Correspondence Address: Kocaeli Üniversitesi Eğitim Fakültesi Umuttepe Yerleşkesi 41380 Kocaeli/Türkiye Phone: (0262) 303 24 76 / Email: fatibkezer@yahoo.com
