

YETENEK VE BAŞARININ ÖLÇÜLMESİNDE TAILORED TEST YÖNTEMİ

Yrd. Doç. Dr. Nilgün KÖKLÜ*

GİRİŞ

Tailored test yönteminin özelliklerine geçmeden, genelde literatürde “adaptive test yöntemi” olarak bilinen bireysel testlerin özelliklerini açıklamak da fayda vardır. Bireyin cevaplama özelliklerine yada başka bir ifade ile yeteneğine uygun maddelerin seçilerek uygulanması esas alınarak kalem-kağıt yada bilgisayar kullanılarak yapılan, bireysel uygulama gerektiren testler genelde adaptive testler adı verilmiştir.

Adaptive testler için üç önemli özellik; cevaplayıcıya sorulacak olan ilk maddenin cevaplayıcı hakkında daha önceki bilgilerin kullanılarak seçilmesi, daha sonra ki maddelerin kişinin verdiği cevaplar doğrultusunda seçilerek sorulması ve yeterli sayıda madde uygulandığında teste son verilmesidir (Weiss, 1976). İşte bu özellikler adaptive testlerin klasik testlerden farkını ortaya çıkarmaktadır. Klasik test uygulamalarında, testte yer alan tüm maddeler, herkese uygulandığı için kişiler, kendi yeteneklerine uygun olmayan güçlükteki maddelerle karşı karşıya kalmaktadırlar.

Adaptive test yönteminde kullanılan kural yukarı-aşağı yöntemi (up-down rule) olarak bilinen yöntem olup, bu yöntemde göre eğer bir cevaplayıcı kendisine yöneltilen bir maddeyi doğru cevaplarsa, ondan sonra uygulanacak olan madde daha zor, yanlış cevaplarsa daha kolay bir madde olmaktadır (Lord 1970; Weiss 1976; Brown ve Weiss 1977; Thorndike 1982).

Son yıllarda bilgisayar teknolojisinin gelişmesi ve kullanımının artması ile, adaptive testlerin uygulanmasında bilgisayar kullanılmaya geçilmiş, bu da bu testlerin avantajlarını arttırmıştır.

* Eğitim Yönetimi ve Planlaması Bölümü Öğretim Üyesi.

Bireysel olarak uygulanan ve madde seçiminde bilgisayar kullanılarak oluşturulan testlerden birisi de "tailored test" (Lord, 1970) dir. Tailored test, kişinin cevaplayacağı test maddelerinin sırasının ve sayısının, onun daha önceki maddelerde göstermiş olduğu performansına bağlı olduğu bir test yöntemi olarak tanımlanabilir (Spinetti ve Hambleton, 1977).

Thorndike (1982: 300-302) tailored test yöntemi için altı uygulama modelini şöyle açıklamaktadır.

1. *Çok Aşamalı Uygulama*: Bu modele göre, cevaplayıcıya uygulanacak olan maddelerin seviye seçimi, onun daha önceki bilgilerine, yaşına veya sınıfına bağlı kalınarak yapılır. Bu testler, birbirini izleyen her bir test kendinden öncekinden daha zor olacak şekilde pekçok testten oluşmaktadır. Testin en düşük seviyesi için, kişi en kolay model ile başlar ve devam eder. Testin sekiz veya on seviyelik serileri yapılabilmekte, bunların her biri kendinden önce gelenden zor olacak şekilde uygulanmaktadır.

2. *İki Aşamalı Uygulama*: Bu modelde, test iki parçadan oluşur. Testin ilk parçası, bütün cevaplayıcılar için aynı olmaktadır. Bu ilk parçadan alınan puan kişinin ikinci olarak alacağı parçayı tayin etmekte kullanılmaktadır.

İkinci parça güçlük düzeyleri farklı pek çok alternatif formlardan oluşmaktadır. İlk parçadan alınan puan yetenek tahmininde bulunmaya yarar ve bu tahminin kişi hakkında daha çok bilgi alınmasını sağlayacak olan ikinci formun hangi güçlük düzeyinde olması gerektiğine yardımcı olur.

3. *Cevaplayıcı Seçimine Dayalı Uygulama*: Bu modelde test maddeleri güçlük düzeylerine göre sıralanarak, cevaplayıcı kendi yaş grubuna yada sınıfına göre uygun olabilecek orta güçlükteki maddeler ile teste başlatılır. Eğer bu maddeler cevaplayıcı için kolay ise, cevaplayıcı daha zor maddelere geçer; eğer onları zor bulursa, daha kolay maddelerin sorulmasına geçilir.

4. *Değişimlemeli Uygulama*: Bu modelde de yine, maddeler güçlük düzeylerine göre sıralanır. Başlangıç maddesi orta güçlükte olacak şekilde (<) işaretinin iki kolundan birine kolay, diğerine ise zor maddeler sırası ile yerleştirilir. Cevaplayıcı teste orta güçlükteki ilk maddeyi cevaplayarak başlar, eğer bu maddeyi doğru cevaplırsa, daha zor maddeleri cevaplamaya çalışır, başarısız olduğunda, kolay maddelerin sıralandığı taraftan ilk kolay madde sorulur. Belli sayıda madde uygulanana kadar teste devam edilir.

5. *Kademeli Uygulama* : Bu model, güçlük düzeyine göre homojen ve ayrılcılığına göre sıralanmış maddelerden oluşan kademelerden meydana gelmektedir. Cevaplayıcı, kendi yaşı, sınıfı yada yada önceki bilgileri esas alınarak, kendisine uygun olabilecek bir kademenin ilk maddesi ile teste başlar, eğer doğru cevaplırsa daha güç kademenin ilk maddesini cevaplar. Başaramadığı zaman, test daha kolay kademeye geçilerek devam eder. Her kademeye geçildiğinde kullanılmamış ilk madde uygulanarak teste devam edilir.

6. *Çok Basamaklı (Piramidal) Model* : Bu modelde, maddeler güçlük düzeylerine göre her bir basamakta sıralanır. Tipik bir örüntü, aşağıda 4- basamaklı bir piramit oluşturacak şekilde gösterilmiştir.

		BASAMAK				
		I	II	III	IV	
Başlangıç				4	7	Zor
			2	5	8	
		1	3	6	9	
					10	

Bütün cevaplayıcılar I. madde ile teste başlar. I. maddeyi doğru cevaplayan, II. basamakta daha zor olan 2. maddeyi cevaplar. I. maddeyi doğru cevaplamayan ise, II. basamaktaki 3. maddeyi (daha kolay) cevaplar. Her bir basamakta, madde doğru olarak cevaplanırsa, ondan sonraki basamakta daha zor olan maddeye geçilir, yanlış cevaplandığında ondan sonraki basamakta daha kolay madde cevaplanır. Böylece 4-basamaklı bir piramit modelde 4 madde cevaplanarak teste son verilir.

Tailored test yönteminin en önemli özelliği, kişinin yetenek düzeyine uygun maddelerin seçilerek sorulmasıdır. Soruların seçilmesinde bilgisayarın kullanılması, testin uygulanmasını kolaylaştırmaktadır. Bu teste, madde istatistikleri hesaplanmaz, önceden geliştirilmiş ve istatistikleri hesaplanmış maddelerden yararlanır. Bunun için madde parametreleri bilinen, geniş madde havuzlarına (item pool) ihtiyaç vardır.

Tailored test yöntemlerinde, klasik test teorisinden faydalanılabilmekle beraber, tailored testin bireylere uygulanmasında, madde havuzlarından, herkese farklı maddelerin seçilip sorulması gerektiğinden, madde parametrelerinin gruba bağımlı olması nedeni ile klasik test teorisinin kullanılması güçleşmektedir. Bu nedenle, tailored test ile ilgili çalışma-

larda, madde istatistiklerinin hesaplanmasında madde parametrelerinin gruba bağımlı olmadığı örtük özellikler teorisi (Latent trait theory)'nin kullanılması daha yaygındır.

Lord, tailored test için bilgisayarın aşağıdaki şartları sağlayacak şekilde programlanması gerektiğini vurgulamaktadır (Kreitzberg ve Jones 1980: 14).

1. Kişinin daha önceki performansı gözönünde bulundurularak testin henüz uygulanmamış maddelerine verebileceği cevaplar tahmin edilmeli,

2. Bu bilgi, daha sonra uygulanacak olan maddenin seçiminde etkin olarak kullanılabilmesi,

3. Testin sonunda, kişinin yetenek düzeyini gösteren bir puan verilebilmelidir.

Tailored testlerin uygulama yöntemlerinin yanısıra, bireye maddelerin nasıl sorulacağına, maddelerin nasıl seçileceğinin yanında, teste nasıl son verileceği konusu da çok önemlidir. Tailored testlerde, teste son vermede iki yönteme başvurulmuştur. Bunlardan birisi, bireyin yetenek tahmininin belli bir güvenilirlik düzeyine ulaşmasıyla teste son vermektir. Urry (1977) ve Jensema (1974)'nin yaptıkları çalışmada, standard hatanın doğrudan tailored test güvenilirliğine bağlı olduğu ve teste son vermede önceden belirlenen sabit bir standard hataya ulaşılmasının gerektiği anlatılmaktadır. Teste son vermede ikinci yöntem ise, önceden belirlenen soru sayısına ulaşmaktır (Jensema 1974; McKinley ve Reckase 1980; Kreitzberg ve Jones 1980; Köklü 1990). Bu çalışmalara bakıldığında, teste son vermek için belirlenen soru sayısının çok az olduğu görülmektedir. Bunun sebebi, az sayıda madde ile yüksek güvenilirliğe ulaşabilmedir.

Tailored Test Yönteminin Avantajları

Tailored test yöntemi ile ilgili olarak yapılan çalışmalar, bu tür testlerin, grup testlerine göre bazı avantajları olduğunu ortaya çıkarmıştır (Kreitzberg ve Jones 1980: 17). Bilgisayara uyarlanmış bireysel testlerin grup testlerine göre daha programlı ve standardize olduğunu, böylece kontrol edilebilirliğini ve çevrenin etkisinin en aza indirildiğini savunmaktadır. Yine Kreitzberg ve Jones (1980: 17-19)'un belirttiğine göre Weiss, bilgisayarın kullanılması ile cevap kağıdının niteliğinin, madde düzeninin, puanlama yönteminin zararlı etkilerinin en aza indirildiğini ve test kitapçıklarına ihtiyaç olmadığından gizliliğinin sağlanmasına katkıda

bulunduğu, bireysel testlerle grup testlerine oranla daha fazla bilgi edinmenin mümkün olduğunu, dönüt verilerek öğrenme güçlüklerinin daha iyi tespit edildiğini ve cevapların anında puanlanabildiğini belirtmektedir.

YETENEĞİN VE BAŞARININ ÖLÇÜLMESİNDE TAILORED TEST YÖNTEMİNİN KULLANILMASI ÜZERİNE ÇALIŞMALAR

Tailored test konusundaki çalışmalar ancak son zamanlarda eğitim araştırmacılarının dikkatlerini çekmiştir. 1950'lerde ve 1960'ların ilk yıllarında çeşitli çalışmalar olmasına rağmen, Lord'un (1970, 1971) test zamanını azaltarak, öğrencinin yetenek düzeyinin hassas olarak ölçülmesi ile ilgili ölçme tekniklerinin geliştirilmesi üzerine yaptığı çalışması, tailored test yönteminin ilgi çekmesine neden olmuştur. Wood (1973) bu tür araştırmaların bir derlemesini yapmıştır.

Tailored test yöntemi üzerine yapılan çalışmalar genellikle yeteneğin ölçülmesi üzerine olmuştur. Bu çalışmalar, tailored test yöntemi ile aynı değişkeni ölçen grup testlerine göre daha az sayıda madde kullanılarak daha güvenilir ve geçerli testlerin elde edilebileceğini gösterir mahiyettedir (Ferguson 1969; Lord 1977; Kreitzberg ve Jones 1980).

Lord (1977) tarafından geliştirilen "Broad Range Tailored Test (BRTT)" bilgisayara uyarlanmış, madde parametreleri örtük özellikler teorisine göre hesaplanmış olup 5. sınıftan üniversiteye kadar olan öğrencilerin sözel yeteneğini ölçebilecek şekilde düzenlenmiştir (Kreitzberg ve Jones 1980: 25-31). Lord bu çalışmasında, yukarı-aşağı yöntemi kullanmış ve BRTT'yi aynı yeteneği ölçen klasik bir test ile karşılaştırdığında, BRTT'nin daha az madde ile daha güvenilir sonuçlar verdiği sonucuna ulaşmıştır.

Kreitzberg ve Jones (1980) BRTT'nin performansı üzerine bir çalışma yapmıştır. Bu testin iki paralel formu 146 lise öğrencisine uygulanmış ve her öğrenci kendi yeteneğine uygun 25 soru cevaplamıştır. Sonuçlar, üç parametrelili lojistik modelin tailored teste uygun olduğunu ortaya çıkarmış ve az sayıda madde ile yüksek güvenilirliğe ulaşıldığını göstermiştir. Aynı öğrencilerin bu teste olan tutumlarına bakıldığında ise, tailored test yönteminin klasik teste göre daha az yorucu ve kullanışlı olduğu ortaya çıkmıştır.

Koch ve Reckase (1978)'de sözel yeteneğin ölçülmesinde tailored test kullanmışlardır. Test-tekrar test ve eşdeğer form güvenilirliklerine bakıldığında, üç parametrelili lojistik modelin bir parametrelili lojistik modele göre daha iyi sonuçlar verdiği görülmüştür.

Başarı testlerinin çok boyutlu olmalarından dolayı, tailored test teknikleri daha çok tek boyutlu olan yetenek testlerinde daha başarılı sonuçlar vermiştir. Yine de 1977'lerden sonra başarının ölçülmesiyle ilgili çalışmalara rastlanmaktadır. Bejar, Weiss ve Kingsburry (1977) tarafından yapılan bir çalışmada, biyoloji başarı testi kullanılmış, ancak bu testin tek faktör (tek konu) ağırlıklı olduğu anlaşılmıştır. Madde karakteristik eğrisi modeli kullanılarak yapılan madde havuzu seçimlerinin yeterli olduğu kanıtlanmıştır. Bu model ile tek faktör başarı testinin uygulamasının tek boyutlu yetenek testinden farklı olmadığı sonucuna ulaşılmıştır.

Koch ve Reckase (1978)'in çalışmasında lisans düzeyindeki eğitimde ölçme ve değerlendirme dersine ait sınıf testlerinin her biri bir parametrelili ve üç parametrelili lojistik modele göre hazırlanmış olup, test-tekrar test yöntemine göre uygulanmıştır. Çalışmanın sonuçları, uygulanan bir parametrelili ve üç parametrelili lojistik modellerin güvenilirlik için yetersiz olduğunu ortaya koymuştur.

Mc Kinley ve Reckase (1980)'in yaptığı çalışmada, test-tekrar test yöntemi kullanılarak 88 öğrencinin oluşturduğu bir gruba, bir parametrelili lojistik ve üç parametrelili lojistik modele göre hazırlanmış tailored başarı testi uygulanmıştır. Ölçmeye Giriş dersine ait teste 20 madde uygulandığında son verilmiştir. Bir parametrelili ve üç parametrelili lojistik test için hesaplanan test-tekrar test güvenilirlikleri arasındaki fark manidar bulunamamıştır. Klasik testle karşılaştırıldığında ise test-tekrar test güvenilirliğinin yüksek olduğu görülmüştür.

Ayrıca bu çalışmada, kapsam geçerliğine bakılmış, testlerin belirtke tablosunda gösterilen konu alanına verilen ağırlıkları yansıttığı kanıtlanmıştır. Tutum ölçeğinin sonuçları da, tailored teste karşı öğrencilerin motivasyonunun fazla olduğunu ve çok az stresli olduklarını göstermiştir.

Reckase (1986)'nın Illinois'de Denizcilik Öğretim Merkezi'nde Radar Teknik Öğretim Kursu ile birlikte yerine getirilen tailored test yöntemini açıklayan projesinde, öğrencilerin tailored teste olan tutumları incelenmiştir. Sonuçlar, bilgisayar ile test almanın herhangi bir güçlüğü olmadığını, öğrencilerin % 62'sinin tailored testten hoşlandığını göstermiştir. Testin nasıl yapılacağına dair verilen komutların çok kolay olduğu, klasik teste göre tailored testin biraz daha hızlı uygulanabildiği sonucuna varılmıştır.

Tailored test ile ilgili Türkiye'de yapılan bir çalışmada ise Köklü (1990) tailored test ile grup testinin başarıyı ölçmede ne derece farklılık

gösterdiğini incelemiştir. Klasik test teorisinin madde parametreleri kullanılarak, değişimlemeli modele göre hazırlanan tailored test ile grup testinin güvenilirlikleri, geçerlikleri ve bu iki yöntemden hangisinin aşamalılık ilişkilerinin ortaya çıkarılmasında daha etkili olduğu incelenmiştir. "Işık, Işığın Yansırması ve Kırılması" konuları için hazırlanan her iki test, T.E.D. Ankara Koleji Ortaokul III. sınıflardan seçilen iki gruba uygulanmıştır. Araştırmada, tailored testin ve grup testinin KR-29 güvenilirlikleri arasında manidar bir farklılık görülmüş, hesaplanan geçerlik katsayıları arasında manidar bir fark bulunmamakla birlikte, tailored testin bir miktar daha geçerli sonuç verdiği izlenimi edinilmiştir. Ayrıca, tailored testin grup testine göre önşart ilişkilerini göstermekte daha etkin bir yöntem olarak kullanılabilceği kanısına varılmıştır.

SONUÇ

Genel olarak, yapılan çalışmalara bakılacak olursa, tailored testin yetenek ve başarının ölçülmesinde kullanılabilceği, hatta grup testlerine oranla daha iyi sonuçlar verdiği, ancak yetenek ölçmede daha iyi olduğu, bununda nedeninin başarı testlerinin çok boyutlu olmasından kaynaklandığı görülmektedir. Tailored testin, örtük özellikler teorisine göre hazırlanan madde parametreleri kullanılarak uygulandığında, üç parametrelili lojistik modelin, bir parametrelili lojistik modele göre daha güvenilir ve geçerli sonuçlar verdiği gözlenmektedir. Ayrıca, başarının ölçülmesinde kullanıldığında ise, aşamalılık ilişkilerini açıklamakta grup testine göre daha iyi olduğu görülmektedir. Diğer göze çarpan bir sonuç ise, öğrencilerin tailored testine karşı olan tutumların gayet olumlu olduğu, bu testin daha az yorucu ve kullanışlı bulunduğudur.

KAYNAKLAR

- Bejar, I.I. D.J. Weiss ve G.G. Kingsbury (1977). Calibration of an Item Pool for the Adaptive Measurement of Achievement. Research Report. Minneapolis: University of Minnesota, Department of Psychology.
- Brown, J.M. D.J. Weiss. (1977) An Adaptive Testing Strategy for Achievement Test Batteries Research Report. Minneapolis: University of Minnesota, Department of Psychology.
- Jensem, Carl J. (1974) "The Validity of Bayesian Tailored Testing." Educational and Psychological Measurement, 34: 757-766.
- Koch, W.R. M.D. Reckase (1978) A Live Tailored Testing Comparison Study of the One- and Three- parameter Logistic Models. Research Report. Columbia: University of Missouri Department of Educational Psychology.
- Kökü, Nilgün (1990) Klasik Test Teorisine Göre Geliştirilen Tailored Test ile Grup Testi Arasında Bir Karşılaştırma Yayınlanmamış Doktora Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.

- Lord, F.M.** (1979) "Some Test Theory for Tailored Testing". W.H. Holtzman (Ed.) Computer Assisted Instruction, Testing and Guidance. New York: Harper and Row, Bölüm 8.
- Lord, F.M.** (1977) "A Broad-Range Tailored Test of Verbal Ability". Applied Psychological Measurement, 1: 95-100.
- Mckinley, E.L ve M.D. Reckase** (1980) A Successful Application of Latent Trait Theory to Tailored Achievement Testing. Research Report, Columbia: University of Missouri.
- Reckase, M.D.** (1986) The Use of Tailored Testing with Instructional Program. Research Report, Iowa City.
- Spineti, J.P., R.K. Hambleton.** (1977) "A Computer Simulation study of Tailored Testing Strategies for Objective-Based Instructional Programs" Educational and Psychological Measurement, 37: 139-158.
- Thorndike, R.L.** (1982) Applied Psychometrics: Boston: Houghton Mifflin Company.
- Urry, V.W.** (1977) "Tailored Testing: A Successful Application of Latent Trait Theory." Journal of Educational Measurement, 14:2:181-195.
- Weiss, D.J.** (1976) The Stratified Adaptive Computerized Ability Reserach Report 73-3. Minneapolis: Department of Psychology.
- Wood, R.** (1973) "Response-Contingent Testing". Review of Educational Research, 43: 529-544.