

TEST GELİŞTİRMEDE YENİ YAKLAŞIMLAR: ÖRTÜK ÖZELLİKLER KURAMI-TEMEL ÖZELLİKLERİ VARSAYIMLARI, MODELLER VE SINIRLILIKLARI

*Nükhet ÇIKRIKÇI-DEMİRTAŞLI**

Bu yazıda, Psikometri alanında test ve ölçek geliştirme çalışmalarının büyük bir kısmının kuramsal dayanağını oluşturan Klasik Test Kuramına alternatif olarak geliştirilen ve 80'li yıllardan itibaren üzerinde yapılan çalışma sayısının arttığı, diğer bir test kuramı olan Öztürk Özellikler Kuramı (Latent Traits Theory) üzerinde durulmaktadır. Bu gerçevede, Kuramın temel özellikleri, varsayımları ve modelleri ile sınırlılıkları ve ülkemizde bu kuramla yapılmış araştırmalar ele alınmaktadır.

Herhangi bir psikoloji kuramı için temel sorun, verilen bir davranış evreni içinde, kritik önemi olan psikolojik özelliklerin gözlenebilir bir biçimde ortaya konması ve tanımlanmasıdır. Örtük özelliklerle ilgili bir test kuramı ise, bireyin davranışını, "treyleler" (tratis) adı verilen psikolojik özellikleri tanımlama yoluyla, belli bir dereceye kadar açıklanabileceğini, bu psikolojik özelliklerin (doğrudan gözleyemediğimiz, başka davranışlar aracılığı ile hakkında çıkarsamalar yaptığımız, yetenek başarı, kişilik, ilgi biçiminde örneklendirilebileceğimiz, bireylere özgü özelliklerin birinde, bireyin yerinin puan olarak kestirilebileceği ve sonrada, benzer koşullarda bireyin davranışını açıklamak ve yordamak üzere elde edilen sayısal değerlerin, kullanılabilceğini farzeder (Lord ve Novick, 1968).

Örtük özellikler kuramı (Latent Traits Theory) veya Madde-Cevap Kuramı (Item-Response Theory) olarak da adlandırılan bu ölçme kuramı, klasik test kuramında yaşanan sınırlılıklara alternatif olarak geliştirilen bir ölçme kuramıdır. Özellikle, klasik test kuramına dayalı test geliştirme ve uygulamasında elde edilen, madde ve test istatistiklerinin, testin uygulandığı grubun yetenek düzeyine göre değişmesi, alternatif kuram arayış-

* Araş. Gör., Dr. Eğitimde Psikolojik Hizmetler Bölümü Ölçme Değerlendirme Anabilim dalı.

larına yol açan önemli bir sınırlılık olmuştur. Örtük özellikler kuramı, bireyin testle ölçüldüğünü düşünülen gözlenemeyen psikolojik özellikleri (yetenek) ile bireyin o testle gösterdiği cevaplama edimi arasındaki doğrusal ilişkiyi ortaya koymaya çalışır (Hambleton ve diğ., 1978; Gustafsson, 1980; Hambleton ve Swaminathan, 1985; Hambleton, 1989). Bir başka deyişle kuram, bir cevaplayıcının bir maddeye doğru cevap verme olasılığını gösterecek olan madde özellikleriyle, cevaplayıcıların yetenek düzeyinin (psikolojik özellik) birlikte ele alınmasında, matematiksel eşitliklerin kullanımı öngörerek, iki özellik arasındaki ilişkiyi, klasik kuramdan farklı olarak, matematiksel biçimde işlevselleştirmeyi amaçlar (Gustafsson, 1988; Wilcox, 1988). Bu bağlamda “örtük özellik” kavramı, yukarıda sözü edilen iki değişken arasındaki doğrusal ilişkiyi matematiksel eşitliklerle tanımladığı için, istatistiksel bir yapıya gönderme yapar. Bu çerçevede örtük özellik kavramı, bireysel farklılıkların ortaya çıkışında başlı başına bir kavram olarak ele alınmamaktadır. (Anastasi, 1988).

Kuramın, Klasik Test Kuramından en önemli farklı yönü, test maddelerinin özelliklerini veren madde parametreleri kestirimlerinin bir cevaplayıcı grubundan diğerine değişmemesi (sample free measurement); cevaplayıcıya ait yetenek parametrelerini kestirimlerinin, bir grup madde örnekleminde diğerine değişmeden (item free measurement) kalmasıdır. (Lord ve Novick, 1968; Hambleton ve diğ., 1978; Crocker ve Algina, 1986; Gustafsson, 1988; Hambleton, 1989). Buna karşılık klasik test kuramında, bu yönde yaşanan sorunun özünde, gözlenen puan, x 'in (veya gerçek puan T 'nin), birimlerinin aralıklı bir ölçek üzerinde yer almamasıdır. T puanı $(0, n)$ aralığında ve belli bir test için tanımlanmıştır. (Hambleton ve diğ., 1978; Douglas, 1988). Bir cevaplayıcının T puanı, aynı yeteneğin ölçüsü, paralel olmayan ölçmeler boyunca değişecektir. Öte yandan, bir cevaplayıcının (θ) yeteneği, tek bir yeteneğin ölçüsü olan bir “madde havuzu” veya “madde evreni” için tanımlanmıştır. Kuramda, belli bir madde örneklemindeki cevaplayıcı edimini, geniş bir madde örneklemini için tanımlanmış yetenek ölçeğine dönüştürmek her zaman mümkündür. Böylece bir cevaplayıcı, bir madde havuzundan alınan herhangi bir madde örneklemini (test) için farklı gerçek puanlar almış iken, her madde örnekleminde, farklı test puanları alacak, ancak, her madde örnekleminde beklenen cevaplayıcı yeteneğin (θ) kestirimi aynı olacaktır.

Bu çerçevede, bir psikolojik özellik olarak, “yetenek” (θ) kavramı bu kuramda, çoğunlukla eşit aralıklı birimleri olan, bir ölçeği öngörür. Örtük özellikler kuramındaki örtük değişken, θ (yetenek), $(+\infty, -\infty)$ aralığında tanımlı ve süreklidir. Bu aralık içinde, olası bütün değerleri alır. Bu bağlamda, klasik test kuramındaki T puanı ile Örtük özelliklerdeki θ , örtük değişken ölçüsü olarak, aynı yeteneği farklı ölçeklerde temsil ederler. Örtük Özellikler Kuramındaki değişmez (invariance) parametre değerlerinin elde edilmesi için, seçilen örtük özellikler modeli ile veriler arasında

bir uygunluğun (fitting) olması, kuramın temel varsayımlarının karşılanması gerekir. Bu kuramda yapılan ölçekleme çalışmalarında, tüm (complete) örtük uzay, ölçme hatası kavramını da içerecek biçimde tanımlanmıştır. (Lord ve Novick, 1968; Hambleton ve Swaminathan, 1985).

A. Temel Varsayımları

Kuramın herhangi bir yetenek ölçümünden elde edilen verilere uygulanmasında ve kuramın iddiaları çerçevesinde madde ve yetenek kestirmelerinin elde edilebilmesi için, üç varsayımın karşılanması gerekir.

- Ölçülen psikolojik özelliğin tek boyutlu olması (unidimensionality)
- Ölçülen psikolojik özelliğin evrende normal dağılım göstermesi
- Maddelere verilen cevapların yerel bağımsızlığı (local independence)

bu çerçevede kuramın, en önemli varsayımları tek boyutluluk (unidimensionality), yerel bağımsızlık (local independence) ve madde-test karakteristik eğrileridir (item/test-characteristic curve). Aşağıda bu boyutlar ana hatları ile tanımlanmıştır.

1. Tekboyutluluk (Unidimensionality)

k'sayıda psikolojik özellik ve n sayıda madde düşünelim. Bu psikolojik özelliklerin her birinin, bireyin n sayıdaki test maddesinden en az birinde gösterdiği cevaplama davranışını etkilediği varsayıldığında, bu psikolojik özelliklere örtük özellikler (latent traits) veya örtük değişkenler (latent variables) adı verilir. Sonra, bu k' boyutlu uzayda, tüm cevaplayıcı evrenin n sayıda madde ile karşılaştığını ve her maddenin her cevaplayıcıya bir kez uygulandığını, belli bir θ (yetenek) değeri için cevaplayıcılardan madde puanlarının cevaplarına ait puanların koşullu dağılımı, tüm cevaplayıcı evrenleri için aynı ise, bireylerin testle ölçülen yeteneklerinin tek boyutlu olduğu varsayılır. Bu varsayım da, kuramın ön gördüğü "tek boyutluluk" varsayımıdır. (Lord ve Novick, 1968 s:358-359; Hambleton ve Swaminathan, 1985 s:25).

Kuşkusuz bu varsayım, psikolojik ölçümlerde tam olarak karşılanmaz. Çünkü test uygulaması sırasında, testteki edimi etkileyebilecek çeşitli bilişsel (akademik başarı, önceki öğrenmeler, genel yetenek düzeyi), duyuşsal özellikler (ilgi, tutum) ve kaygı, heyecan, fizyolojik bir rahatsızlığın olup olmaması ve en önemlisi ölçülen psikolojik özelliğin tanımlanmasından gelen sorunlar olabilir (Hambleton ve Swaminathan, 1985). Ancak, burada test edimini "başat" olarak etkileyebilecek faktör gözönüne alınmaktadır. Bu da testle ölçülen "yetenektir/psikolojik özelliktir." Tekboyutlu test geliştirme veya bir testin tek boyutluğunu incelemede, bazı yöntemlerden söz edilir. Bunlardan biri, "Lumsden yöntemi" olarak

bilinen, faktör analizine dayalı, tekboyutlu test geliştirmeye yarayan bir yöntemdir. (Gustafsson, 1980; Hamleton ve Swaminathan, 1985 s: 18-19). Tekboyutluluğu belirlemede ikinci bir yol, kesikli iki değişken arasındaki ilişkinin ölçüsünü veren Tetrakorik Korelasyon katsayılarının hesaplanmasıdır. Maddelerin tekboyutluluğu için yeterli koşul, maddelerarası Tetrakorik Korelasyon katsayısı matrisinin tek bir faktöre sahip olmasıdır (Hambleton ve Swaminathan, 1985 s:19).

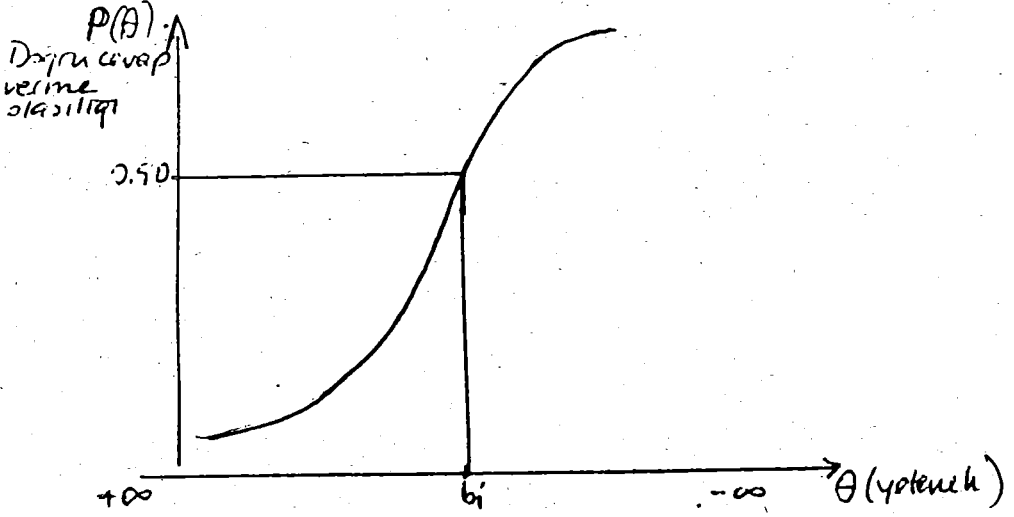
2. Yerel Bağımsızlık (Local Independence)

Yerel bağımsızlık, testle ölçülen yetenek, tekboyutlu olduğunda, aynı θ yetenek değerlerine sahip bir cevaplayıcı evreninde madde puanlarının şartlı dağılımının birbirinden bağımsız olmasıdır (Lord ve Novick, 1968; Hambleton ve diğ. 1978; Hambleton ve Swaminathan, 1985). Buna göre, yerel bağımsızlık, bir cevaplayıcının bir maddeye doğru cevap verme olasılığının testteki diğer maddelere vereceği cevaplardan etkilenmeyeceğini ifade eder (Crocker ve Algina, 1986). Yani, bir maddeye bir maddeye verilen doğru veya yanlış bir cevabın başka bir maddeye verilen cevabı etkilememesi o maddeye verilen cevabın yerel bağımsız olduğunu gösterir (Hambleton ve Swaminathan, 1985, 1985).

Yerel bağımsızlık varsayımı, tekboyutluluk varsayımı ile paraleldir. Bu paralellikten dolayı, yerel bağımsızlığın test edilmesinde de faktör analizi teknikleri kullanılmaktadır. Faktör analizi yöntemlerinden başka, Lord, x^2 tekniğinden de yararlanmayı önermiştir. Buna göre, giderek azalan standard sapma değerlerine göre, dar aralıklarda kalan yetenek gruplarındaki cevaplar her madde çifti için ele alınarak, madde puanlarının bağımsızlığının bir ölçüsü olarak x^2 istatistiği hesaplanarak, manidarlığı test edilir. (Hambleton ve Swaminathan, 1985 s:24). Bu süreç cevaplayıcıları 0 yetenek ölçeğinde farklı bölgelere yerleştiren kadar devam eder.

3. Madde Karakteristik Eğrisi (Item Characteristic Curve)

Madde puanının, θ yetenek ölçeği üzerinde regrasyonuna "madde karakteristik fonsiyonu" (MKF), bu fonksiyona ait eğriye de "madde karakteristik eğrisi" (MKE) adı verilir (Lord ve Novick, 1968). Bu matematiksel fonksiyon, o yetenek düzeyindeki cevaplayıcıların, herhangi bir test maddesinde, başarılı olma olasılıklarını verir. Bir cevaplayıcının bir maddeye doğru cevap verme olasılığı, sadece o maddenin MKE'sinin biçimine bağlıdır. Böylece ilgilenilen cevaplayıcı evreninde, cevaplayıcı yetenek dağılımı evrenden bağımsızdır. Buna göre, bir cevaplayıcının, bir maddeye cevap verme olasılığı, aynı yetenek düzeyinde başka kaç tane cevaplayıcı olduğuna bağlı değildir. Bir başka deyişle, MKE'sinin biçimi cevaplayıcı evrenindeki yeteneğin dağılımına dayalı değildir. Bu, MKE'sinin değişmezlik özelliği ve bu eğriyi tanımlayan parametrelerin de gruptan gruba değişmezliğini getirir (Hambleton ve diğ., 1978). Şekil 1'de 3 parametrelili lojistik modele ait bir madde karakteristik örneği verilmiştir.



Şekil 1: Madde Karakteristik Eğrisi

B. Örtük Özellikler Kuramında Modeller

Örtük özellikler kuramında, eğitimsel ve psikolojik test verilerinin analizinde kullanılan modeller, gözlenen parametrelerle (test maddesine verilen cevap), gözlenemeyen parametreler (ölçme konusu olan psikolojik özellik) arasındaki ilişkiyi tanımlayan olasılık modelleridir. Burada sözü edilen olasılık boyutu, modelin matematiksel formunun getirdiği bir boyuttur. Ancak, (Douglas, 1988). Kuramdaki her model:

1- Cevaplayıcının maddedeki edimi ile psikolojik özellik (yetenek) arasındaki ilişkiyi veren matematiksel bir eşitlikten,

2- Örtük özellikler kuramı çerçevesinde belirlenmiş, tanımlanmış temel varsayımlardan oluşur (Hambleton ve Swaminathan, 1985).

Kuramın ilk ortaya konusu ile birlikte üzerinde çalışılmaya başlanmasından bu yana, geçen zaman içinde, sayıları giderek artan çeşitli modeller geliştirilmiştir. Modeller, önceleri geliştiricisinin adıyla anılırken, (Rasch Model, Samejima Derecelendirilmiş Model gibi) bu durum bazı karışıklıklara (aynı kişinin birden fazla modele katkısı olma durumunda yaşanan sorunlar) sebep olmuş, sonraları ise, modelin matematiksel fonksiyonu esas alınarak isimlendirme (Normal Ogive Model, Lojistik Model gibi) yapılmıştır.

Bu noktada "model" teriminden ne kastedildiğini açıklamak yararlı olacaktır. Model, belli bir psikolojik özelliğin ölçülmesinde, bir test maddesine verilen cevapların dağılımının, koşullu olasılığı veya o maddeye

ve yetenek kestirimlerinde, bu kestirimlerle parametreleri arasındaki korelasyon, madde sayısının az olduğu koşullarda bile yüksek bulunmuştur. Madde kestirimlerindeki yanlışlığın ihmal edilebilir düzeyde olduğu, modelin bu ihlallere karşı güçlü (robust) kaldığı, öte yandan, şans hatasının parametre değerlerindeki yayılımı küçülttüğü, dolayısıyla da modelin, $c_i = \theta$ varsayımının ihlal edilmesinden daha fazla etkilendiği belirtilmektedir (Dinero ve Hertel, 1977; Viver, 1986; Davison ve Chen, 1991). Ayrıca Rasch modelinin, örneklemeden bağımsız madde parametreleri, maddelerden bağımsız yetenek parametreleri vermede başarısız olduğu yönünde tartışmalar da vardır (Chaplin, 1976 ve Wood, 1976: Adı geçen kaynak; Kline, 1986; 1986). Bunun yanında, Rasch modeli farklı örneklem büyüklüklerinde çalışıldığında, özellikle küçük örneklerde ($n < 200$), 2PL ve 3PL modele göre daha kullanışlı bulunmuştur (Wright, 1977; Whitley, 1977; Lord 1983; Vijver, 1986; Henning, 1989). Modelin küçük örneklerdeki kullanışlılığı, kestirim işlemlerinin daha az bilgi gerektirmesi ve işlemlerin az zaman alması yüzünden araştırmalarda sıkça kullanılmaktadır.

b) 2 Parametrelili Lojistik Model (2PL)

Birnbaum (1968), tarafından önerilen bu modelde, madde karakteristik eğrisi, 2 parametrelili lojistik dağılım fonksiyonudur. Modelin eşitliği,

FORMÜL 3 FİLM OLARAK GİRECEK

biçimindedir. Bu modelde, 2 parametrelili normal ogive eğrisi yerine önerilen, lojistik dağılımların en önemli avantajı, bu dağılımlara ait eğrilerin daha kolay çalışılabilir ve daha uygun (tractable) olmalarıdır. $P_i(\theta)$, b_i , a_i , ve θ esasen normal ogive modelindeki anlama sahiptir. D sabiti bir ölçme faktörüdür, $D=1.7$ olduğunda ve bu gösterildiğinde, $P_i(\theta)$ yetenek değerleri, 2p normal tipik özelliği, a_i ve b_i parametrelerini 2 Parametrelili model çerçevesinde kestirmesi, şans parametresini işleme koymasındır. (Hambleton ve diğ., 1978; Hambleton ve Swaminathan, 1986; Crocker ve Algina, 1985).

c) 3 Parametrelili Lojistik Model (3PL)

3 PL model, 2PL modelde, c_i , şans parametresi notasyonu eklenmesiyle ortaya çıkmıştır. 3 PL modelin eşitliği aşağıdaki gibidir:

Eşitlikte yeralan alan sembollerinin anlamları önceki modellerdeki anlamı taşıırken onlardan farklı olarak, c_i , şans parametresi tanımlanmıştır. c_i parametresi, madde cevap eğrisi ile x eksenindeki uzaklıktır ve düşük yetenekli cevaplayıcıların i maddesini şansla doğru cevaplama olasılığını gösterir. Modelde, c_i parametresinin bulunma amacı, yetenek ölçeğinin en alt ucunda, madde karakteristik eğrilerinin modelle uyuşmaması (misfitting) durumunda, bu faktörün, diğer faktörler arasında, test edimini etkileyen bir faktör olarak kontrol altında tutulmak istenmesidir (Lord ve Novick, 1968; Hambleton ve diğ., 1978; Crocker ve Algina, 1985). 3 PL model, diğer iki modele göre, kestirim işleminde daha çok bilgi kullanır, işlemler daha uzun, karmaşık ve zaman alıcıdır.

Ülkemizde Örtük Kuramı ile İlgili Yapılmış Araştırmalar

Ülkemizde, Örtük Özellikler kuramı konusuyla ilgili araştırmaların geçmişi eski olmamakla birlik, son yıllarda bu konuda yapılan araştırma sayısında bir artış gözlenmektedir. Bu konuda ülkemizde yapılan ilk araştırmada (Baykul, 1979), klasik test kuramı ve 3 PL modele göre, bir matematiksel yetenek testi için, kestirilen madde değerleri, testin geçerlilik ve güvenilirliğini sağlama bakımından karşılaştırılmış, 3 PL daha kararlı ölçme sonuçları verdiği saptanmıştır.

Üniversite öğrenci seçme sınavlarında kullanılan testlerde, objektif ölçmeyi sağlama ve artırmada Rasch modelinin katkısı, klasik test kuramıyla karşılaştırılarak bir araştırma çerçevesinde incelenmiştir (Berberoğlu, 1988). Bu araştırmanın bulguları, Rasch modeline dayalı kestirimlerin seçme testleriyle yapılan ölçmelerdeki objektifliği sağladığını, bu kestirimlerin hesaplanan geçerlilik ve güvenilirlik katsayılarının, klasik kuramdaki istatistiklerden hesaplanana göre daha yüksek olduğunu göstermiştir.

Diğer yandan, Gelbal (1994) tarafından yapılan araştırma ise, yine seçme amacıyla kullanılan iki testin Klasik Test Kuramı ve Rasch Modeline göre hesaplanan güçlük değerleri arasındaki ilişki yüksek bulunmuştur. Buna göre de, iki kuram çerçevesinde kestirilen madde güçlük değerlerinin önemli derecede farklı olmadıkları, bu yaklaşımlardan elde edilen madde güçlük değerlerinin birbirlerinin yerine kullanılmasının fark yaratmayacağı bulgusuna ulaşılmıştır.

Demirtaşlı-Çıkrıkçı (1995) tarafından yapılan araştırmada, bir genel yetenek testi olan Raven Standard İlerlemeli Matrisler Testi (RSİMT)'nin klasik test kuramı ve Rasch modeli çerçevesinde analiz edilerek, test maddelerinin Rasch modeline uyumu ile test maddelerinin özellikleri iki kuram çerçevesinde incelenmiştir. Bu araştırmanın sonuçları, RSİMT'nin Rasch modeline uyumunun zayıf olduğunu, verilerin normal dağıldığı koşulda, uyumunun arttığını göstermiştir. Ayrıca RSİMT maddelerinin iki

kurama göre yapılan madde analizlerinde, maddelerin güçlük sıralarının değişmediği bulunmuş, bu sonuç, Gelbal (1994)'ın temel bulgusu ile paralellik göstermektedir. Aynı çalışmada, testin yapısı ile ilgili olarak, testle ölçülen özelliğin tekboyutlu olması ve maddelerin güçlük düzeylerine göre kolaydan zora dizilme özelliği gösterdiğine, ilişkin iddiaların büyük ölçüde doğrulandığı bulunmuştur. Kriter dayanıklı testlerin geliştirilmesinde, Rasch Modeli ve yargıcı kararları yaklaşımıyla seçilen maddelerin farklılık gösterip göstermediği ve bu iki yöntemle geliştirilen testlerin geçerlilik ve güvenilirliklerinin farklı olup olmadığı incelenmiştir. (Gücüm, 1990). Sonuç olarak, Rasch Modeli ve yargıcı kararları yaklaşımı arasında, kriter dayanıklı test maddeleri seçiminde manidar bir fark gözlenmemiştir. ÖSYM tarafından yapılan ÖSS testleri, Rasch'a dayalı test eşitleme (test equating) yöntemi, 2 PL model ve doğrusal eşitleme yöntemleriyle eşitlenerek, testler için en uygun eşitleme yönteminin bulunmasını amaçlamıştır. (Kelecioğlu, 1994). Sonuçlar, Türkçe testleri için en uygun yöntemin doğrusal eşitleme; Sosyal Bilimler ve Matematik testleri için Rasch Modeli ile eşitleme; Fen Bilimleri testleri için de eşit yüzdelikli eşitleme olduğu görülmüştür. Ayrıca Kaptan (1993) tarafından yapılan çalışmada bireyselleştirilmiş test uygulaması (adaptive testing) ile kâğıt-kalem test uygulaması çeşitli sorular çerçevesinde karşılaştırılmıştır. Sonuçlar, iki uygulama için harcanan sürenin bireyselleştirilmiş test uygulamasında daha kısa olduğunu; iki uygulama yöntemiyle kestirilen yetenek düzeyleri arasında manidar fark olmadığını göstermiştir.

C. Örtük Özellikler Kuramının Uygulama Alanları ve Sınırlılıkları

Örtük Özellikler Kuramı türünde bir kuram, aslında eğitimsel ölçme tarihinin ilk dönemlerinde, 1930'larda 40'larda kabul edilmiştir. O günlerde de yetenek ile maddeye verilen cevap arasındaki ilişki bir normal ogive modeliyle daha iyi tanımlanabileceği varsayılmıştı. (Carroll, 1988). Bu modelin doğurguları, modelin gerektirdiği matematiksel kullanımların geliştirilmesi yüzünden, elde edilememiştir. Örtük Özellikler Kuramına dayalı çalışmaların eğitimsel ve psikolojik ölçme işlemlerinde, son yıllarda giderek genişleyen uygulama alanları vardır. Bunlardan başlıcaları, "test eşitleme (test equating), test (başarı ve yetenek testlere) ve ölçek (kişilik ve tutum ölçekleri) geliştirme, soru bankası oluşturma (item banking), bireyselleştirilmiş test hazırlama ve uygulama (tailored testing) çalışmalarıdır. Her kuramda olduğu gibi, bu kuram ve buna dayalı modellerin de sınırlılıkları, güçlükleri ve sorunlu boyutları vardır (Carroll, 1988 s: 248-249; Hambleton, 1989). Bunlar şunlardır:

1. İlki madde ve yetenek kestirimlerindeki sorunlardır. Klasik Maximum Likelihood işlemleri, geniş veri grupların ve fazla hesaplama işlemlerini gerektirir. Sonuçlar bazen anlamsız ve düşük güvenilirlikte çıkabilmektedir.

2. Örtük Özellikler Kuramında kestirim işlemleri, seçilen modellere göre yapılmaktadır. Seçilen modeldeki parametre sayısı arttıkça, kestirim ve hesaplama güçlükleri, sorunları doğmaktadır.

3. Diğer bir sorun, tekboyutluluk varsayımının akılcı bir varsayım gibi görünmemesindedir. Geleneksel olarak, boyut belirlemede faktör analizi tekniklerinden yararlanılmaktadır. Ancak, özellikle 0-1'le puanlarının maddelerde, atlanmış maddelerin bulunması analizde güçlük yaratır.

4. Bu kurama dayalı uygulamalar pahalıdır; özel bilgisayar paket programlarını ve bilgisayar kullanımını gerektirir.

SONUÇ

Örtük özellikler kuramına dayalı ölçme kuram ve uygulamaları, sayılan sınırlılıklarına rağmen, temelinde Matematiğin getirdiği kuramsal zenginliği barındırdığı için gelişmeye devam etmektedir. Bu gelişmeler çerçevesinde, ülkemizde, bu kuramla yapılacak Psikometri ile ilgili çalışma ve araştırmaların, özellikle psikoloji ve eğitim boyutunda hız ve ağırlık kazanması beklenmektedir.

KAYNAKLAR

- Anastasi, A.** (1988). *Psychological testing*. (6th ed.), New York: Mc Millan Publishing Company.
- Baykul, Y.** (1979). Örtük özellikler ve klasik test kuramları üzerine bir karşılaştırma. Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü, yayınlanmamış Doktora tezi, Mayıs, Ankara.
- Berberoğlu, G.** (1988). Seçme amacıyla kullanılan testlerde Rasch modelinin katkısı. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, yayınlanmamış Doktora tezi, Mayıs, Ankara.
- Carroll, J.B.** (1988). Future developments in measurement. In J.P. Keeves (Ed.) *Educational research, methodology and measurement an international handbook* (s. 247-252) Pergamon Press.
- Crocker, L.ve Algina, J.** (1986). *Introduction to classical and modern test theory*. N.Y: CBS College Publishing Company.
- Davison. M.L. ve Chen, T.** (1991). *Parameter invariance in the Rasch model* (ERIC Document Reproduction Service No. ED 333 043)
- Demirtaşlı Çıkrıkçı, Nühket** (1995). Rasch modelinin Raven stardard ilerlemeli matrisler testine uygulanması ve Klasik test kuramı ile karşılaştırılması. Ankara Üniversitesi Sosyal Bilimler Enstitüsü yayınlanmamış Doktora tezi, Haziran
- Dinero, T.E. ve Haertel, E.** (1977). Applicability of the Rasch model with varying item discriminations. *Applied Psychological Measurement*, 1, 581-592.

- Divgi, D.R.** (1986) Does the Rasch model really work for multiple choice items? Not if you look closely. *Journal of Educational Measurement*, 23, 283-298.
- Douglas, G.** (1988) Latent trait measurement models. In J.P. Keeves (Ed.) *Educational research, methodology and measurement: an international handbook*, (p 282-286): Pergamon Press.
- Gelbal, S.** (1994). P madde güçlük indeksi ile Rasch modelinin b parametresi ve bunlara dayalı yetenek ölçüleri üzerine bir karşılaştırma. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora Tezi, Ağustos, Ankara.
- Gustafsson, J.** (1980). Testing and obtaining of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33, 205-233
- Gustafsson, J.** (1988). Models of intelligence in J.P. Keeves (Ed.) *Educational research methodology and measurement. An international handbook*, (s 437-441): Pergamon Press.
- Gücüm, B.** (1990) Kriter dayanaklı testler için madde seçiminde iki farklı yöntem olarak Rasch modeli ve yargıcı kararlarının karşılaştırılması Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora tezi, Haziran, Ankara.
- Hambleton, R.K., Swaminathan, H., Cook, L.L., Eignor, D.R. and Gifford, J.A.** (1978). Developments in latent theory: models, technical issues, and applications. *Review of Educational Research*, Fall 48 (4), 467-510
- Hambleton, R.K. ve Swaminathan, H.** (1985). *Item response theory: principles and applications*. (2nd.eds.) Boston; Kluwer-Nijhoff Publishing
- Hambleton, R.K** (1989). *Item response theory introduction and bibliography*. (Research Report No 196) (ERIC Document Reproduction Service No ED 310 137.
- Henning, G.** (1989). Does the rasch work for multiple choice items? take another look: a response to Divgi. *Journal of Educational Measurement*, 26, 91-97.
- Kaptan, F.** (1993). Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü yayımlanmamış Doktora tezi, Ankara.
- Kelecioğlu, H.** (1994). Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine çalışma. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü yayımlanmamış Doktora tezi, Ağustos, Ankara.
- Kline, Paul** (1986). Computerized testing, tailored testing, Rasch scaling and cognitive process studies. In *A handbook of test construction: introduction to Psychometric design*. P. 193-205. Methuen & Co. Ltd.
- Lord, F.M. ve Novick, M.R.** (1968). *Statistical theories of mental test scores*. NY: Addison-Wesley Publishing Company.
- Lord, F.M.** (1983). Small n justifies Rasch model in D.J.Weis (Ed.). *New horizons in testing: latent trait test theory and computerized adaptive testing* (s.51-61). NY: Academic Press.

- Thissen, D ve Steinberg, L.** (1986). A taxonomy of item response models. *Psychometrika*, 51(4), 567-577.
- Vijver Van de, F.J.R.** (1986). The robustness of Rasch estimates. *Applied Psychological Measurement*. 10, 45-67.
- Whitley, S.E. ve Dawis, R.V.** (1974). The nature of objectivity with the Rasch model. *Journal of Educational Measurement*, 11, 163-178.
- Whitley, S.E.** (1977). Models, meanings and misunderstandings: some issues in applying Rasch's theory. *Journal of Educational Measurement*, 14, 227-235.
- Wilcox, R.R.** (1988). True score models. In J.P. Keeves (Ed.). *Educational research, methodology and measurement: an international handbook*. (s.267-269). Pergamon Press.
- Wright, B.D.** (1988). Measurement models. In J.P. Keeves (Ed.). *Educational research methodology and measurement: an international handbook* (s.286-292). Pergamon Press.