

p*MADDE GÜÇLÜK İNDEKSİ İLE RASCH MODELİNİN *b* PARAMETRESİ ve BUNLARA DAYALI YETENEK ÖLÇÜLERİ ÜZERİNE BİR KARŞILAŞTIRMA

Selahattin GELBAL**

ABSTRACT: The purpose of this study is to find out the similarities of both item parameters and ability parameters estimated by the Rasch model of latent traits theory and the classical test theory. The data used in this study are the responses to Turkish and mathematic tests given by the fifth grade primary school students. There found a high level of correlations between the unit normal values of item difficulty indexes and the *b*g parameter estimated by the Rasch model. In addition to this, there also appeared a high level of correlations among the ability levels estimated by the Rasch model, the unit normal values of item difficulty indexes that were used instead of the parameters in Rasch model and classical test theory. The correlations were higher in higher ability levels.

KEY WORDS: Rasch Model, Classical Test Theory, Item Difficulty, Item Parameters

ÖZET: Bu araştırmanın amacı, örtük özellikler teorisinin Rasch modeliyle ve klasik test teorisi ile kestirilen madde ve yetenek parametrelerinin benzer yönlerini ortaya koymaktır. Bu amaçla, ilkokul 5. sınıf için hazırlanmış Türkçe ve matematik testlerine öğrencilerin verdikleri cevaplardan yararlanılmıştır. Sonuçta, madde güçlük indekslerinin birim normal değerleri ile Rasch modeliyle kestirilen *b*g parametreleri arasında oldukça yüksek korelasyonlar bulunmuştur. Ayrıca, Rasch modeliyle ve Rasch modelinde *b*g parametresi yerine madde güçlük indekslerinin birim normal değerleri ve klasik test teorisi ile kestirilen yetenek ölçüleri arasında, özellikle ileri yetenek düzeylerinde yüksek korelasyonlar bulunmuştur.

ANAHTAR SÖZCÜKLER: Örtük Özellikler, Klasik Test Teorisi, Madde Güçlüğü

1. GİRİŞ

Bireylerin çeşitli alanlardaki özelliklerini ölçmede, yazılı ve sözlü yoklamalar, kısa cevaplı, sınıflama gerektiren ve çoktan seçmeli testler, ödev ve projeler günümüzde sıkça kullanılanlar arasındadır. Bu araçlardan hangisine başvurulacağı ölçülen değişkenin niteliğine, araçların özelliğine, uygulamanın yapılacağı öğrenci sayısına ve ekonomik sebeplere bağlıdır. Ölçme araçlarından elde edilen sonuçlar bireylerin ölçülen özellikleri hakkında değer yargılarına varmada kullanılır. Varılan değer

yargılarının isabetliliği, ölçme araç veya yöntemlerinin niteliklerine bağlıdır. Bir ölçme araç ve yönteminin nitelikleri güvenilirlik, geçerlik ve kullanılabilirlik kavramları altında gruplandırılmıştır (Turgut 1984:27). Bunlardan güvenilirlik ve geçerlik ölçme araçlarının vazgeçilmez iki özelliğidir. Bireyler hakkındaki kararların isabetlilik derecesi ölçütün uygunluğu yanında, ölçme sonuçlarının veya ölçme araç veya yöntemlerinin güvenilirlik ve geçerliğine bağlıdır. Eğitimdeki ölçmelerde yukarıda sözü edilen araçlardan çoktan seçmeli testler, bu iki özellik yönünden diğerlerine göre üstünlük sağlar. Bu bakımdan çoktan seçmeli testler, özellikle geniş gruplar üzerinde yapılan ölçme çalışmaları için halihazırda en geniş uygulama alanı bulan araç durumundadır. Çoktan seçmeli testler, istatistiksel olarak üzerinde çok çalışılmış ve teoriler geliştirilmiş ölçme araçlarıdır.

Bir testin puanlanabilen en küçük birimine madde adı verilir (Turgut 1978:2). Testin özellikleri onu oluşturan maddelerin özelliklerine bağlıdır. Bu bakımdan madde özelliklerini kontrol altında tutmak suretiyle test özellikleri kontrol altında tutulabilir. Testin özelliklerinin kontrol altında tutulmasında test özellikleri ile test özelliklerini madde özelliklerine bağlayan fonksiyonlardan yararlanır. Bu durum aşağıdaki eşitlikle ifade edilebilir.

$$\text{Test özellikleri} = f(\text{madde özellikleri}) \quad (1)$$

Böylece istenen özelliklere uygun testler yapma imkanı doğar. İstenen özelliklere uygun testler yapma, eğitimdeki ölçmelerde test geliştirme olarak adlandırılmaktadır. Test geliştirmede en çok başvurulan iki teoriden biri klasik test teorisi, diğeri örtük özellikler (latent traits veya item response) teorisidir.

Klasik test teorisinde testin özellikleri arasında testin aritmetik ortalaması, standart kayması, ortalama güçlüğü, güvenilirliği ve geçerliği önemli bir yer tutar. Bu teorideki belli başlı madde istatistikleri arasında madde güçlük indeksi (*p*_j), madde ayrımcılık gücü indeksi (madde geçerliği) (*r*_{jx}), madde güvenilirliği (*r*_j), madde standart ka-

* Araştırma verilerinin kullanılmasına izin veren UNICEF Ankara Temsilciliğindeki yetkililere teşekkür ederim.

** Dr. Selahattin Gelbal, Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı Araştırma Görevlisi.

yması (sj), maddeler arası kovaryans (cjk) ve maddeler arası korelasyon (rjk)'lar vardır.

Örtük özellikler teorisindeki madde özellikleri de seçilen örtük özellik modeline göre, (1) maddenin en iyi ölçtüğü yetenek düzeyi (bg) veya bununla beraber, (2) maddenin kalitesi hakkında bilgi veren ag veya bu ikisiyle beraber, (3) maddenin şansa doğru cevaplandırılma olasılığı olan cg parametresidir. Birinci maddede belirtilen bg parametresi bir parametrelili lojistik ve Rasch modellerinin, ikinci maddede sözü edilenler iki parametrelili lojistik modelin, üçüncü maddede sözü edilenler de üç parametrelili lojistik modelin parametreleridir.

Klasik test teorisindeki madde istatistikleri ile örtük özellikler teorisinin madde parametreleri arasındaki farklılıklardan biri klasik test teorisinde pj ve rjx'in testin geliştirildiği gruptan elde edilmesi karşısında, örtük özellikler teorisinin bg ve ag parametrelerinin seçilen modele göre matematiksel bir dağılım fonksiyonundan elde edilmesidir. Birçok yazara göre örtük özellikler teorisinin klasik teoriye üstünlüğü bu fonksiyon yardımıyla madde özelliklerinin gruptan bağımsız olarak hesaplanabilmesidir (Lord ve Novick 1968:379, Hambleton ve Swaminathan 1985:11).

Örtük özellikler teorisinin klasik teoriye göre yukarıda belirtilen avantajının yanında bu iki teori arasında benzerlikler de vardır. Nitekim klasik teorideki güçlük indeksi (pj) ile örtük özellikler teorisindeki maddenin ölçtüğü en iyi yetenek düzeyi parametresi (bg) ve klasik teoride madde ayırıcılık gücü indeksi olan rjx ile örtük özellikler teorisindeki maddenin kalitesini gösteren ag parametresi karşılıklı olarak aynı anlamlara gelmektedirler ve bu iki parametre çiftinin her birinden diğerine geçiş için eşitlikler Lord ve Novick (1968: 376-378) tarafından verilmiştir. Bu eşitlikler örtük özellikler teorisinin klasik test teorisine benzerliğini ifade etmektedir. Bu benzerliğe Weiss (1983: 23) de başka bir yönden değinmekte; örtük özellikler teorisinin klasik test teorisinden çıkarıldığını ve klasik test teorisinin çok basit bir örtük özellik modeli olduğunu belirtmektedir. Bu araştırma bir yönüyle, klasik test teorisinde normalize edilmiş güçlük indeksleriyle örtük özellikler teorisindeki bg parametresinin değerleri arasındaki benzerlik veya farklılığın çalışılması üzerinedir. Aynı zamanda, klasik test teorisine kestirilen yetenek ölçüleri, Rasch modeliyle ve Rasch modelinde bg parametresi yerine madde güçlük indeksinin birim normal değerleri (zg) kullanılarak kestirilen yetenek ölçülerinin karşılaştırılması da konu edilmektedir.

Yetenek ölçülerinin gruba bağımlılığı ve bunun sonucu olarak yorumlanmasında karşılaşılan güçlükler, yete-

nek ölçülerinin gruptan bağımsız olarak kestirilmesi çalışmalarına yol açmıştır. Örtük özellikler teorisine dayalı yetenek kestirme çalışmalarının böyle bir problem-den doğduğu söylenebilir.

Örtük özellikler teorisinin esası, bireylerin yeteneklerinin gösterildiği psikolojik uzay ile bu uzayı temsil etmek amacıyla hazırlanan madde uzayı arasındaki bağıntılarla ilgilidir. Örtük özellikler teorisi, test geliştirmenin yanında bireylerin çalışılan özelliklerini, bir birim ve izafi bir başlangıç noktasına göre ölçülerini hesaplamaya da yarar. Bireylerin doğrudan gözlenemeyen yetenek ölçüleri (θ), test maddelerine verdikleri cevaplarla gözlenmeye çalışılır. Bu işlemde, $P(\theta)$ ile belirtilen ve adına madde karakteristik fonksiyonu denilen, hipotetik fonksiyonlardan yararlanır.

$$P(\theta) = c_g + \frac{1 - c_g}{1 + e^{-1.7a_g(\theta - b_g)}} \quad (1.2)$$

Bu fonksiyonlardaki madde parametrelerinin sayısına göre değişik modeller ortaya çıkar.

Maddenin doğru cevaplandırılma olasılığı hesaplanırken madde karakteristik fonksiyonunda ag, bg ve cg parametrelerinin üçü birden kullanıldığında model, üç parametrelili lojistik model adını almakta ve madde karakteristik fonksiyonu, (1.2) eşitliği ile yazılmaktadır (Hambleton ve Swaminathan 1985: 37). Bu modelde şans başarısının sıfır ($c_g=0$) sayılmasıyla iki parametrelili lojistik model; maddelerin ag parametrelerinin birbirine eşit ve 1 olarak alınmasıyla bir parametrelili ya da Rasch modeli elde edilir (Hambleton ve Swaminathan 1985:37). Bir teste bütün maddelerin ayırıcılık gücü indekslerinin 1'e eşit olması çok zordur. Rasch modelindeki bu problem daha sonra ag parametresinin 1 olarak alınması yerine, ag parametrelerinin ortalamasının kullanılması suretiyle giderilmeye çalışılmış ve Rasch modeline ait denklem, (1.3) eşitliği ile ifade edilmiştir (Hambleton ve Swaminathan 1985: 47).

$$P(u_{vg} = 1) = \frac{1}{1 + e^{-Da(\theta_v - b_g)}} \quad (1.3)$$

(1.3) nolu eşitlikteki ag'nin bir ortalama değer olması sebebiyle maddelerin tek tek ag özelliklerini değil, testin tüm maddeleri için bir ortak özellik ifade eder. Böylece ag bir madde özelliği değil bir madde grubunun ortak özelliği olarak düşünülmektedir.

Gerek Rasch modelinde gerekse diğer modellerde bg parametresi, yukarıda da belirtildiği gibi, klasik teorideki

madde güçlük indeksine karşılık geldiği belirtilmiş ve bg parametresi için dönüşüm Lord ve Novick (1968:376-378)'te verildiği belirtilmişti. Belirtilen eşitlikler dikkatlice incelendiğinde, klasik test teorisindeki pg güçlük indeksinin madde karakteristik fonksiyonunun integralsiz halinin belli bir değerinden ag'ye kadar olan belirli integraline karşı getirildiği görülmektedir. Rasch ve diğer örtük özellikler modellerinin hepsinde tek boyutluluk (unidimensionality) ve normal dağılım bir sayıltı olarak yer aldığına göre, klasik teorideki pg'nin dönüşüm eşitlikleriyle hesaplanması ile klasik teorideki pg güçlük indeksi yerine Henryssen'in belirttiği gibi buna karşı gelen olasılığın kullanılması arasında belli bir ilişkinin bulunması beklenir. Bu araştırmada böyle bir ilişkinin var olup olmadığı, varsa bunun nasıl bir ilişki olduğu ilkökul 5. sınıf öğrencileri için Türkçe ve matematik alanlarında hazırlanmış olan iki test üzerinde çalışılmaktadır. Araştırmanın problemi aşağıdaki gibi ifade edilmiştir.

1.1. Problem Cümlesi

İlkokul 5. sınıf öğrencilerinin Türkçe ve matematikteki temel bilgi ve becerilerini yoklayan iki test yardımıyla, Rasch modeli kullanılarak kestirilen bg parametreleri ile klasik test teorisine kestirilen ve birim standart (zg) değerlere dönüştürülen madde güçlük indeksleri arasında bir ilişki var mıdır? Ayrıca, Rasch modeli, klasik test teorisi ve Rasch modelinde bg parametresi yerine madde güçlük indekslerinin zg değerleri kullanılarak kestirilen yetenek ölçüleri arasında nasıl bir ilişki vardır?

Araştırmanın problemi aşağıdaki alt problemlere bölünerek incelenmiştir.

Alt Problemler

1. İlkokul 5. sınıf Türkçe testindeki maddelerin Rasch modeline göre kestirilen bg parametreleri ile madde güçlük indekslerinin birim normal değerleri (zg) arasında nasıl bir ilişki vardır?

2. İlkokul 5. sınıf matematik testinde maddelerin Rasch modeli ile kestirilen bg parametreleri ile madde güçlük indekslerinin birim normal değerleri (zg) arasında nasıl bir ilişki vardır?

3. İlkokul 5. sınıf Türkçe dersi için hazırlanan testte, madde güçlük indekslerinin birim normal değerlerinin Rasch modelinde bg parametreleri yerine kullanılarak kestirilen yetenek ölçüleri, test ham puanları ve Rasch modeli ile kestirilen yetenek ölçüleri arasında nasıl bir ilişki vardır?

4. İlkokul 5. sınıf Matematik dersi için hazırlanan testte, madde güçlük indekslerinin birim normal değerlerinin

Rasch modelinde bg parametreleri yerine kullanılarak kestirilen yetenek ölçüleri, test ham puanları ve Rasch modeli ile kestirilen yetenek ölçüleri arasında nasıl bir ilişki vardır?

1.2. Sayıltılar ve Sınırlılıklar

Araştırmada Rasch modeli ile madde ve yetenek parametreleri kestirilirken, modelin kendi sayıltıları araştırma içinde kontrol edilmiştir. Bu sayıltıların dışında, öğrencilerin testleri cevaplandırırken yeterli motivasyona sahip oldukları ve başkalarından etkilenmedikleri de varsayılmaktadır. Araştırma, İlkokul 5. sınıf Türkçe ve matematik testlerine öğrencilerin vermiş oldukları cevaplar ve bu testlerdeki soruların ölçtükları davranışlarla sınırlıdır.

1.3. Araştırmanın Amacı

Bu araştırmanın amacı, klasik test teorisi ve örtük özellikler teorisinin Rasch modeliyle elde edilen madde ve yetenek parametrelerinin benzerliklerini ortaya koymaktır. Ayrıca, yetenek puanlarının dağılımının bu özellikleri etkileyip etkilemediğini görebilmektir.

1.4. İlgili Araştırmalar

Baykul (1979), klasik test kuramı ile geliştirdiği bir matematik yetenek testini, örtük özellikler kuramının üç parametrelili lojistik modeli ile geliştirdiği bir matematik yetenek testinin güvenilirlik ve geçerliklerini ve aynı öğrenci grubuna ait yetenek ölçülerini karşılaştırmıştır. Karşılaştırmada, puanların şans başarısından arıtılmış ve arıtılmamış hallerinin her ikisinde de, üç parametrelili lojistik model ile elde edilen puanlardan hesaplanan KR-20 güvenilirlik katsayısının klasik kuramla hesaplanandan daha yüksek çıktığını; fakat KR-21 güvenilirlikleri arasında manidar bir fark olmadığını görmüştür. Bu iki kuramdan elde edilen puanların genel yetenek puanları ile korelasyonuna bakıldığında, klasik kuramdan elde edilenin daha yüksek korelasyon verdiği fakat öğretmen notuyla korelasyonları arasında manidar bir fark olmadığını görmüştür.

Douglass, Khavari ve Farber (1979), klasik ve örtük özellikler madde analizi yöntemlerini karşılaştırmışlardır. Her iki yöntemle seçilen maddelerle ölçüt puanları arasındaki korelasyonların birbirine benzediği görülmüştür.

Thissen ve Wainer (1982), örtük özellikler teorisinin bir, iki ve üç parametrelili lojistik modelleriyle yapılan standart hataları karşılaştırmışlardır. Örneklem sayısındaki artış, yetenek parametrelerinin kestirilmesinde yapılan standart hatayı fazla etkilemediği görülmüştür. Standart hata, üç parametrelili lojistik model ve çok düşük yetenek düzeyinde en büyük; orta yetenek düzeyinin bir

ve iki parametrelili lojistik modellerinde yapılan standart hataların aynı ve üç parametrelili modelden daha düşük olduğu gözlenmiştir.

Swaminathan ve Gifford (1983:14-30), üç parametrelili örtük özellik modelinde madde parametrelerinin kestirilmesini incelemişlerdir. Yetenek dağılımı normal olan 10 maddelik testte a parametresinin gerçek değeri ile kestirilen değerleri arasındaki korelasyon 0,36 bulunmuştur. Ayrıca kestirilen ortalama gerçek değerlerin ortalamasından daha büyük bulunmuştur. Örneklemin ve test uzunluğunun artmasıyla korelasyonun arttığı, kestirilen ortalama ile gerçek ortalamasının birbirine yaklaştığı görülmüştür. 80 maddelik test ve 1000 kişilik cevaplayıcı durumunda a parametresinin kestirilen ve gerçek değerleri arasındaki en yüksek korelasyon 0,88 bulunmuştur. Bu eğilim, yetenek dağılımının her üçünde de aynı şekilde gözlenmiştir. Yetenek parametresinin kestirimi, test uzunluğu sabit tutulduğunda kestirilen ve gerçek değerleri arasındaki korelasyonu, örneklem büyüklüğünün etkilemediği görülmüştür. Diğer taraftan her üç dağılımda da test uzunluğunun büyük oranda artırılması yeteneğin kestirilen ve gerçek değerleri arasındaki uyumu etkilediği gözlenmiştir.

Hambleton ve Cook (1983:30-49), test uzunluğunun, yetenek dağılımının ve örneklem büyüklüğünün etkilerini bir iki ve üç parametrelili modellerde başarıya olan etkisini araştırmışlardır. Araştırmada, düşük yetenek düzeylerinde 20 maddelik test ve 3PL modelde kişilerin başarı sıraları arasında önemli ölçüde farklılık bulunmuştur. Tek biçimli (uniform) dağılımda korelasyonun 0,08 değiştiği (0,75-0,83), normal dağılımda da 0,08 değiştiği (0,65-0,73) görülmüştür. 40 maddelik testte düşük yetenek düzeyinde kişiler daha doğru bir sıraya konulmaktadır. Testin uzunluğu iki kat arttığında korelasyonda da bir artış görülmüştür.

Goldman ve Raju (1986), değişik büyüklüklerde alınan örneklemelerin kişi ve madde parametrelerinin doğruluğuna etkisini araştırmışlardır. Bir ve iki parametrelili model ile kestirilen yetenek parametrelerinin, farklı örneklemelerde evren parametresiyle çok iyi korelasyon verdiği gözlenmiştir. Madde güçlük parametreleri, bir parametrelili model için 250 denekli örnekleme, iki parametrelili model için ise 500 denekli örnekleme eşit bulunmuş; madde ayrıcalık indeksi için ise en az hatalı kestirim 1000 denekli örnekleme gerektirmiştir.

Yen (1987:275-291), Örtük özellikler teorisinde BILOG 2.2 ve LOGIST 5.0 ile kestirilen madde ve test parametrelerinin etkililiğini ve doğruluğunu karşılaştırmıştır. 10 maddelik testte BILOG'la kestirilen madde karakteristik fonksiyonu (ICF) ve test karakteristik fonksiyonu (TCF) daha doğru kestirilmekte, 20 ve 40 maddelik testlerde ise iki programın da eşit doğrulukta kestirdiği görülmüştür. Her iki programla kestirilen

madde parametreleri (b, a ve c) arasındaki korelasyonlara bakıldığında LOGIST ile kestirilenlerin daha yüksek olduğu görülmüştür. Yeteneğin normal dağıldığı 20 maddelik testte her iki programla kestirilen özellikler eşit doğrulukla kestirildiği; 40 maddelik testte LOGIST programının parametreleri daha iyi kestirdiği görülmüştür.

Seçme amacıyla kullanılan testlerde Rasch modelinin katkılarını inceleyen Berberoğlu (1988). Aynı sayıdaki madde ile Rasch modelinin kullanılmasındaki standart hatanın, ölçmenin standart hatasından daha az olacağını; fakat, Rasch modelinin yordama geçerliliğini artırmayacağını ifade etmektedir.

Cook, Eignor ve Taft (1988:31-45) örtük özellikler teorisindeki b parametresinin kestirimindeki kararlılığın klasik madde güçlük indeksi ile karşılaştırmıştır. Test formları 25004000 öğrenciye 3 kez uygulanarak, klasik teorideki güçlük indeksleri D (delta) değerlerine dönüştürülmüştür. Sonuçlara bakıldığında örtük özelliklerdeki b parametresinin D değerlerine göre daha az kararlılık gösterdiği görülmüştür. Bunun nedenleri arasında 3PL modelin kullanılmasının etkili olabileceği, araştırmacılar tarafından belirtilmektedir.

2. Yöntem

2.1. Araştırmanın Türü

Bu araştırmada, bir taraftan klasik test teorisindeki madde güçlük indeksleri ile Rasch modelinin bg parametreleri karşılaştırılmakta; diğer taraftan ise klasik test teorisine (ham puanlar), Rasch modeliyle (θ yetenek ölçüleri) ve Rasch modelinde bg parametresi yerine klasik test teorisinden elde edilen madde güçlük indekslerinin birim normal değerleri kullanılarak kestirilen yetenek ölçüleri karşılaştırılmaktadır. Bu araştırma, survey tipi araştırmalarda olduğu gibi, örneklem bilgilerinin bir evrene genellemesi olmayıp daha çok durum belirlemeye yönelik olduğundan bir temel araştırma niteliğindedir.

2.2. Verilerin Toplanması ve Analizi

Araştırmanın verileri Unicef, Mili Eğitim Bakanlığı (M.E.B.) ve Hacettepe Üniversitesi (HÜ) Eğitim Fakültesince yürütülen "İlköğretimde Temel Öğrenme İhtiyaçlarının Karşlanması" konulu araştırmada kullanılan, ilköğretim 5. sınıf Türkçe ve matematik derslerinde temel davranışları yoklayan testlere öğrencilerin verdikleri cevaplardır. Diğer bir deyişle, çalışmada Unicef, M.E.B. ve H.Ü. Eğitim Fakültesince yapılan bir araştırmanın verileri kullanılmıştır. Veri toplama aracı olarak kullanılan testler, araştırmanın pilot çalışmasında araştırmacılar tarafından klasik test teorisine dayanılarak geliştirilmiştir (Fidan ve Baykul 1991:4674). Testlerin KR-20 formülü ile hesaplanan güvenilirlikleri, Türkçe testi için 0,86 ve matematik testi için 0,78 olarak bulunduğu rapor edilmiştir. Testlerin geçerliği için, içlerinde deney-

imli ilkokul öğretmenleri ile üniversiteden seçilen program geliştirme ve ölçme uzmanları ile eğitim felsefecisi ve bilim adamlarından oluşan uzman kanısına başvurulmuştur. Araştırmanın verileri Adıyaman, Kastamonu, Tekirdağ ve Şanlıurfa illeri ve bu illerin çeşitli ilçelerindeki ilkokulların 5. sınıflarındaki öğrenciler üzerinden alınmıştır.

İlk olarak, verilerin Rasch modelinin sayıtlarını sağlayıp sağlamadığına bakılmıştır. Türkçe ve matematik testinin her ikisinden de alınan puanların normal dağılmadığı, kıkare testi ile bakılmıştır (Ünver ve Gammam 1986). Kontrol sonucunda her iki teste ait ham puanların normal dağılmadığı görülmüş ve bu şartı sağlayabilmek için bazı öğrencilerin cevap kağıtları araştırma kapsamı dışına alınmıştır. Normallik şartı bu şekilde sağlandıktan sonra testlerin tek boyutlu olup olmadığına bakılmıştır.

Bir testteki tüm maddeler aynı özelliği ölçüyorsa, bireyin testten aldığı puan, onun bu boyuttaki (uzaydaki) yerini gösterir. Bir ölçme aracının tek boyutlu olup olmadığına kontrolü konusunda Hambleton ve Swaminathan (1985:21), faktör analizi tekniklerinin kullanılabilmesini; Crocker ve Algina ise (1986:343) tüm grup üzerinden testi oluşturan bütün maddeler arasındaki korelasyonun hesaplanması ile bakılabileceğini belirtmişlerdir. Ayrıca, Lord (1980:21), tek boyutluluğun belirlenmesinde, tetrakorik korelasyonlarla hesaplanan korelasyon matrisiyle yapılan faktör analizinde, birinci faktör yüküne ait öz değerin (eigen value) ikinci faktör yüküne ait öz değere çok farklı olmasının ve ikinci faktör yüküne ait öz değerin diğerlerinden çok farklı olmasının bir ölçü olabileceğini belirtmektedir. Bu araştırmada tek boyutluluk, maddeler arasındaki tetrakorik korelasyonlar ve bu korelasyonlara uygulanan faktör analizi ile kontrol edildi. Faktör analizi sonucunda her iki testte de maddelerin çoğunun bir boyutta faktör yüklerinin en çok olduğu; ayrıca, Türkçe testinde, birinci boyuttaki faktör yüklerine ait öz değerin 8,09'den ikincisinde 1,95'e, matematikte ise bu değerler, 5,378'den 2,467'e düştüğü görülmüştür. Bu ani düşüşler ve maddelerin büyük bir çoğunluğunun birinci boyutta yüksek yük vermesi, Lord'un yukarıda belirtilen görüşü de dikkate alınarak testlerin her ikisi için de tek boyutlu olduğunu gösterir nitelikte olduğu söylenebilir. Fakat Türkçe testinin diğerine göre tek boyutluluğu daha fazla sağladığı anlaşılmaktadır. Testin tek boyutluluğu ile yakından ilişkili olan başka bir sayıtlı, maddelerin birbirinden bağımsız olarak cevaplandırıldığıdır.

Maddelerin birbirinden bağımsız olarak cevaplandırılması, maddelerden birine verilen doğru veya yanlış bir cevabın başka bir maddeye verilecek olan

cevabı etkilememesidir. Maddelerin yerel bağımsız olup olmadığına kontrolü için değişik yöntemlerden söz edilmiştir. Crocker ve Algina yerel bağımsızlığın test edilmesi için belli yetenek düzeylerindeki cevaplayıcılardan elde edilen madde puanları arasındaki ikili korelasyonlara bakılmasını tavsiye etmişlerdir (1986:342-343). Bu araştırmada, yerel bağımsızlık için Crocker ve Algina'nın belirttiği fikirler dikkate alınmıştır. Farklı yetenek düzeyi aralıkları için tüm madde çiftleri arasındaki tetrakorik korelasyonlara bakılmıştır. Her iki testteki maddeler için, tüm grup üzerinden hesaplanan korelasyonların, daha homojen olan diğer gruplardan hesaplanan korelasyonlardan daha büyük olduğu gözlenmektedir. Ayrıca, korelasyonların, 100 serbestlik derecesinde ve 0,01 manidarlık düzeyinde 0 (sıfır)'dan farklı olabilmesi için, 0,25'ten büyük olması gerekir (Akhun 1988:43). Türkçe ve matematik testlerinin her ikisinde de homojen gruplarda maddeler arasındaki tetrakorik korelasyonlardan çoğunun belirtilen kritik korelasyondan daha küçük olduğu gözlenmektedir. Bütün bu sonuçlara bakarak testlerdeki maddelerin yerel bağımsızlık şartını sağladığı söylenebilir.

Rasch modelinde maddelerin ag parametrelerinin 1'e eşit olduğu varsayılmaktadır. Bu varsayımın araştırmada kullanılan testlerin maddeleri için yerine gelip gelmediği kıkare testi ile yoklanmıştır. Türkçe testindeki 40 maddeden 25'inin, matematik testinde de 47 maddeden 28'inin ag parametresiyle ilgili şartı 0,01 manidarlık düzeyinde sağlamadığı gözlenmiştir. Yapılan araştırmaların bir çoğunda, Rasch modeline uygun olan maddelerin sayısının çok az olduğu görülmektedir (McPherson 1983). Bu sonuçlara göre, Türkçe ve matematik testini oluşturan maddelerin Rasch modeline uygun olanlarının sayısı yarıdan az olmasına rağmen, yukarıda belirtilen araştırmalarda ifade edilen yüzdelerden daha iyi bir sonuç alındığı ifade edilebilir.

3. Bulgular ve Yorumlar

Bu bölümde araştırmanın bulguları ve bulgulara dayalı yorumlar alt problemlerdeki sıraya göre açıklanmıştır.

1. Alt Problem: İlkokul 5. sınıf Türkçe testindeki maddelerin Rasch modeline göre kestirilen bg parametreleri ile madde güçlük indekslerinin birim normal değerleri arasında nasıl bir ilişki vardır? Bu probleme cevap bulabilmek için klasik test teorisine ve Rasch modeline göre kestirilen madde güçlük indeksleri, test puanlarının normal dağılmadığı ve dağıldığı her iki durumda da elde edilmiş ve sonra bunlar arasındaki korelasyon katsayıları ve bunlara dayalı temel istatistikler Tablo 3.1'de verilmiştir.

Tablo 3.3

Matematik Testinden Alınan Puanlar Normalleştirildikten
Sonra Testi Oluşturan Maddelerin İstatistiksel
Özellikleri Arasındaki Korelasyonlar

| | p | z | b | S.H.(b) |
|-----------|--------|------|--------|---------|
| z | 0,998* | | | |
| b | 0,998* | | 1,000* | |
| Art. Ort. | 0,39 | 0,39 | 0,39 | 0,035 |

* P<0.01

Tablo 3.3'de matematik testi maddelerine ait güçlük indekslerinin birim normal değerleri (zg) ile bg değerlerine ait aritmetik ortalamaların birbirine eşit ve 0,39 olduğu; ayrıca, bu iki değer arasındaki korelasyonun 1,000 olduğu görülmektedir. Madde güçlük indekslerinin p değerleri ile diğer özellikler arasındaki korelasyonlar da 0,998 olarak bulunmuştur. Aritmetik ortalamaların birbirlerinin aynı olması ve değerler arasındaki korelasyonların yüksek oluşu, Türkçe testinde olduğu gibi matematik testinde de maddelerin zg ve bg değerlerinin birbirleri yerine kullanılabilirliğini gösterir. Test puanları normal dağılmadığı durumda, aynı işlemler tekrarlanmış ve sonuçlar Tablo 3.4'de verilmiştir.

Tablo 3.4

Matematik Testinden Alınan Puanlar
Normalleştirilmeden Önce Testi Oluşturan Maddelerin
İstatistiksel Özellikleri Arasındaki Korelasyonlar

| | p | z | b | S.H.(b) |
|-----------|--------|--------|------|---------|
| z | 0,998* | | | |
| b | 0,998* | 1,000* | | |
| Art. Ort. | 0,40 | 0,28 | 0,30 | 0,030 |

* P<0.01

Tablo 3.4 incelendiğinde, matematik testi puanlarının normal dağılmadığı durumda da madde güçlük indekslerinin (pg) zg ve bg parametreleriyle; ayrıca zg ve bg parametrelerinin kendi aralarında oldukça yüksek korelasyon görülmektedir. Bu sonuçların, Türkçe testindekiler ile tutarlı olduğu görülmektedir. Ayrıca, matematik testinin tek boyutluluk sayılısı açısından zayıf olmasının da bu benzerliği çok fazla etkilemediği söylenebilir.

Maddelerin Rasch modelinde bg parametrelerinin kestirilmesinde yapılan standart hata, örneklem sayısı (N=1556) daha az olan fakat test puanları dağılımının normal olduğu durumda, aritmetik ortalama 0,035'dir. Örneklem sayısının daha çok olduğu (N=2077) ancak puanlarının dağılımının normal olmadığı durumda bg nin kestirilmesinde yapılan ortalama standart hata 0,030 olarak bulunmuştur. Bu sonuçlar da, Türkçe testinde olduğu gibi matematik testinde de, bg parametresinin kestirilmesinde yapılan standart hatanın, test puanlarının normal dağılıp dağılmamasından çok örneklemin büyüklüğüne bağlı olduğunu gösterir niteliktedir.

Maddelerin iki ayrı teori ile kestirilen özelliklerinin birbirine yakın sonuçlar vermesi, birinin diğeri yerine kullanılabilirliğini gösterir. Ayrıca, örtük özellikler teorisinin önemli üstünlüğü madde parametrelerinin gruba bağımlı olmamasıdır. zg ve bg parametrelerinin birbirlerine çok benzemesi, klasik teori ile elde edilen madde özelliklerinin de Rasch modeliyle karşılaştırıldığında aynı üstünlüklere sahip olabileceğini söylemeye imkan vermektedir. Bunun yanında Türkçe ve matematik testlerinde maddelerin özellikleri, puanların normalleştirildiği durumda sırasıyla 1566 ve 1556 kişilik örneklem grubu üzerinden, puanların normalleştirilmediği durumda ise yine aynı sırayla 2072 ve 2077 kişilik örneklem grubu üzerinden hesaplanmıştır. Örneklemin büyük olması maddelerin özelliklerinin daha az hata ile kestirilmesine ve bu nedenle de iki farklı yöntemle kestirilen özelliklerin birbirine benzer sonuçlar vermesine neden olabilir.

Araştırmanın diğer bölümünde yetenek ölçüleri karşılaştırılmış ve aşağıda sunulmuştur.

3. Alt Problem: İlkokul 5. sınıf Türkçe için hazırlanan testte, madde güçlük indekslerinin birim normal değerlerinin Rasch modelinde bg parametreleri yerine kullanılarak kestirilen yetenek ölçüleri, test ham puanları ve Rasch modeli ile kestirilen yetenek ölçüleri arasında nasıl bir ilişki vardır? Araştırmanın bu alt problemde, üç değişik yöntemle yetenek düzeyleri hesaplanmıştır. Yetenek düzeylerinden birincisi, Rasch modeli ile kestirilen θ yetenek parametreleri; ikincisi madde güçlük indekslerinin birim normal değerleri (zg)'nin Rasch modelinde bg parametresi yerine kullanılarak kestirilen yetenek parametreleri (zg yeteneği) ve üçüncüsü de klasik test teorisi ile elde edilen test ham puanları (X)'dir. Bu yetenek puanları arasındaki Pearson momentler çarpımı korelasyon katsayıları, yüksek, orta ve düşük yetenek düzeylerinde hesaplanarak üçüncü alt problemdeki soruya cevap aranmıştır. Hesaplanan korelasyon katsayıları Tablo 3.5'te verilmiştir.

Her üç düzeyde de korelasyonların oldukça yüksek olmasına rağmen, yetenek düzeylerinin artmasıyla puanlar arasındaki korelasyonlarda da bir artış görülmektedir. Özellikle, Rasch modeli ile kestirilen yetenek puanları ile Rasch modelinde bg parametresi yerine zg kullanılarak kestirilen yetenek puanları arasındaki korelasyon, alt yetenek düzeyinde 0,923'iken orta ve üst yetenek düzeylerinde 1,00'e çıkması dikkat çekicidir. Bu durum, zg parametrelerinin kullanılmasıyla kestirilen yetenek düzeylerinin Rasch modeli ile kestirilen yetenek ölçülerini, özellikle orta ve yüksek yetenek düzeylerinde daha iyi temsil edileceğini gösterir.

Tablo 3.5

Türkçe Testinin Farklı Yetenek Düzeylerine Göre ve Üç Ayrı Yöntemle Hesaplanan Yetenek Ölçüleri Arasındaki Korelasyonlar

| Yetenek Düzeyleri | Yetenek Türleri | yet. (bg) | Test P. (X) |
|-------------------|--------------------|-----------|-------------|
| Alt | Test P. (X) | 0,937* | 0,835* |
| | θ yet. (zg) | 0,923* | |
| Orta | Test P. (X) | 0,968* | 0,970* |
| | θ yet. (zg) | 1,000* | |
| Üst | Test P. (X) | 0,989* | 0,989* |
| | θ yet. (zg) | 1,000* | |

* P < 0.01

4. Alt Problem: İlkokul 5. sınıf matematik dersi için hazırlanan testte, madde güçlük indekslerinin birim normal değerlerinin Rasch modelinde bg parametreleri yerine kullanılarak kestirilen yetenek ölçüleri, test ham puanları ve Rasch modeli ile kestirilen yetenek ölçüleri arasında nasıl bir ilişki vardır? Bir önceki alt problemde Türkçe testi için yapılanlar, araştırmanın son alt probleminde matematik testi için yapılmıştır. Matematik testine cevap veren öğrencilerin klasik test teorisine göre elde edilen test ham puanları (X), Rasch modeliyle (θ) ve Rasch modelinde zg parametresi kullanılarak kestirilen yetenek parametreleri bulunduktan sonra, bunlar arasındaki korelasyon katsayıları, değişik yetenek düzeyleri için hesaplanmıştır. Hesaplanan katsayılar Tablo 3.6'da verilmiştir.

Tablo 3.6

Matematik Testinin Farklı Yetenek Düzeylerine Göre ve Üç Ayrı Yöntemle Hesaplanan Yetenek Ölçüleri Arasındaki Korelasyonlar

| Yetenek Düzeyleri | Yetenek Türleri | θ yet. (bg) | Test P. (X) |
|-------------------|--------------------|--------------------|-------------|
| Alt | Test P. (X) | 0,663* | 0,486* |
| | θ yet. (zg) | 0,769* | |
| Orta | Test P. (X) | 0,722* | 0,611* |
| | θ yet. (zg) | 0,928* | |
| Üst | Test P. (X) | 0,873* | 0,929* |
| | θ yet. (zg) | 0,963* | |

* P < 0.01

Tablo 3.6 incelendiğinde, farklı yetenek grupları üzerinden elde edilen puanlar arasındaki korelasyonlarda, Türkçe testine göre oldukça düşüş gözlenmiştir. Bu düşüş, özellikle alt grup olarak aldığımız düşük yetenek düzeyinde daha belirgindir. Bütün yetenek düzeylerinde en yüksek korelasyonlar Rasch modeliyle kestirilen yetenek ile Rasch modelinde zg parametreleri kullanılarak kestirilen yetenek arasında bulunmuştur. Matematik testinde de, Türkçe testindeki kadar olmasa da yetenek düzeylerinin artmasıyla korelasyonlarda da bir artış gözlenmektedir. Bu sonuçlara, göre yetenek düzeyleri daha çok homojenleştirildiğinde üst gruptaki korelasyonlarda, artışın, alt gruptakinde ise azalmanın olması beklenir.

Matematik testi sonuçları Türkçe testindeki sonuçlarla karşılaştırıldığında genelde bir benzerlik görülürken, özellikle alt yetenek düzeyinde farklılaşma görülmektedir. Orta ve üst yetenek düzeylerinde ise Türkçe testindeki korelasyonlar kadar yüksek korelasyonlar bulunmamıştır. Bu durum, matematik testinin Türkçe testine göre normallik ve tek boyutluluk sayıtlarını daha az karşılamasından kaynaklanması sebebiyle açıklanabilir.

4. Sonuç ve Öneriler

Araştırmanın bu bölümünde, bulgulardan çıkan sonuçlara ve bu sonuçlar doğrultusunda yapılan önerilere yer verilmiştir.

4.1. Sonuçlar

Araştırmada, örtük özellikler teorisinin Rasch modeli ile elde edilen madde ve yetenek parametrelerinin, klasik

test teorisi ile elde edilen madde özellikleri ve yetenek ölçülerinin karşılaştırılması yapılmıştır.

Araştırmanın ilk iki alt problemi, Türkçe ve matematik testlerinde Rasch modeliyle kestirilen madde parametre (bg)'lerinin klasik test teorisi ile belirlenen madde güçlük indekslerinin karşılaştırılmasına yöneliktir. Rasch modeliyle kestirilen bg parametrelerinin birim normal dağılıma sahip olduğu bilinmektedir. Madde güçlük indeksleri 0,00 ve 1,00 arasında değerler alabilen ve aynı zamanda maddenin doğru cevaplandırılma olasılığını veren bir istatistiktir. Bu istatistik, birim normal değerlere dönüştürüldüğünde Rasch modeliyle kestirilen ve maddenin en iyi ölçtüğü yetenek düzeyini veren bg parametresine ne derece benzediği araştırılmıştır. Araştırmanın bulgularına bakıldığında, bu iki özellik arasında çok az bir farklılığın olduğu görülmüştür. Bu farklılık ancak yüzde ve binde düzeyindedir. Ayrıca, maddelerin her iki yöntemle kestirilen özelliklerini betimleyen istatistiklerden aritmetik ortalama ve standart kaymaları birbirine çok yakın bulunmuştur. İkisi arasında benzerlik ölçüsü olarak hesaplanan Pearson momentler çarpımı korelasyon katsayısı 1,00'dir. Bu durum, Türkçe ve matematik testindeki maddelerin klasik test teorisi ile kestirilen madde güçlük indekslerinin birim normal değerlerinin, Rasch modeli ile kestirilen bg parametrelerinin aynı olduğu sonucuna götürür.

Test puanları normal dağılmadığında, maddelerin her iki yöntemle elde edilen özellikleri arasındaki korelasyonun, test puanlarının normal dağıldığı durumdaki ile aynı olduğu görülmüştür. Bu ise, test puanlarının normal dağılıp dağılmamasının klasik test teorisi ve örtük özellikler teorisinin Rasch modeliyle kestirilen madde özellikleri (madde güçlük indeksi ve bg parametresi) arasındaki benzerliği büyük ölçüde etkilemediği sonucuna götürür.

Rasch modeli ile maddelerin bg parametreleri kestirilirken yapılan standart hatanın, test ham puanlarının normal dağılıp dağılmamasına değil parametrenin kestirilmesinde kullanılan örneklemin büyüklüğüne bağlı olduğu görülmektedir. Örneklemin daha büyük olduğu, ancak test ham puanlarının dağılımının normal olmadığı durumda, bg parametresinin kestirilmesinde yapılan standart hatanın, test ham puanlarının normal dağıldığı fakat daha küçük örneklem üzerinden kestirilen bg parametresinin kestirilmesinde yapılan hatadan daha küçük bulunmuştur. Bu ise, beklentiye uygun olarak daha büyük örneklemelerde madde parametrelerinin daha az hata ile kestirilebileceğini işaret etmektedir.

Değişik yöntemlerle kestirilen yetenek ölçüleri arasındaki korelasyonların özellikle Türkçe testinde oldukça yüksek olduğu bulunmuştur. Bu benzerlik, değişik yetenek düzeylerinde kontrol edilmiş ve yetenek düzeyi arttıkça korelasyonlarda da artış görülmüş. Bu artışın Türkçe testinde daha yüksek olduğu görülmüştür. Bu durum, matematik testinin tek boyutluluk ve test puanlarının normallik şartlarını sağlamada daha zayıf olmasıyla açıklanabilir.

4.2. Öneriler

1. Klasik test teorisi ile kestirilen madde güçlük indeksleri ve güçlük indekslerinin birim normal değerlerinin (zg) örtük özellikler teorisinin Rasch modeli ile kestirilen bg parametreleri ile çok yüksek korelasyon vermesi, bunlar arasındaki benzerliği göstermektedir. Bu nedenle, zg parametrelerinin bg parametreleri yerine kullanılabilmesi önerilebilir.

2. Parametre kestirilirken, örneklem mümkün olduğunca büyük alınmalıdır.

3. Klasik test teorisi ile elde edilen test ham puanlarının ve madde güçlük indekslerinin birim standart değerleri (zg) Rasch modeli eşitliğinde kullanılarak kestirilen yeteneklerin Rasch modeliyle kestirilen yetenek ölçüleri yerine kullanılması tavsiye edilebilir.

4. Korelasyonların yetenek düzeyleri ile artması, benzerliğin düşük yetenek düzeylerinde az olduğunu gösterir. Özellikle düşük yetenek düzeylerinde yeteneklerin birbirleri yerine kullanılmasında dikkatli olunmalıdır.

KAYNAKÇA

- AKHUN, İlhan (1988). İstatistiksel Formüller ve Tablolar (Geliştirilmiş Üçüncü Baskı). Ankara:
- BAYKUL, Yaşar (1979). "Örtük Özellikler ve Klasik Test Kuramları Üzerine Bir Karşılaştırma", (Yayımlanmamış Doktora tezi). Ankara: Hacettepe Üniversitesi.
- BERBEROĞLU, Giray (1988). "Seçme Amacıyla Kullanılan Testlerde Rasch Modelinin Katkıları", (Yayımlanmamış doktora tezi), Ankara: Hacettepe Üniversitesi.
- COOK, L., D. R. EIGNOR ve H. L. TAFT (1988). "A Comparative Study of the Effects of Recency of Instruction on the Stability of IRT and Conventional Item Parameter Estimates", Journal of Educational Measurement, Vol. 25, No. 1, s. 31-45
- CROCKER, L. ve J. ALGINA, (1986). Introduction to Classical and Modern Test Theory, CBS College Publishing, New York.

- DOUGLAS, F. M.; A. K. KHAVARI ve P. D. FARBER (1979). "A Comparison of Clasical and Latent Trait Item Analysis Procedures", Educational and Psychological Measurement, Vol. 39, s. 337-352.
- FIDAN, N. ve Y.BAYKUL (1991). "İlköğretimde Temel Öğrenme İhtiyaçlarının Karşılanması", Ankara.
- GOLDMAN, S. H. ve N. S. RAJU (1986). "Recover of One and Two Parameter Logistic Item Parameters: An Empirical Study", Educational and Psychological Measurement, Vol. 46, s. 11-21.
- HAMBLETON, R. K. ve H. SWAMINATHAN (1985). Item Response Theory: Priciples and Applications, Kluwer. Nijhoff Publishing a Member of the Kluwer Academic Publisher Group.
- HAMBLETON, R. K. ve L. L. COOK (1983). "Robustness of IRT Models and Effects of Test Length and Sample Size on the Precision of θ ", New Horizons in Testing, Academic Press, A Subsidiary of Hacourt Brace Jovanovich, Publishers, Edited by David J. Wiess.
- LORD, F. M. ve M. R. NOVICK (1968). Statistical Theories of Mental Test Scores, AddisionWesley Publishing Company, Educational Testing Service.
- LORD, Frederic M. (1980). Aplications of Item Response Theory to Pratical Testing Problems, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.
- MCPHERSON, D. (1983). "Using the Rasch Model to Evaluate Test Item for Grade 4 and Grade 7 Michigan Educational Assesment Program Criterion Referenced reading Tests", Dissertation Abstracts International, 44: 734A.
- MISLEVY R. J. ve R. D. BOCK (1990). BOLOG 3.0 Item Analysis and Scoring with Binary Logistic Models, Second Edition, Scientific Software, Inc., Moresville, USA.
- SWAMINATHAN, H. ve J. A. GIFFORD (1983). "Estimatiion of Parameters in the Three Parameter Latent Trait Model", New Horizons in Testing, Academic Press, A Subsidiary of Hacourt Brace Jovanovich, Publishers, Edited by David J. Wiess.
- THISSEN, D. ve H. WAINER (1992). "Some Standard Errors in Item Response Theory" Psychometrika, 47, s. 397-412.
- TURGUT, M. Fuat (1978). "EBB673 Eğitimde Ölçme Teknikleri" (Yayımlanmamış Ders Notu No. 3).
- TURGUT, M. Fuat (1984). Eğitimde Ölçme ve Değerlendirme Metodları, Ankara: Saydam Matbaacılık.
- ÜNVER, Ö ve H. GAMGAM (1986). Uygulamalı İstatistiksel Yöntemler, Ankara.
- WIESS, David J. (1983). "Latent Trait Theory", New Horizons in Testing, Academic Press, A Subsidiary of Hacourt Brace Jovanovich, Publishers, Edited by David J. Wiess.
- YEN, Wendy M. (1987). "A Comparison of the Efficiency and Accoracy of Bilog and Logist", Psychometrika, Vol. 52, No. 2, s. 275-295.