

# Erişî Testlerine Madde Seçiminde Klasik Test Kuramı ve Rasch Modelinin Karşılaştırılması

Yard. Doç. Dr. Giray BERBEROĞLU (\*)

Eğitim ve öğretim süreci içerisinde kullanılan testler ve ölçekler bir çok amaca hizmet etmektedir. Amaç, sınıf içinde ele alınan hedef ve davranışların yoklanması ve elde edilen sonuçlara göre eğitim ve öğrenim sürecinin tekrar gözden geçirilmesi ya da not vermek olduğunda, sınıf içerisinde kullanılan erişî testlerinden elde edilen puanların da sözkonusu kararlardaki isabeti arttıracak nitelikte olması kaçınılmazdır. Bir testten elde edilen puanların niteliği, o testteki maddelerin niteliği ile doğrudan ilgilidir. Bu nedenle, test geliştirme süreci içerisinde öncelikle testte yer alan maddelerin nitelikleri incelenmektedir.

Madde analizinde sıkça kullanılan yöntemler klasik test kuramına dayanmaktadır. Burada sözkonusu olan yöntem, test puanları ile düşük korelasyon veren maddelerin testten atılarak, tanımlanan amacı ölçen homojen bir ölçek geliştirmektir.

Klasik test kuramına alternatif olarak Örtük Özellik Kuramı (Lafent Trait Theory veya Item Response Theory) modelleri önerilmektedir (Hambleton ve Swaminathan, 1985). Bu kuram içerisinde ele alınan Rasch modeli bir test maddesine cevap verme olasılığını, maddenin güçlüğüne ve bireyin yetenek düzeyine bağlı bir fonksiyonla açıklamaktadır (Özçelik, Berberoğlu, 1989). Madde seçimi ise  $X^2$  uyum testi ile yapılmaktadır. Bu modele göre kabul edilebilir bir maddede, bireyin yetenek düzeyi arttıkça doğru cevap verme olasılığının da artması beklenmektedir. Böyle bir maddeye her bir yetenek diliminde doğru cevap verilme olasılığı lojistik fonksiyon eğrisini takip eden bir dağılım göstermekte ve bu eğriye madde karakteristik eğrisi adı verilmektedir (Keeves, 1988, s. 407).  $X^2$  uyum testi ile, ölçekleme sonucu madde güçlüğü ve yetenek düzeyinde elde edilen teorik dağılımla, o maddeye verilen cevapların oluşturduğu gözlenen puan dağılımı karşılaştırılmakta, bu iki dağılım arasındaki uyum arttıkça  $X^2$  değeri küçülmektedir.  $X^2$  uyum testi ile elenen bir maddenin, (1) yazımında teknik bir hata olduğu; (2)

(\*) O.D.T.Ü. Eğitim Fakültesi Öğretim Üyesi.

ayırıcılığın düşük ya da testteki diğer maddelerden farklı olduğu; (3) test ile ölçülen yeterlik dışında bir boyutu ölçtüğü; (4) şansla cevaplanabildiği; (5) testteki diğer maddelerden güç olduğu ya da yetenek dağılımı çok geniş bir kitleye uygulandığı, düşünülmektedir (Özçelik, Berberoğlu, 1989, s. 42).

Klasik kuramın ve Rasch modelin madde seçiminde birbirinden farklı yöntemler öne sürmesi birçok araştırmacıyı bu iki yöntemi karşılaştırmaya yönlendirmiştir.

Helsley, Suber ve Ryan (1977) Rasch modeli ile elenen maddelerin, klasik kuram yöntemleri ile de eleneceğini öne sürmüşler ve Rasch modelini klasik test kuramının matematiksel bir uzantısı olarak göstermişlerdir.

Douglass, Khavari ve Farber (1979) ise, Mac Andrew alkolizm ölçeğinde klasik test kuramını ve Rasch modelini karşılaştırmışlar ve bu iki yöntemle seçilen maddeler arasında önemli farklılıklar olduğunu göstermişler ve Rasch modelinin klasik test kuramının matematiksel bir uzantısı olmadığını belirtmişlerdir.

Masters (1984, 1988), klasik test kuramı ile Rasch modelinin bir çok durumda birbirine benzediğini, klasik kuram tarafından elenen bir maddenin Rasch modeli tarafından da eleneceğini göstermiştir,  $X^2$  uyum testinin sınıf içi testlere madde seçiminde uygun bir yöntem olabileceğini belirtmiştir.

Bazı araştırmacılar erişim testlerine madde seçiminde Rasch modelini diğer örtük özellik modelleri ile karşılaştırmışlar ve Rasch modelinin diğer modellere göre daha çok sayıda maddeyi elelediğini göstermişlerdir (Bock, 1979; Divgi, 1981; Albanese ve Forsyth, 1984).

Bu çalışmalardan da görüldüğü gibi, birçok yönleri ile klasik test kuramı ve Rasch modeli benzer özellikler göstermekte ancak madde seçiminde yöntemler arası bazı farklılıkları dikkate çekmektedir.

Bu araştırmanın amacı, sınıf içinde kullanılan erişim testlerinin geliştirilmesinde klasik test kuramına ve Rasch modeline göre seçilen maddeleri karşılaştırmaktır.

## Yöntem

Bu çalışmada kavrama ve uygulama düzeyinde, dört seçeneğe sahip 43 maddeden oluşan ve güvenirlik, geçerlik, madde yazım teknikleri konularını kapsayan bir Ölçme ve Değerlendirme testi kullanılmış, madde seçimi amacı ile sözkonusu test Orta Doğu Teknik Üniversitesi Eğitim Fakültesi, Yabancı Diller Eğitimi Bölümü 3. sınıfta olan ve Eğitimde Ölçme ve Değerlendirme dersini alan 64 kişilik bir gruba uygulanmıştır.

Tablo I'de test puanlarının ortalama, ortanca standart sapma, çarpıklık ve basıklık ölçüleri verilmektedir.

**Tablo I. Test puanlarının Ortalama, Ortanca, Standart Sapma, Çarpıklık, Basıklık Ölçüleri**

	Ham Puanlar
k	43
N	64
X	24.83
Ortanca	25
Standart Sapma	6.04
Çarpıklık	- 0.727
Basıklık	2.969

Tablo I'de görüldüğü gibi, testen elde edilen ham puan dağılımları negatif yönde çarpıktır. Bu ilk uygulamada elde edilen KR (20) güvenilirlik katsayısı 0.77 olarak bulunmuştur.

Klasik test kuramına göre madde seçimi madde test korelasyonuna (biserial correlation) göre yapılmış, madde-test korelasyonu 0.20'den az olan maddeler testten çıkarılması gereken maddeler olarak düşünülmüştür (Ebel, 1965, s. 399, Douglas, Khavari, Farber, 1979).

Rasch modeline göre yapılan madde seçimi  $X^2$  uyum testi ile yapılmış,  $\alpha = 0.05$  düzeyinde anlamlı  $X^2$  değerine sahip olan maddeler testten çıkarılması gereken maddeler olarak düşünülmüştür.

Analiz sonucunda her iki yöntemle göre testten çıkarılması gereken maddelerin birbirinden farklı maddeler olup olmadığı karşılaştırılmıştır.

Klasik madde analizi ve Rasch modeli analizleri MicroCAT (tm) - Testing System (1986) paket programları kullanılarak yapılmıştır.

### **Bulgular**

Klasik madde analizine göre her bir madde için elde edilen madde güçlük ölçüsü (p) ile, madde test korelasyonu ya da madde ayırıcılık ölçüsü (r) ve Rasch modeline göre her bir madde için elde edilen madde güçlük ölçüsü (b<sub>i</sub>) ile  $x^2$  değerleri Tablo II'de gösterilmiştir.

Tablo II. Her Bir Maddenin p, r, b<sub>j</sub> ve X<sup>2</sup> değerleri.

Madde No.	p	r	b <sub>j</sub>	X <sup>2</sup>
1	0.58	0.39	-0.607	11.629
2	0.75	0.64	-1.936	8.367
3	0.23	-0.08 **	1.927	25.093*
4	0.84	0.64	-2.943	8.445
5	0.81	0.47	-2.563	8.646
6	0.53	0.33	-0.285	15.704
7	0.78	0.60	-2.233	7.424
8	0.78	0.48	-2.233	7.430
9	0.33	0.27	1.142	15.715
10	0.39	0.14 **	0.682	24.103
11	0.53	0.64	-0.285	13.500
12	0.45	0.48	0.247	9.877
13	0.53	0.54	-0.285	13.464
14	0.70	0.71	-1.535	14.979
15	0.81	0.68	-2.563	12.343
16	0.64	0.35	-1.053	10.208
17	0.50	0.41	-0.072	6.812
18	0.50	0.64	-0.072	17.856
19	0.36	0.19 **	0.908	25.408*
20	0.11	0.65	3.429	10.611
21	0.75	0.26	-1.936	11.811
22	0.81	0.37	-2.563	10.363
23	0.70	0.36	-1.535	10.408
24	0.77	0.53	-2.081	10.578
25	0.48	0.31	0.034	10.021
26	0.38	0.40	0.794	8.517
27	0.84	0.54	-2.943	13.802
28	0.67	0.44	-1.288	12.264
29	0.34	0.53	1.024	7.982
30	0.47	0.24	0.149	7.132
31	0.23	0.25	1.927	21.877*
32	0.64	0.31	-1.053	9.835
33	0.47	0.01**	0.140	21.518*
34	0.53	0.12 **	-0.285	14.603
35	0.47	0.64	0.140	20.675
36	0.41	0.50	0.571	17.024
37	0.56	0.22	-0.499	20.336
38	0.58	0.47	-0.607	6.477
39	0.78	0.60	-2.233	7.628
40	0.78	0.52	-2.233	8.577
41	0.86	0.48	-3.159	9.478
42	0.63	0.39	-0.939	11.713
43	0.50	0.29	-0.072	8.797

\*  $\alpha = 0.05$  düzeyinde modele uyum vermeyen maddeler

\*\* r değeri 0.20'nin altında olan maddeler

Tablo II'de görüldüğü gibi klasik test kuramına göre testten elenmesi gereken maddeler 3, 10, 19, 33, 34 numaralı maddelerdir. Rasch modeline göre testten elenmesi gereken maddeler ise 3, 10, 19, 31, 33 numaralı maddelerdir. Her iki yönetime göre aynı sayıda madde elenmiştir. 31 ve 34 numaralı maddeler hariç her iki yönetime göre testten çıkarılması gereken maddeler aynıdır.

## Sonuç

Bu araştırmada sınıf içinde kullanılan bir erişim testinin geliştirilmesinde klasik test kuramına ve Rasch modeline göre seçilen maddelerin birbirinden farklı olup olmadığına bakılmıştır.

Tablo II'deki bilgiler yakından incelendiğinde klasik test kuramına ve Rasch modeline göre testten elenmesi gereken maddeler arasında büyük ölçüde benzerlik vardır. Ancak bir madde klasik kurama göre elenmiş buna rağmen  $X^2$  uyum testine göre teste konulması uygun görülmüş (madde no: 34), aynı şekilde bir madde de  $X^2$  uyum testine göre elenmiş, ancak klasik kurama göre teste konulması uygun görülmüştür (madde no: 31). Bu maddelerden 31 numaralı maddenin  $X^2$  değeri tablo değerine çok yakın bir değerdedir ve çok az bir farkla modele uyum göstermektedir. Maddenin  $p$  ve  $b_j$  değerlerine bakıldığında, testteki diğer maddelere göre oldukça zor olduğu,  $r$  değerinin ise 0.20'ye yakın olduğu dikkati çekmektedir. Bu nedenle 31 numaralı maddede şans başarısının sözkonusu olduğu ve bunun da  $X^2$  uyum testi ile belirlendiği söylenebilir (Wright, Panchapakesan, 1969). Ancak 34 numaralı madde düşük  $r$  değerine sahip olmasına rağmen Rasch modeline göre teste konulması uygun görülmüştür. 34 numaralı maddenin  $r$  değeri düşük olmasına rağmen gözlenen puanlardan elde edilen madde karakteristik eğrisinin lojistik fonksiyona uyduğu ve düşük yetenek düzeyindeki bireylerin maddeye doğru cevap verme olasılıklarının yetenek düzeyi yüksek bireylere nazaran daha düşük olduğu gözlenmiştir.

Tablo II'de, genel olarak zor gözükken ve  $r$  değerleri düşük olan maddelerin her iki yöntem tarafından da elendiğini, madde ayırıcılık ölçüleri ile  $X^2$  uyum testi arasında büyük bir benzerlik olduğunu gösteren kanıtlar vardır.  $r$  değeri yükseldikçe gözlenen puan dağılımlarının madde  $r$  değeri yükseldikçe gözlenen puan dağılımlarının madde karakteristik eğrilerini takip ettiği ve  $X^2$  uyum testine göre modele uyum gösterdiği söylenebilir. Ancak 34 numaralı maddede de görüldüğü üzere yalnızca  $r$  değeri maddenin Rasch modeline uyum verip vermeyeceğinin bir göstergesi olmamaktadır. Rasch modeline göre elenen maddeler testten atılıp klasik madde analizi yeniden yapıldığında ortalama madde güçlük ölçüleri ( $p$ ) 0.58'den 0.61'e yükselmekte, ortalama madde ayırıcılık ölçüleri de ( $r$ ) 0.42'den 0.46'ya

yükselmektedir. Ayrıca, testin güvenilirlik katsayısı da maddeler atıldıktan sonra 0.77'den 0.80'e yükselmektedir.  $X^2$  değerleri ve  $r$  değerleri arasında -.048'lik bir korelasyon vardır. Bu da madde ayırıcılık ölçüleri yükseldikçe  $X^2$  değerlerinin küçüldüğünü ve klasik kuram tarafından düşük ayırıcılık ölçüsü nedeni ile testten elenen maddelerin büyük olasılıkla Rasch modeli tarafından da eleneceğini göstermektedir. Bu bulgular Helsley, Suber, Ryan (1977), ve Masters'ın (1988) bulunduğu sonuçları destekler niteliktedir.

Bu araştırmanın bulgularından şu sonuçlara gidilebilir: Klasik test kuramı ve Rasch modeline göre madde seçiminde bazı yöntem farklılıkları vardır, ancak bu çalışmada kullanılan teste seçile maddelerde klasik test kuramı ve Rasch modeli arasında büyük farklılıklar gözlenmemiştir. Ayrıca  $X^2$  uyum testi tek boyutlu sınıf içi testlerinin geliştirilmesinde uygun bir yöntem olarak kullanılabilir.

Söz konusu araştırma probleminin daha büyük örneklem gruplarında ve farklı testlerde incelenmesinde yarar görülmektedir.

#### KAYNAKLAR

Albanese, M.A., Forsyth, A.R. (1984) "The One-Two-and Modified Two-Parameter Latent Trait Models: An Empirical Study of Relative Fit". **Educational and Psychological Measurement**. 44. ss. 229-245.

Bock, R.D. (1979). "A Feasibility Study of the One- two-and three parameter Item-Response Models for the Analysis and Reporting of the California Assessment Data". Alındığı Kaynak: Albanese ve Forsyth. **Educational and Psychological Measurement**. 1984: 44.

Divgi, D.R. (1981) "Model-free Evaluation of Equating and Scaling". Alındığı Kaynak: Skaggs ve Lissitz. **Review of Educational Research**. 1986: 56.

Douglas, F.M., Khavari, K.A., Farber, P.D. (1979). "A Comparison of Classical and Latent Trait Item Analysis Procedures". **Educational and Psychological Measurement** 39. ss. 337-352.

Abel, R.L. (1965). **Essentials of Educational Measurement**. New Jersey: Prentice-Hall, Inc.

Hambleton, K.R., Swaminathan, H. (1985). **Item Response Theory: Principles and Applications**. Boston: Kluwer. Nijhoff Publishing.

Helsley, T.L., Suber, J.S., Ryan, J.P. (1977). "Item Homogeneity Unidimensionality and Test Reliability: Comparing Classical and Latent Trait Psychometric Procedures". Alındığı Kaynak: Douglass Khavari, Farber. **Educational and Psychological Measurement**. 1979: 39. s. 339.

Keeves, J.P. (1988). "Scaling Achievement Test Scores". Keeves, J.P. (Ed.) (1988). **Educational Research Methodology and Measurement: An International Handbook**. New York: Pergamon Press.

Masters, G.N. (1984) "DICOT: Analyzing Classroom Tests With the Rasch Model" **Educational and Psychological Measurement**. 44. ss. 145-150.

Masters, G.N. (1988). "Item Discrimination: When More is Worse". **Journal of Educational Measurement**. 25-1. ss. 15-29.

MicroCAT (tm) - Testing System (1986). Assessment Systems Corporation.

Özçelik, D.A., Berberoğlu, G. (1989). "Ölçme Objektiflik ve Rasch Modeli". **Eğitim ve Bilim**. 13-72. ss. 34-44.

Thorndika, R.L. (1982). **Applied Psychometrics**. Boston: Houghton Mifflin Company.

Wright, B.D., Panchapakesan, N. (1969). "A Procedure for Sample-free Item Analysis". **Educational and Psychological Measurement**. 29: 23-48.