

DAVRANIŞ ÖLÇÜMÜNDE YAPISAL GEÇERLİK GÖSTERGESİ

Ali BAYKAL
Boğaziçi Üniversitesi

ÖZET

Bu çalışmada, iki ya da daha çok boyutlu bir testin yapısal geçerliğini kendi içindeki verilerle sayısallaştıracak bir gösterge önerilmektedir. Test maddelerinin, ölçülen boyutlara dağılımı ile ölçmeleri beklenen boyutlara dağılımı arasındaki tutarlılığın, dağılımdaki toplam belirsizliğe oranı testin yapısal geçerliği olarak tanımlanmıştır. Tanımı gereği, yapısal geçerlik göstergesi 1.00 ile 0.00 değerleri arasında değişir. Maddeleri beklenen boyutlara dağılımı ile gözlenen faktörlere dağılımı birbirine tam olarak denk düştüğünde gösterge 1.00 değerini alır. Göstergenin en küçük kuramsal değeri 0.00'dur. Bu değer maddelerin beklenen ve gözlenen boyutlara çapraz dağılımının tümüyle rastlantısal, gelişigüzel ve seçkisiz olduğunu yansıtır.

Anahtar sözcükler: Yapısal geçerlik, çok boyutlu testler

ABSTRACT

Items of a multi-dimensional test can be cross-tabulated with respect to expected versus observed categories of factors. Some entropy (uncertainty) measures can be calculated within such a contingency table. Construct validity of the test is defined as the ratio of uncertainty removed by the observed distribution over the total uncertainty observed in the distribution. This ratio yields an index varying between unity and zero. Unity implies perfect construct validity, i.e., distribution of items into observed factors match perfectly well to the factors set before. Index will yield zero when the distribution of items into observed factors is completely independent of their distribution within the expected categories.

Key words: Construct validity, multi-dimensional tests

GİRİŞ

Ölçme bazı araçlarla ve belirli işlemlerle gerçekleştirilir. Davranış bilimlerindeki ölçmelerde en yaygın kullanılan araçlar testlerdir. Testlerin geçerli olmasını sağlamanın ve saptamanın önemi açıktır. Geçerlik, genel anlamda ölçülebilen ile ölçülmek istenen arasındaki tutarlılık (özdeşlik, uzlaşma, örtüşme, çakışma, bağdaşma) düzeyidir. Başka bir deyişle, ölçme sonuçlarının ölçmenin amacını gerçekleştirme derecesidir. Bu demektir ki, bir ölçmenin geçerliği amacından bağımsız olarak değerlendirilemez. Bir ölçme için geçerli olan bir test başka bir amaç için yeterince geçerli olmayabilir.

Geçerliğin çeşitli türleri içinde yapısal geçerlik özel önem taşır. Yapısal geçerlik ölçülen her birim davranış

için tek tek saptanması gereken bir niteliktir. Örneğin, bir testte 10 maddeden 9 tanesi içerik ve düzey olarak ölçülmek istenilen davranışlara uygun iken bir tanesi aykırı, çelişkili, tutarsız ya da ilişkisiz olabilir.

Yapısal geçerliğin niceliğini ölçen tek ve kesin bir yöntem yoktur. Ölçülmek istenen boyutun yapısal özellikleri iyi bilindiğinde bu nitelik aynı yapıda bir araçla karşılaştırılarak ölçülebilir. Örneğin, uzunluk boyutu önceden belirlenmiş bir başka uzunlukla (arşın, metre, mil) karşılaştırılır. Ağırlık yine önceden belirlenmiş bir başka ağırlıkla tartılır. Davranış bilimlerinde ölçülen niteliklerin yapıları ise ancak öznelere davranışları ile tanımlanabilmektedir. Özellikle, bilişsel ve duyuşsal alandaki pek çok niteliğin gözlenebilir davranışlara yansımaları ise çok çetrefildir. Bu yüzden;

1. Testin içerdiği bilişsel ya da duyuşsal süreçleri çözümlenerek,

* Bu çalışma Boğaziçi Üniversitesi Araştırma Fonunca desteklenen 92D103 kodlu proje ürünlerinden biridir.

Yazışma Adresi: Prof. Dr. Ali Baykal, Boğaziçi Üniversitesi Eğitim Bilimleri Bölümleri, 80815 Bebek, İstanbul

2. Ölçülen grupta elde edilen verileri ölçüt grup verileri ile karşılaştırarak,

3. Öntest-sontest karşılaştırmaları yaparak,

4. Geçerli olduğu kabul edilen başka ölçütlerle karşılaştırarak,

yapısal geçerlik hakkında destekleyici kanıt toplamak gerekir. Bu yöntemlerin ilk üçünde yapısal geçerlik için nicel bir gösterge elde edilemez. Dördüncü yöntemde istatistiksel ilişki katsayıları kullanılabilir. Ancak istatistiksel ilişki katsayılarının kendi yetersizlikleri görmezlikten gelinse bile ilişkinin niceliği yapısal özdeşlik dışındaki nedenlerden kaynaklanabilir. Bir başka deyişle iki değişken arasındaki ilişki koşut geçerliğin ölçüsü olabilir. Bu olasılık da yok sayılsın ve iki ölçüm arasındaki ilişkinin tamamıyla yapısal özdeşlikten ortaya çıktığı kabul edilsin. Daha da ötesi, ilişki katsayısının 1.00, yani "mükemmel" olduğu tasarlansın. Böyle bir durumda da yeni testi

eski testten farklı bir şey ölçmediği itiraf edilmiş olur. İlişkinin tam olmaması durumunda ise ölçüt alınan eski testle açıklanamayan değişkenlik, yeni testin yapısal yetersizliğine bağlanacaktır. Bu çalışmada, iki ya da daha çok boyutlu bir testin yapısal geçerliğini kendi içindeki verilerle sayısallaştıracak bir gösterge önerilmektedir. Çalışmanın amacı bu göstergelyi tanıtmak, örneklendirmek ve irdelemektir.

Bir test, ölçmeyi amaçladığı yapıyı ölçebildiği oranda geçerlidir. Beklenen yapı ile gözlenen yapının özdeşlik derecesi testin yapısal geçerliğidir. Test maddelerinin, ölçülen boyutlara dağılımı ile ölçmeleri beklenen boyutlara dağılımı arasındaki tutarlılığın, dağılımdaki toplam belirsizliğe oranı testin yapısal geçerliğidir.

Yapısal geçerlik için önerilen gösterge, maddelerin beklenen ve gözlenen boyutlara çapraz dağılımındaki belirsizlik (entropi) değerlerine bağlı olarak (1) numaralı bağlantı ile tanımlanmıştır.

g: Önerilen yapısal geçerlik göstergesi

(1)

$$g = \frac{I(X, Y)}{H(X, Y)}$$

- K : Toplam madde sayısı
 B : Testte tanımlanan boyut sayısı
 N(X) : Tanımlanan boyut kategorilerindeki madde sayıları
 N(Y) : Gözlenen boyut kategorilerindeki madde sayıları.
 N(X, Y) : Maddelerin, beklenen ve gözlenen boyutlara çapraz dağılım tablosundaki örtüşen sayıları.
 I(X, Y) : Çapraz belirsizliğin, beklenen ve gözlenen dağılımların belirsizliğini çözümlenme ölçüsü.
 (2)
 I(X, Y) = H(X) + H(Y) - H(X, Y)
 H(X, Y) : Çapraz Belirsizlik: Test maddelerinin, tanımlanan ve beklenen boyutlara göre çapraz dağılımındaki toplam belirsizlik ölçüsü.

(3)

$$H(X, Y) = \log K - \frac{1}{K} \sum_{X=1}^B \sum_{Y=1}^B N(X, Y) \log N(X, Y)$$

H (X) : Beklenen belirsizlik: Test maddelerinin test yapımcılarınca tanımlanan boyutlara dağılımındaki belirsizlik ölçüsü.

$$(4) \quad H(X) = \log K - \frac{1}{K} \sum_{X=1}^B N(X) \log N(X)$$

H (Y) : Gözlenen Belirsizlik: Test maddelerinin faktör analizi ya da uzman kanışı v.b. yöntemlerle tanımlanan boyutlara dağılımındaki belirsizlik ölçüsü.

$$(5) \quad H(Y) = \log K - \frac{1}{K} \sum_{Y=1}^B N(Y) \log N(Y)$$

ÖRNEK : Beşer seçenekli 80 soruluk bir yetenek testi yapılmıştır. Test, aşağıdaki genel yetenek boyutlarını ölçmek için hazırlanmıştır.

ÇPS : Çeşitli Sayısal Problemleri Çözebilme.

SMG : Sözel Muhakeme Gücü.

GAK : Geometrik Alan Kestirimi.

MDSŞ : Matris, Dizi, Şifre Çözebilme.

CEB : Cebirsel Problemleri Çözebilme.

TAN : TANGRAM Kalıplarını Çözebilme.

TET : TETRİS Bileşimlerini Çözebilme.

KAL : Kaleidoskopik Şekilleri Çözümleyebilme.

Test uygulanmadan önce, her maddenin ölçmesi beklenen boyut zaten belirlenmiştir. Uygulamadan sonra, test verileri üzerinde faktör analizi yapılmıştır. Her maddenin hangi boyutu ölçtüğü, faktör analizi yerine, uzman kanışı ya da benzeri bir yolla saptanmış olabilirdi. Testteki her maddenin ölçmesi beklenen ve faktör analizi sonucunda gözlenen boyutları Tablo 1'de listelenmiştir.

Maddeler, gözlenen faktörlerdeki yerleri ile sınav öncesinde saptanmış yerlerine göre çaprazlanınca Tablo 2 elde edildi.

Tablo 2'deki dağılım üzerinde (2), (3), (4) ve (5) sayılı belirsizlik bağıntılarından yararlanılarak (1) sayılı bağıntıda tanımlanan "yapısal geçerlik" göstergesi hesaplanmıştır.

Tablo 1

80 Maddelik bir testte maddelerin beklenen ve gözlenen boyutlara dağılımı

MADDE	BEK	GÖZ.	MADDE	BEK.	GÖZ	MADDE	BEK.	GÖZ	MADDE	BEK.	GÖZ.
1.	ÇSP	F1	21.	GAK	F6	41.	CEB	F5	61.	TAN	F3
2.	ÇSP	F1	22.	GAK	F6	42.	CEBC	F6	62.	TAN	F3
3.	ÇSP	F1	23.	GAK	F6	43.	EB	F5	63.	TAN	F6
4.	ÇSP	F4	24.	GAK	F6	44.	CEB	F1	64.	TAN	F3
5.	ÇSP	F1	25.	MDŞ	F4	45.	SMG	F8	65.	TET	F2
6.	ÇSP	F5	26.	MDŞ	F1	46.	SMG	F8	66.	TET	F2
7.	ÇSP	F3	27.	MDŞ	F4	47.	SMG	F8	67.	TET	F2
8.	ÇSP	F1	28.	MDŞ	F5	48.	SMG	F8	68.	TET	F2
9.	ÇSP	F1	29.	MDŞ	F1	49.	SMG	F1	69.	TET	F2
10.	ÇSP	F8	30.	MDŞ	F4	50.	SMG	F4	70.	TET	F2
11.	ÇSP	F6	31.	MDŞ	F8	51.	SMG	F4	71.	TET	F2
12.	ÇSP	F6	32.	MDŞ	F4	52.	SMG	F8	72.	TET	F2
13.	ÇSP	F1	33.	MDŞ	F4	53.	SMG	F1	73.	KAL	F7
14.	ÇSP	F6	34.	MDŞ	F1	54.	SMG	F8	74.	KAL	F7
15.	ÇSP	F1	35.	MDŞ	F4	55.	SMG	F8	75.	KAL	F7
16.	GAK	F6	36.	CEB	F5	56.	SMG	F8	76.	KAL	F7
17.	GAK	F6	37.	CEB	F5	57.	TAN	F3	77.	KAL	F7
18.	GAK	F6	38.	CEB	F1	58.	TAN	F3	78.	KAL	F7
19.	GAK	F6	39.	CEB	F5	59.	TAN	F3	79.	KAL	F7
20.	GAK	F6	40.	CEB	F5	60.	TAN	F7	80.	KAL	F7

Bulgular aşağıdadır:

H (X)	: 0.892 Beklenen belirsizlik
H (Y)	: 0.888 Gözlenen Belirsizlik
H (X, Y)	: 1.201 Çapraz (Toplam) Belirsizlik
I (X, Y)	: 0.579 Çapraz Dağılımda Tutarlılık

Yapısal geçerlik göstergesi : $g = 0.482$

Tablo 2

Beklenen faktörler-gözlenen faktörler (Yetenek testi)

Gözlenen		F1	F2	F3	F4	F5	F6	F7	F8	TOPLAM
B ÇSP		8		1	1	1	3		1	15
E GAK							9			9
K MDŞ		3			6	1			1	11
L CEB		2				6	1			9
E SMG		2			2				8	12
N TAN				6			1	1		8
E TET			8							8
N KAL								8		8
TOPLAM		15	8	7	9	8	14	9	10	80

Göstergenin Alabileceği Sınır Değerler: Tanımı gereği, yapısal geçerlik göstergesi 1.00 ile 0.00 değerleri arasında değişir. Maddelerin beklenen boyutlara dağılımı ile gözlenen faktörlere dağılımı birbirlerine tam olarak denk düştüğünde gösterge 1.00 değerini alır. Bu durum Tablo 3'de örneklendirilmiştir. Burada görüldüğü gibi, ÇSP boyutunu ölçmesi beklenen 15 maddenin hepsi F1 faktöründe toplanmıştır. Ya da ölçmeci tarafından ÇSP küme-

sine konan 15 madde uzmanlar tarafından da F1 kümesi altında derlenmiştir. Burada yeniden belirtilmesi gerekir ki, göstergenin hesaplanabilmesi için faktör analizi zorunlu değildir. Maddelerin beklenen boyutlara dağılımı ölçmecilerin öznel tercihleridir. Bu tercihi yapmak onların hakkıdır. Ancak, tercihlerini uygulama öncesinde açıklamaları da bilimsel nesnellüğün gereğidir. Maddelerin gözlenen boyutlara dağılımı uygulama sonrasında

Tablo 3
Beklenen faktörlerle gözlenen faktörlerin tam tutarlılığı

Gözlenen	F1	F2	F3	F4	F5	F6	F7	F8	TOPLAM		
B ÇSP	15								15		
E GAK				9					9		
K MDS			11						11		
L CEB		9							9	H (X)	= 0.892
E SMG						12			12	H (Y)	= 0.892
N TAN					8				8	H (X, Y)	= 0.892
E TET								8	8	I (X, Y)	= 0.892
N KAL							8		8	g	= 1.000
TOPLAM	15	9	11	9	8	12	8	8	80		

faktör analizi ile saptanabileceği gibi, uygulama öncesinde uzman kanıları ile de belirlenebilir.

Göstergenin en küçük kuramsal değeri 0.00'dır. Bu değer maddelerin beklenen boyutlara dağılımı ile gözle-

nen boyutlara dağılımı arasında hiçbir bağlantı, ilişki, tutarlılık olmadığı durumda ortaya çıkar. Maddelerin beklenen ve gözlenen boyutlara çapraz dağılımının tümüyle rastlantısal, gelişigüzel ve seçkisiz olduğunu yansıtan bir

Tablo 4
Beklenen faktörlerin gözlenen faktörlerden tam bağımsızlığı

Gözlenen	F1	F2	F3	F4	F5	F6	F7	F8	TOPLAM		
B ÇSP	2	2	2	2	2	2	2	1	15		
E GAK	2	1	1	1	1	1	1	1	9		
K MDS	1	1	2	2	1	1	1	2	11	H (X)	= 0.892
L CEB	1	2	1	1	1	1	1	1	9	H (Y)	= 0.903
E SMG	1	1	1	1	2	2	2	2	12	H (X, Y)	= 1.783
N TAN	1	1	1	1	1	1	1	1	8	I (X, Y)	= 0.012
E TET	1	1	1	1	1	1	1	1	8	g	= 0.007
N KAL	1	1	1	1	1	1	1	1	8		
TOPLAM	10	10	10	10	10	10	10	10	80		

örnek Tablo 4'te verilmiştir. 0.00 değerinin tam olarak elde edilebilmesi için çapraz tablodaki tüm-kutucuklardaki frekansların birbirine eşit olması gerekir. Örneğimizdeki 80 sorununun 64 kutucuğa kesirsiz ve eşit olarak ayrılacağından örnekte tam 0.00 değeri çıkmamıştır.

Önerilen Göstergenin Anlamlılık Sınaması

Yapısal geçerlik göstergesi, bir testin maddelerinin ölçmesi beklenen boyutlardaki dağılımı ile gözlenen boyutlara dağılım arasındaki tutarlılığın gözlenen toplam belirsizliğe oranı olarak tanımlanmıştır.

Elde edilen sayısal değerlerin ne denli anlamlı olduğunun belirlenmesi için aşağıdaki denence kurulabilir. (Attneave 1959, s. 63-67).

$$D_0 : I_g(X, Y) = 0$$

$$D_A : I_g(X, Y) \geq 0 \text{ s.d.} = (sx - 1) (sy - 1)$$

Bu denencelerin sınanmasında da aşağıdaki bağıntıdan yararlanılabilir (Garner & McGill, 1956, 221). Hemen belirtmeli ki, gösterge Ki Kare dağılımının sayıtları ile sınırlı değildir.

$$X_g^2 = 1.386 * N * I_g(X, Y)$$

X_g^2 : Gözlenen Ki Kare değeri

N : Toplam gözlem sayısı $I_g(X, Y)$: Çapraz belirsizliğin, beklenen ve gözlenen dağılımların belirsizliğini çözümlene ölçüsü.

sx : Satır sayısı (Beklenen boyut sayısı)

sy : Dilim sayısı (Gözlenen boyut sayısı)

s.d. : Dağılımın serbestlik derecesi

$X^2_{.0}$: Ölçüt Ki Kare değeri.

Göstergenin kullanımı için testin en az iki boyutlu olması gereği açıktır. Göstergenin, tek boyutlu testlerde ve madde düzeyinde uygulanabilmesi için de çalışmalar yapılmaktadır.

Tablo 5
Uygulama Sonuçlarını Anlamlılık Düzeyleri.

Veri Kaynağı	$I_g(X, Y)$	N	s.d.	X_g^2	$X^2_{.0}$.05	.01
Tablo II	0.58	80	49	64.32	62	*	

KAYNAKLAR

- Attneave, F. (1959). *Applications of information theory to psychology*. New York: Henry Holt and Company.
- Baykal, A. (1980-1981). Entropy as a measure of error in achievement testing, *Boğaziçi University Journal: Education*. VIII-IX, 1-18.
- Baykal, A. (1992). *TANGRAM and TETRIS items in the measurement of ability*, Paper presented at The Third European Conference of The European Council for High Ability, Munich.
- Cronbach, L.J. (1971). Test validation. Thorndike, R.L. (Ed.) *Educational measurement*. Washington: American Council on Education.
- Garner, W.F. & McGill, W.J. (1956). The relation between information and variance analysis. *Psychometrika*. 21-3, 219-228.

- Guilford, J.P., (1954). *Psychometric methods*, New York: Mc Graw-Hill.
- Keeves, J.P. (Ed.) (1988). *Educational research. methodology, and measurement*. Oxford: Pergamon Press.
- Shannon, C.B. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: The University of Illinois Press.
- Thorndike, R.L. (Ed.), (1971). *Educational measurement*. Washington: American Council on Education.
- Zeller, R.A. (1988). Validity. In J.P.Keeves (ed.), *Educational research, methodology, and measurement* (pp. 322-329) Oxford: Pergamon Press.