

P_{CT}'nin Değişik n, k ve Yüzdelik Dilimlerdeki Değişimi ile ρ , Yarıya Bölme ve Değişik İndekslerle Karşılaştırılması

Adnan Erkuş
Mersin Üniversitesi

Özet

Ölçme araçlarının en önemli niteliklerinden olan geçerlik, son yıllarda ölçme ve karar geçerliği olarak iki grupta ele alınmaktadır. Bu çalışmada, karar geçerliği altındaki sınıflama ve sıralama geçerliği için önerilen P_{CT} indeksinin, ikili ve çoklu puanlanan ölçeklerde, değişik n, k ve % manipülasyonlarında ne tür değişim gösterdiği; ρ , yarıya bölme katsayısı ve κ ile Po gibi sınıflama geçerliği katsayılarıyla ilişkisi incelenmiştir. İkili puanlanan (Beck Umutsuzluk Ölçeği) ve çoklu puanlanan (Boyun Eğici Davranışlar Ölçeği) ve tekboyutlu duruma getirilen iki ölçeğin uygulandığı 161 kişilik örneklemden elde edilen veriler üzerindeki analiz sonuçları; indeks için en iyi kesme alanının %27'lik alan olduğunu; ikili ölçek için n azaltıldığında sabit k düzeyleri için indeks değerinin azaldığını, çoklu ölçekte ise tersine arttığını; n sabit k azaldığında ise n manipülasyonunun tam tersi bir değişim olduğunu göstermiştir. P_{CT} indeksine Spearman-Brown düzeltmesi yapılarak geleneksel yarıya bölme katsayısı ve ρ ile karşılaştırılmış ve indeksin değerinin iki katsayı arasında yer aldığı gözlenmiştir. Geleneksel sınıflama tutarlığı indeksleri olan κ ve Po ile yapılan karşılaştırmada, P_{CT} indeksinin iki katsayı arasındaki değerlere sahip olduğu bulunmuştur. Analiz sonuçları, P_{CT} indeksine düzeltme yapılması gerektiğini, indeksin hem sınıflama hem de sıralama geçerliği indeksi olarak kullanılabileceğini, ayrıca ölçme aracının güvenilirliğine ilişkin de bilgi verebileceğini ortaya koymuştur. Bu bulgulara rağmen, indeks üzerinde başka çalışmalar yapılması gerektiği de açıktır.

Anahtar kelimeler: P_{CT} indeksi, sınıflama ve sıralama geçerliği, ölçme geçerliği, karar geçerliği

Abstract

Recently, a new validity classification has been proposed by Murphy and Davidshofer (2001) and Erkuş (2002 and 2003): Validity of measurement and validity for decisions. In this study, it has been examined how P_{DC} which is proposed as the index of classification and ordinal validity changes depending on various n, k, and % manipulations and its relation with ρ , split-half coefficient and the classification consistency indices (κ and Po). Analyses on obtained data collected from 161 individuals by using Hopelessness Scale and Submissive Acts Scale have shown that 27%-cutoff area is the best percent; the value of index decreases as n decreases (when k is constant) for dichotomous data, unlike for polytomous data; when n is constant k decreases, P_{DC} value is opposite to n manipulation. When P_{DC} is corrected according to Spearman-Brown formulae (for 2*%46-cutoff area) and is compared with ρ and split-half coefficient, it has been observed that P_{DC} has values between two conventional coefficients. Moreover, when P_{DC} is compared with κ and Po which are the classification consistency indices, the results have pointed out that it has values between κ and Po. As a result, it is suggested that P_{DC} can be used as both classification and ordinal validity index; 46%-Spearman-Brown correction for P_{DC} is appropriate; at the same time, P_{DC} includes information in relation to reliability of the scales. But, it needs further studies especially on its distribution function.

Key words: P_{DC} index, classification and ordinal validity, validity of measurement, validity for decisions

Ölçme araçlarının en önemli psikometrik niteliklerinden biri geçerliktir (validity). Ölçme aracının geliştirilme amacına hizmet etme derecesi olarak nitelenebilecek geçerlik, son yıllarda ölçme aracının “neyi ölçtüğü” (ölçme geçerliği) ve ölçme sonucuna dayanarak “verilecek kararların tutarlığı” (karar/değerlendirme geçerliği) açısından iki ana grupta incelenmektedir (Erkuş, 2002; Erkuş, 2003; Murphy ve Davidshofer, 2001). Örneğin, bir depresyon ölçeği, öncelikle kapsamında ölçmeyi amaçladığı depresyonu ölçmeli (ki, bu “ölçme boyutu”nda amaca hizmet etme derecesini gösterir); ikincisi, güvenilirliği ve geçerliği bilinen yollarla gösterilen bu depresyon ölçeğine dayanarak, bireyler hakkında “depresyonlu-depresyonsuz” şeklinde verilecek kararları da tutarlı bir şekilde ölçebilmelidir (ki, bu da ölçeğin asıl nihai amacı olan “değerlendirme boyutu”nda amacına hizmet etme derecesini gösterir); bu iki amaca yönelik olarak da ölçeğin geçerli olduğuna kanıt toplanmalıdır. Murphy ve Davidshofer’dan (2001) farklı olarak, Erkuş’un (2003) sınıflamasında, bilinen geçerlik belirleme yolları *ölçme geçerliği* altında (yordama geçerliği hem ölçme hem de karar geçerliği altında), ölçüt-dayanaklı değerlendirme bağlamında geliştirilen indeksler (Berk, 1980; Erkuş, 2000a) karar geçerliğinin ilk alt kategorisi olan “*sınıflama geçerliği*” altında, Spearman’ın Sıra Farkları Korelasyon Katsayısı (p) gibi indeksler de, yine karar geçerliğinin ikinci alt sınıfını oluşturan “*sıralama geçerliği*” altında yer almıştır. Ölçme geçerliği boyutundan çok, karar geçerliği boyutunda sınıflamada bir sorun olduğu görünmektedir. Bilindiği gibi, ölçme araçlarına dayanarak verilecek olası kararlar (“assessment” anlamında) seçme, yerleştirme, tanı koyma gibi kararlardır. Ancak, dikkat edilecek olursa, bu kararların “*kurumsal kararlar*” olduğu ve aynı ölçme aracının çeşitli kurumsal amaçlar için kullanılabilmesi görülecektir. Oysa ölçme açısından bakılacak olursa, hangi kurumsal karar olursa olsun, ölçme sonucunda “sınıflama” ve “sıralama” gibi iki “*istatistiksel karar*” verilebilir. Sınıflama tutarlığına ilişkin olarak, bambaşka bir alanda, ölçüt-dayanaklı değerlendirme bağlamında çok sayıda indeks geli-

ştirilmiş olmakla birlikte, bu indekslerin yeri konusunda uzlaşamamıştır; yeni bir geçerlik sınıflaması yapılarak bu indeksler sınıflama geçerliği altında hak ettikleri yere oturtulabilirler. Bu bağlamda, Erkuş (2000b) tarafından hem sınıflama hem de sıralama geçerliği bilgisi verebileceği öne sürülen “Çift Tutarlık İndeksi” (P_{CT}) geliştirilmiştir. P_{CT} homojen ölçek veya alt ölçeklerin tekler-çiftler şeklinde (veya eşdeğer iki yarıyı sağlayacak herhangi bir başka yöntem de olabilir) iki yarıya bölünüp, iki yarıdaki puanların büyükten küçüğe doğru sıralandıktan sonra alt ve üst %27’lik gruplardaki frekansların uyumuna bakmaya dayanmaktadır:

$$P_{CT} = 1 - \left[\frac{(f_{\bar{U}_T} - f_{\bar{U}_C}) + (f_{A_T} - f_{A_C})}{N_{\bar{U}+A}} \right] =$$

$$P_{CT} = 1 - (p_{\bar{U}_T - \bar{U}_C} + p_{A_T - A_C}) \quad \dots (1)$$

Bu çalışmada, P_{CT} indeksinin ikili puanlanan (dichotomously) bir ölçek (Beck Umutsuzluk Ölçeği) ile çoklu puanlanan (polytomously) bir ölçekten (Boyun Eğici Davranışlar Ölçeği) elde edilen verilere dayanarak, iki farklı şekilde puanlanan ölçekler için değişik örneklem büyüklüğü (n), madde sayısı (k) ve yüzdelik dilimlerdeki durumu; “sıralama geçerliği” indeksi olarak kullanıldığında Spearman’ın sıra farkları korelasyon katsayısı ve geleneksel yarıya bölme güvenilirlik katsayısı ile, ayrıca geleneksel sınıflama geçerliği indeksleri (Po ve k) ile karşılaştırmalı olarak ele alınacaktır.

Yöntem

Örneklem

Beck Umutsuzluk Ölçeği (BUÖ) ve Boyun Eğici Davranışlar Ölçeği (BEDÖ)’ni yanıtlayan öğrenci, öğretmen, sağlık çalışanı ve depresif hastadan oluşan 161 kişilik örneklemin yaş ortalaması 30.30 (S=9.18)’tur.

Veri Toplama Araçları

Çalışmada verileri kullanılan BUÖ, Beck, Lester ve Trexler (1974) tarafından geliştirilmiş ve

uyarlaması Seber (1991) ve Durak (1993) tarafından yapılmış, 1-0 şeklinde puanlanan 20 maddeden oluşan ve bireylerin geleceğe yönelik olumsuz beklentilerini ölçmeyi amaçlayan bir ölçektir (Savaşır ve Şahin, 1997). BEDÖ ise, Gilbert ve Allan (1994) tarafından geliştirilmiş ve uyarlaması Şahin ve Şahin (1992) tarafından yapılmış, 1-5 arası puanlanan 16 maddeden oluşan, depresyonla ilgili boyun eğici sosyal davranışları ölçmeyi amaçlayan bir ölçektir (Savaşır ve Şahin, 1997). P_{CT} indeksinin homojen yapıda, güvenilir ve geçerli ölçek verilerine uygulanması gerektiğinden, her iki ölçek için de madde analizi yapılarak ölçekler tekboyutlu duruma getirilmişlerdir (Erkuş, 2004). Erkuş'un çalışmasında, BUÖ'nün bir maddesinin (12. madde) sorunlu olduğu saptanmış ve madde atıldıktan sonra madde-toplam ölçek korelasyonlarının 0.39 ile 1.00 arasında değiştiği, KR-20 içtutarlık katsayısının .95'e çıktığı, ölçegin toplam varyansın %32'sini açıkladığı tek faktörlü olduğu ve umutsuzluk düzeyi düşük ve yüksek %27'lik uç grupları ayırt ettiği (t = 8.153; sd: 86; p<0.000) bulunmuştur. Aynı çalışmada, BEDÖ'nün bir dizi analizden sonra 7 maddeye indirilerek tekboyutlu duruma (açıklanan varyans %38) geldiği, madde-toplam ölçek korelasyonlarının 0.34 ile 0.65 arasında değiştiği, Cronbach Alfa içtutarlık katsayısının 0.72 olduğu ve %27'lik uç grupları ayırt ettiği (t = 8.115; sd: 86; p<0.000) bulunmuştur. Bu çalışmada, BUÖ'nün 19 maddelik ve BEDÖ'nün 7 maddelik homojen duruma getirilen formlarından elde edilen veriler kullanılmıştır.

İşlem ve Veri Analizi

Her iki ölçek için de n (gözlem sayısı) önce 161, sonra puanlar sıralanıp birer atlayarak sistematik bir şekilde 80 kişiye indirilmiş; k (madde sayısı) BUÖ için önce 19 ve sonra 10'a, BEDÖ için önce 7 sonra 4'e birer atlayarak sistematik bir şekilde azaltılmış; indeksin sabit kesme alanlarını oluşturan % grupları da %27, %10 ve %5 olarak alındıktan sonra indeks değerleri hesaplanmıştır. Dolayısıyla, madde sayısı (k) iki, gözlem sayısı (n) iki ve yüzde manipülasyonu üç düzeyde olmak

üzere; her iki ölçek için 24 değişik kombinasyon için P_{CT} değerleri hesaplanmıştır. Geleneksel yarıya bölme katsayıları ise, iki n, iki k düzeyinde olmak üzere iki ölçek için toplam 8 değişik durum için hesaplanmış ve indekslerin birbirlerine ve değişik manipülasyonlara göre değişimi incelenmiştir. Geleneksel sınıflama geçerliği indekslerinden Po ve κ ise; kesme puanının, puan ranjının alt %10'u, alt %27'si ve alt %50'si, üst %27'si ve üst %10'u olmak üzere iki ölçek için 10 kez hesaplanmış ve değişik kesme puanlarına göre gösterdikleri değişim incelenmiştir. Po (Hambleton ve Novick, 1973) ve κ (Swaminathan, Hambleton ve Algina, 1974) Subkoviak'ın (1988) geliştirdiği yaklaştırma formülüyle, geleneksel yarıya bölme katsayıları SPSS paket programıyla, P_{CT} değerleri ise elle hesaplanmıştır.

P_{CT} indeksinin, geleneksel yarıya bölme katsayılarıyla karşılaştırılmasında ortaya çıkacak yanlışlığı gidermek için, indeksin %46'lık diliminin eşdeğer olduğu sayılısıyla Spearman-Brown düzeltmesi aşağıdaki formül yardımıyla yapılmıştır:

$$R_{KK} = 2(0.46/0.54) P_{CT} / 1 + [2(0.46/0.54) - 1] P_{CT} = [(2.3478)P_{CT}] / [1 + (1.3478)P_{CT}] \dots(2)$$

Bulgular

Bu kısımda, öncelikle P_{CT} indeksinin değişik n, madde sayısı ve % gruplarındaki değişimi, sonra geleneksel yarıya bölme katsayıları ve geleneksel sınıflama geçerliği indeksleriyle karşılaştırılması ele alınacaktır.

a) P_{CT} İndeksinin Değişik n, k ve % Gruplarında Aldığı Değerler

P_{CT} indeksinin değişik n, madde sayısı ve % gruplarında aldığı değerler, iki ayrı şekilde puanlanan ölçek için ayrı ayrı Tablo 1 ve Tablo 2'de verilmiştir.

Tablo 1'in incelenmesinden anlaşılacağı gibi, %10 ve %5 kesme alanlarında, ölçek maddelerinin 1-0 şeklinde puanlanmasından dolayı, özellikle alt gruplarda varyans 0.00 olmakta ve bireylerin birbirlerine göre sıralaması da anlamını yitirmektedir.

Tablo 1*İkili Puanlanan Ölçek (BUÖ) Verileri İçin Değişik n, k ve % Gruplarında P_{CT} İndeksinin Aldığı Değerler*

		n = 161				n = 80			
		k = 19		k = 10		k = 19		k = 10	
		Alt	Üst	Alt	Üst	Alt	Üst	Alt	Üst
	n	44	44	44	44	22	22	22	22
%27	\bar{X} Tekler	0.2500	5.4318	0.1136	3.7045	0.5909	5.0909	0.4545	3.5000
	\bar{X} Çiftler	0.3409	5.3182	0.0000	2.0000	0.3636	5.3182	0.0000	2.0000
S^2	Tekler	0.1920	3.7390	0.1030	0.6320	0.2530	4.4680	0.2600	0.4520
	Çiftler	0.2300	3.0130	0.0000	1.5810	0.2420	2.8940	0.0000	1.6190
	P_{CT}	0.6818		0.7159		0.5227		0.6136	
	n	16	16	16	16	8	8	8	8
%10	\bar{X} Tekler	0.0000	7.4375	0.0000	4.5625	0.0000	7.3750	0.0000	4.2500
	\bar{X} Çiftler	0.0000	7.1875	0.0000	3.2500	0.0000	7.1250	0.0000	3.2500
S^2	Tekler	0.0000	2.5290	0.0000	0.2630	0.0000	3.1250	0.0000	0.2140
	Çiftler	0.0000	1.3630	0.0000	1.5333	0.0000	1.8390	0.0000	1.6430
	P_{CT}	0.4375		0.7500		0.3750		0.7500	
	n	8	8	8	8	4	4	4	4
%5	\bar{X} Tekler	0.0000	8.7500	0.0000	5.0000	0.0000	8.7500	0.0000	4.5000
	\bar{X} Çiftler	0.0000	8.1250	0.0000	4.2500	0.0000	8.2500	0.0000	4.2500
S^2	Tekler	0.0000	1.3570	0.0000	0.0000	0.0000	2.2500	0.0000	0.3333
	Çiftler	0.0000	0.6960	0.0000	0.7860	0.0000	0.9170	0.0000	0.9170
	P_{CT}	0.6875		0.6250		0.6250		0.7500	

P_{CT} indeksinin ise, toplam n sayısındaki azalıştan olumsuz yönde etkilenmesine rağmen, her n grubunun kendi içinde madde sayısı azaltıldığında artış eğiliminde olduğu görülmektedir.

İkili puanlanan ölçekteki değişimin tam tersine, %27'lik kesme alanı için çoklu puanlanan ölçekte toplam n azaltıldığında indeks değerinde bu kez artış; her n grubunun kendi içinde madde sayısı azaltıldığında ise indeks değerinde azalış gözlenmektedir. İndeks değerlerindeki bu farklı değişim, ölçeklerin puanlanma biçimleri ve uç grup varyanslarıyla ilişkili görünmektedir. İkili puanlanan ölçek verileri için k azaldıkça indeks değerinin artması, uç gruplardaki toplam puanların sıralamasının kararlı kalmasıyla (ki, varyansların 0.00 olması buna destek vermektedir) açıklanabilir. Çoklu puanlanan öl-

çek verileri için ise, tam tersine madde sayısı azaltıldığında, uç grup toplam puanlarında sıralama duyarlılığı artmakta ve bu da indeks değerinde düşüşe yol açmaktadır. Her iki ölçek için en iyi ayırt edici yüzdelik dilimin ise %27'lik sabit kesme alanları olduğu anlaşılmaktadır ki, bu sonuçlar Fan (1954) ve Brennan'ın (1972) bulgularıyla uyumaktadır.

b) P_{CT} 'nin Geleneksel Yarıya Bölme Katsayısıyla İlişkisi

P_{CT} 'nin geleneksel yarıya bölme katsayısı ile ilişkisi için, ölçeklerin tekler-çiftler arasındaki Pearson Momentler Çarpımı Korelasyon Katsayısı ve bir anlamda sıralama tutarlığı (geçerliliği) için de Spearman'ın Sıra Farkları Korelasyon Katsayıları tüm grup, tüm n ve madde sayıları için hesaplanmış ve katsayılarla, iki kat eşdeğer uzunluk için

Tablo 2Çoklu Puanlanan Ölçek (BEDÖ) Verileri İçin Değişik n, k ve % Gruplarında P_{CT}İndeksinin Aldığı Değerler

		n = 161				n = 80			
		k = 7		k = 4		k = 7		k = 4	
Grup		Alt	Üst	Alt	Üst	Alt	Üst	Alt	Üst
n		44	44	44	44	22	22	22	22
%27	\bar{X} Tekler	5.0682	11.9545	2.0682	6.4545	5.1364	12.2273	2.0000	6.7273
	\bar{X} Çiftler	4.4091	10.2955	2.3864	6.6591	4.5000	10.4545	2.4091	6.6818
S ²	Tekler	0.5770	6.6490	0.0650	2.0680	0.6000	5.4220	0.0000	1.5410
	Çiftler	0.5260	4.6780	0.2430	1.6720	0.4520	5.6880	0.2530	1.3700
P _{CT}		0.5341		0.3977		0.5454		0.4318	
n		16	16	16	16	8	8	8	8
%10	\bar{X} Tekler	4.3125	14.5625	2.0000	8.0000	4.3750	14.5000	2.0000	8.1250
	\bar{X} Çiftler	3.6250	12.6250	2.0000	7.9375	3.7500	13.1250	2.0000	2.0000
S ²	Tekler	0.2290	6.9290	0.0000	1.4670	0.2680	5.7140	0.0000	0.9820
	Çiftler	0.2500	3.4500	0.0000	1.5290	0.2140	2.9820	0.0000	1.9290
P _{CT}		0.3750		0.2500		0.3125		0.1250	
n		8	8	8	8	4	4	4	4
%5	\bar{X} Tekler	4.0000	16.6250	2.0000	8.8750	4.0000	16.2500	2.0000	8.7500
	\bar{X} Çiftler	3.2500	14.2500	2.0000	8.8750	3.5000	14.5000	2.0000	8.5000
S ²	Tekler	0.0000	4.8390	0.0000	0.6960	0.0000	4.9170	0.0000	0.9170
	Çiftler	0.2140	0.7860	0.0000	1.2680	0.3333	1.0000	0.0000	3.0000
P _{CT}		0.2500		0.3125		0.2500		0.1250	

Spearman-Brown düzeltmesi yapılmıştır. P_{CT} indeksi için de, eşitliğin orijinal %27'lik kesme alanları için düzeltme yapılmadan ve Eşitlik 2'de verilen %46'luk orta dilim için eşdeğerlik sayılısıyla Spearman-Brown düzeltmesi yapıldıktan sonra ayrı ayrı hesaplanıp indeksler karşılaştırılmıştır.

Geleneksel yarıya bölme katsayılarının düzeltilmesiz değerleri, P_{CT} indeksinin karşılaştırılabilmesi için verilmiştir; düzeltme yapıldıktan sonraki katsayılar incelendiğinde, P_{CT} indeksi için düzeltme yapılmasının ne kadar yerinde olduğu görülmektedir. Gerek r_{TC}, gerekse ρ_{TC}, madde sayısı azaldıkça düşmekte; P_{CT} indeksi ise, tam tersine artmaktadır. n azaltıldığında ise tüm katsayılarda düşüş gözlenmektedir. Öte yandan, P_{CT} indeksi, madde sayısı azaltılmadığında r_{TC} ile ρ_{TC} arasında değerlere sa-

hip olmaktadır. Bu sonuçlar, Erkuş'un (2000b) önerdiği gibi, P_{CT} indeksinin ölçme aracının güvenirliğini gösterdiğine de kanıt verir niteliktedir.

Tablo 3P_{CT}, r_{TC} ve ρ_{TC}'nin İkili Puanlanan Ölçek (BUÖ) Verileri İçin Karşılaştırılması

		Düzeltilmesiz		Düzeltilmeli	
		k		k	
		19	10	19	10
n = 161	r _{TC}	0.742	0.597	0.852	0.748
	ρ _{TC}	0.609	0.546	0.757	0.706
	P _{CT}	0.682	0.750	0.834	0.876
n = 80	r _{TC}	0.685	0.591	0.813	0.743
	ρ _{TC}	0.525	0.461	0.688	0.631
	P _{CT}	0.523	0.614	0.720	0.788

Tablo 4
P_{ÇT}, r_Ç ve ρ_Ç'nin Çoklu Puanlanan Ölçek (BEDÖ) Verileri İçin Karşılaştırılması

		Düzeltilmesiz	Düzeltilmeli
		k	k
		7	7
n = 161	r _Ç	0.602	0.752
	ρ _Ç	0.597	0.748
	P _{ÇT}	0.534	0.729
n = 80	r _Ç	0.601	0.751
	ρ _Ç	0.564	0.737
	P _{ÇT}	0.545	0.738

Çoklu puanlanan ölçek verileri için tüm indeksler, madde sayısı azaltıldıkça düşmekte; n azaltıldığında ise diğer iki indeks düşerken P_{ÇT} yükselmektedir. Bu durum, P_{ÇT}'nin ikili puanlanan ölçeklerde madde sayısına, çoklu puanlanan ölçeklerde de örneklem büyüklüğüne duyarlı olduğunu göstermektedir. Öte yandan, tüm kombinasyonlarda P_{ÇT}'nin diğer indekslere göre daha yüksek değerler aldığı gözlenmektedir (bkz. Tablo 4).

c) P_{ÇT}'nin Geleneksel Sınıflama Tutarlılığı (geçerliliği) İndeksleriyle Karşılaştırılması

P₀ ve κ sınıflama geçerliliği indeksleri, sadece ikili puanlanan ölçek verilerine uygulandığı için, P_{ÇT} indeksiyle karşılaştırma yapmada BUÖ verileri kullanılmıştır. BUÖ'den alınabilecek toplam puan 0-19 arasındadır; her iki geleneksel sınıflama tutarlılığı indeksi için bu puan ranjının alt %10'u (1.90), alt %27'si (5.13), %50'si (9.50), üst %27'si (13.87) ve üst %10'u (17.10) kesme noktası (cutoff score)

alınarak, Subkoviak (1988) yaklaşımıyla iki indeks değeri hesaplanmıştır. P_{ÇT} indeks değerleri ise daha önce verilmişti; ancak P_{ÇT} kesme noktasını değil de sabit kesme alanlarını dikkate aldığından, puan yerine değişik k ve n manipülasyonlarındaki değerlerinin kullanıldığına dikkat edilmelidir (bkz. Tablo 3).

Şans başarısından arınık olmayan P₀, doğal olarak κ ve P_{ÇT}'ye göre, daha yüksek değerler almaktadır; yine her iki indeks de literatürle uyumlu bir şekilde kesme noktalarından etkilenmektedir. Dikkat edilirse, P_{ÇT} indeksi, her iki indeksin arasında değerlere sahiptir. Madde sayısının azaltılması ile kesme puanının azaltılması aynı anlamı taşımadığı için indekslerin karşılaştırılmasında daha fazla yorum yapılmasını engellemektedir (bkz. Tablo 5).

Tartışma

1.a) P_{ÇT} indeksinde kullanılan %27'lik sabit kesme alanının gerek ikili gerekse çoklu puanlanan ölçekler için en uygun kesme alanı olduğu ileri sürülebilir. İkili puanlanan ölçek verilerinde bu konuda bir sıkıntı var gibi görünmekle birlikte, Tablo 1 daha dikkatli incelendiğinde, yüzde azalışı ile madde sayısı azalışının adeta bir ortak etki yaratarak indeks değerini olduğundan daha fazla büyüttüğü (inflation) görülebilir. Bunun nedeni, ölçek maddelerinin ikili puanlanmasından dolayı, alt ve üst gruplarda alınan toplam puan ranjinin sifira yaklaşmasıdır ki, bu da yanıltıcı bir sonuca yol açmaktadır. Oysa çoklu puanlanan ölçek verileri (üstelik k görece çok az olmasına rağmen) üzerinde, P_{ÇT}, %

Tablo 5
Değişik Kesme Puanlarındaki P₀ ve k Değerlerinin P_{ÇT} Değerleri ile Karşılaştırılması

	C = %10 alt	C = %27 alt	C = %50	C = %27 üst	C = %10 üst
κ	0.69	0.71	0.68	0.58	0.58
P ₀	0.90	0.86	0.91	0.98	0.98
P _{ÇT}	n=161	k=19	0.83		
		k=10	0.88		
	n=80	k=19	0.72		
		k=10	0.79		

ve madde sayısının azalmasına bağlı olarak hemen hemen düzenli bir düşüş göstermektedir. Bilindiği gibi uç gruplar yöntemi, *bir tek maddenin* geçerliği (ayırt ediciliği) için yıllardır güvenle kullanılmaktadır (Tezbaşaran, 1996); bu sonuçlar aynı yöntemin sınıflama geçerliği için de kullanılabilir olduğunu göstermektedir. Fan (1954) ve Brennan'ın (1972), varyansı maksimum yapan uç grubun %27'lik dilim olduğu yönündeki bulgularına bu çalışma destek verir niteliktedir.

b) P_{CT} indeksinin, örneklem büyüklüğünden (n) nasıl etkilendiği, (a)'daki açıklama da dikkate alınarak sadece %27'lik kesme alanları için incelendiğinde, ikili ve çoklu puanlanan ölçeklerde değişimin farklı yönlerde olduğu görülmektedir: İkili puanlanan ölçek verileri için n yarı yarıya azaltıldığında, sabit k düzeyleri için indeks değeri *azalmakta*, çoklu puanlanan ölçek verileri için ise tam tersine *artmaktadır*. Bu durum, indeksin, ikili ve çoklu puanlanan ölçekler için farklı n'lere duyarlı olduğunu göstermektedir. İndeks hesaplamalarında, minimum örneklem büyüklüğünü belirlemek için başka çalışmalar (indeksin dağılım fonksiyonunun çalışmasına paralel olarak) yapılması gerekmesine rağmen, bu çalışma özelinde ikili puanlanan ölçekler için daha büyük örneklem gerektirdiği ileri sürülebilir.

c) n sabit olduğunda, ancak madde sayısı azaltıldığında ise indeks değeri ikili puanlanan ölçekte *artmakta*, çoklu puanlanan ölçekte ise tam tersine *azalmaktadır*. Bu değişim, n manipülasyonunun tam tersidir. Madde sayısı azaltıldığında ikili puanlanan ölçekte indeks değerinin artması, daha önce de değinildiği gibi, uç gruplardaki toplam puan ranjının çoklu puanlanan ölçeğe göre çok küçülmesinden kaynaklanıyor görünmektedir. Ancak, daha doyumlu bir açıklama için madde sayısı eşit olan ikili ve çoklu iki ayrı ölçeğin karşılaştırılması uygun olacaktır.

2. Geleneksel yarıya bölme katsayıları ile P_{CT} arasında, yöntem benzerliği açısından bir ilişkinin olacağı açıktır. Ancak, bu ilişkinin irdelenebilmesi

için, P_{CT} indeksine %46'lık dilim ve iki kat uzunluk için Spearman-Brown düzeltmesinin yapılması gerekmektedir. Ölçeklerin tekboyutlu duruma getirilmesinin avantajı ve söz konusu dilimlerin (ve hatta sıralamaların) eşdeğerliği sayılısıyla Eşitlik 2'de verilen düzeltme formülünün kullanılmasının ne kadar yerinde olduğu Tablo 4 ve Tablo 5'teki verilerin incelenmesinden anlaşılmaktadır. Madde sayısı azaltılmadan P_{CT} değerlerinin r_{TC} ile r_{TC} arasında yer aldığı, madde sayısı azaltıldığında ise diğer iki katsayıdan oldukça yüksek değerler aldığı görülmektedir. Bu durum P_{CT}'nin belirli bir madde sayısı düzeyine kadar güvenilirlik katsayısı olarak yorumlanabileceğine bir kanıt olabilir. Madde sayısı azaltıldığında indeks değerinin diğer katsayılara göre artması ise, P_{CT}'nin aynı zamanda uç grup sınıflamalarını dikkate alan (geçerlik) ve bu nedenle diğer geleneksel katsayıların korelasyonun istatistiksel tanımı gereği varyans arttıkça artmaları, P_{CT}'nin ise tam tersine varyans daraldıkça artmasıyla açıklanabilir. ρ sıra farklarını dikkate aldığı için, yeni geçerlik sınıflamasında karar geçerliği altında yer alan sıralama geçerliği indeksi olarak önerilmişti. ρ için iki uygulamadaki sıra farkları kullanılabilirdiği gibi, bu çalışmada olduğu gibi, tek uygulama fakat iki eşdeğer yarı arasındaki düzeltilmiş sıra farkları da kullanılabilir. Bu şekilde hesaplanan r'nun, sıralama geçerliği indeksi olarak kabul edilebileceğini elde edilen değerler göstermektedir.

3. P_{CT}'nin geleneksel sınıflama tutarlığı (geçerliği) indeksleriyle karşılaştırılması, onun aynı zamanda sınıflama geçerliği bilgisi taşıyıp taşımadığının test edilmesini içermektedir. Literatürde, ölçüt-dayanaklı değerlendirme altında, 1963-1994 yılları arasında geliştirilmiş olan bir çok indeksin nasıl adlandırılacağı ve yerinin ne olacağı konusunda hala bir görüş birliği bulunmamaktadır. Bu durum, ölçme ile değerlendirme, güvenilirlik ile geçerlik kavramlarının birbiri yerine kullanılmasına yol açmaktadır. Yeni bir geçerlik sınıflaması, bu indekslerin karar (ya da değerlendirme) geçerliği altında ve "sınıflama geçerliği" olarak yer almasını

önermektedir. Gerçekten de, ölçme araçlarından elde edilen verilere dayanarak en temel ve yaygın olarak (istatistiksel boyutta) sınıflama kararları verilmektedir: Depresyonlu-depresyonsuz, geçer-kalır, işe kabul edilir-edilmez gibi. Bireyler hakkında bu kadar ciddi bir karar olan sınıflamanın ne kadar tutarlı yapıldığı, en az kararın kendi kadar ciddidir ve bu indeksler bu bilgiyi (bazı zaaflı yanlarına rağmen/bkz: Erkuş, 2000a) vermektedirler. P_{CT} indeksini bu indekslerden ayıran en önemli yanı, keyfi (arbitrary) bir kesme noktasını değil, sabit ve kullanılabilirliği yıllar içinde kanıtlanmış %27'lik kesme alanlarını kullanmasıdır. Bu açıdan P_{CT} indeksinin bu indekslerle karşılaştırılması son derece önemlidir. Ancak, söz konusu indekslerin "oynak" kesme puanlarına dayanması, P_{CT} indeksinin sabit kesme alanlarına dayanmasından (bir başka deyişle, yöntemlerinden kaynaklanan farklılıktan) dolayı, karşılaştırma sonuçları üzerinde belirli bir yere kadar yorum yapılabilmesine olanak sağlamaktadır. Tablo 5'teki bulgular incelendiğinde, Po ve k'nın kısmen literatürle uyumlu bir şekilde, kesme noktasının pozisyonuna göre değişiklik gösterdiği gözlenmektedir. Düzeltilmiş P_{CT} değerlerinin ise, Po ile κ arasında yer alması dikkat çekicidir. P_{CT} indeksinin, bu iki indeks ve diğer sınıflama tutarlılığı indeksleriyle daha sağlıklı bir karşılaştırması, indeksin dağılım fonksiyonunun çalışılmasıyla olabilir. Ancak, bu durumda bile P_{CT} indeksinin, özellikle kullandığı yöntemden dolayı, bireyleri tutarlı sınıflamada kullanılabileceği ileri sürülebilir.

Sonuç olarak, bir ölçme aracının, bilinen geçerlik yollarıyla geçerli olduğunun (ölçmeyi amaçladığı şeyi ölçüp ölçmediğinin) gösterilmesi, onun ciddi kararlarda kullanılması için yeterli değildir; psikolojik ölçme araçları bireyler hakkında ciddi kararlar vermek için kullanılırlar ve bu kararların da geçerli olup olmadığı araştırılmak ve gösterilmek durumundadır. Bu açıdan, yeni geçerlik sınıflaması ve onun karar geçerliğine kanıt bulma çabasının bir sonucu olan P_{CT} indeksinin, başta dağılım fonksiyonu olmak üzere diğer özelliklerinin araştırılması dikkate değer görünmektedir.

Kaynaklar

- Beck, A. T., Lester, D. & Trexler, L. (1974). The Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, 42, 861-874.
- Berk, R. A. (1980). A consumers' guide to criterion-referenced test reliability. *Journal of Educational Measurement*, 17(4), 323-349.
- Brennan, R. L. (1972). A generalized upper-lower item discrimination index. *Educational and Psychological Measurement*, 32, 289-303.
- Durak, A. (1993). *Beck Umutsuzluk Ölçeği'nin geçerliği üzerine bir çalışma*. Yayınlanmamış Yüksek Lisans Tezi. Ankara Üniversitesi, Sosyal Bilimler Enstitüsü.
- Erkuş, A. (1999). *Ölçme araçlarının tutarlı ölçme ve sınıflama yapıp yapmadığını belirlemeye ilişkin bir çalışma*. Yayınlanmamış Doktora Tezi. Ankara Üniversitesi SBE.
- Erkuş, A. (2000a). Ölçüt-dayanaklı değerlendirme bağlamında geliştirilen sınıflama tutarlılığı indeksleri ve bazı sorunlar. *Türk Psikoloji Yazıları*, 3(5), 1-21.
- Erkuş, A. (2000b). Yeni bir indeks önerisi: Çift tutarlılık indeksi (P_{CT}). *Türk Psikoloji Dergisi*, 15(46), 63-71.
- Erkuş, A. (2002). Yeni bir geçerlik kavramlaştırması ve sınıflaması önerisi. *XII. Psikoloji Kongresi* 13 Eylül 2002, ODTÜ, KKM.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*. Türk Psikologlar Derneği Yayınları, No: 24
- Erkuş, A. (2004 basımında). Beck Umutsuzluk Ölçeği ve Boyun Eğici Davranışlar Ölçeği'nin psikometrik niteliklerinin gözden geçirilmesi. *Mersin Üniversitesi Tıp Fakültesi Dergisi*.
- Fan, C. T. (1954). Note on construction of an item analysis table for the high-low-27-percent group method. *Psychometrika*, 19(3), 231-237.
- Gilbert, P. & Allan, S. (1994). Assertiveness, submissive behavior and social comparison. *British Journal of Clinical Psychology*, 33, 295-306.
- Hambleton, R. K. & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 10(3), 159-170.
- Murphy, K. R. & Davidshofer, C. O. (2001). *Psychological testing: Principles and applications (5th ed.)*. New Jersey: Prentice Hall.
- Savaşır, I. & Şahin, N. H. (1997). *Bilişsel-davranışçı terapilerde değerlendirme: Sık kullanılan ölçekler*. Türk Psikologlar Derneği Yayınları, No: 9.

- Seber, G. (1991). *Beck Umutsuzluk Ölçeği'nin geçerlik ve güvenilirliği üzerine bir çalışma*. Yayınlanmamış Doçentlik Tezi. Anadolu Üniversitesi Tıp Fakültesi Psikiyatri Bölümü, Eskişehir.
- Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25(1), 47-55.
- Swaminathan, H., Hambleton, R. K. & Algina, J. (1974). Reliability of criterion-referenced tests: A decision-theoretic formulation. *Journal of Educational Measurement*, 11, 263-267.
- Şahin, N. H. & Şahin, N. (1992). Adolescent guilt, shame, and depression in relation to sociotropy and autonomy. *The World Congress of Cognitive Therapy*, Toronto, June 17-21.
- Tezbaşaran, A. A. (1996). *Likert tipi ölçek geliştirme klavuzu*. Ankara: Türk Psikologlar Derneği yayınları.

Summary

Change of P_{DC} in Various n , k and Percent Manipulations and Its Comparison with ρ , Split-Half Coefficient and Various Indices

Adnan Erkuş
Mersin Üniversitesi

Murphy and Davidshofer (2001) and Erkuş (2002 and 2003) proposed a new validity classification named as validity of measurement and validity for decisions. In a different way from Murphy and Davidshofer (2001), Erkuş has suggested that validity for decisions includes the classification validity and the ordinal validity. P_{DC} proposed by Erkuş (2000) carries information of both the classification validity (P_o , κ , etc.) and the ordinal validity (Spearman's rank order correlation with repeated applications or adjusted two half correlation):

$$P_{DC} = 1 - \left[\frac{(f_{U_o} - f_{U_e}) + (f_{L_o} - f_{L_e})}{N_{U+L}} \right] =$$
$$P_{DC} = 1 - (P_{U_o-U_e} + P_{L_o-L_e}) \dots (1)$$

In this study, it has been examined how P_{DC} changes depending on various n , k and % manipulations and its relation with ρ , split-half coefficient and the classification consistency indices (P_o and κ) based on data of two scales which are dichotomously and polytomously scored. Moreover, Spearman-Brown correction has been examined for 46% of P_{DC} :

$$R_{KK} = 2(0.46/0.54) P_{CT} / [1 + 2(0.46/0.54) - 1] P_{CT} = [(2.3478)P_{CT}] / [1 + (1.3478)P_{CT}] \dots (2)$$

Method

Analyses have been realized on data collected by using Hopelessness Scale (HS) and Submissive Acts Scale (SAS) from 161 participants being medical workers, students, teachers, and depressive patients (Age mean = 30.03, $S = 9.18$).

27, 10, and 5 percentage groups have been manipulated for P_{DC} . Moreover, the sample size has been manipulated systematically by omitting every other individual; the items have been manipulated from 19 to 10 for HS, from 7 to 4 for SAS systematically by omitting every other item. Finally, P_{DC} was evaluated for 24 combinations ($k = 2$, $n = 2$, percentage = 3; for two scales). ρ and split-half coefficient were evaluated for 8 combinations ($n = 2$, $k = 2$; for two scales). P_o and κ were evaluated for 10 combinations (cutoff point = 10%, low, 27% low, 50%, 27% up and 10% up for two scales). P_o and k were calculated with Subkoviak's (1988) approximation. Spearman-Brown correction was applied for 46% of P_{DC} (Equation 2).

Results

a) Change of P_{DC} as various n , k and % manipulations

For dichotomously scored scale (HS), P_{DC} decreases as n decreases for constant k levels in

*Address for Correspondence: Doç. Dr. Adnan Erkuş, Mersin Üniversitesi, Eğitim Fakültesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı Yenişehir Yerleşkesi 33169 Mersin, Turkey.
E-mail: adnanerkus@yahoo.com

every percentage; but, P_{DC} increases as k decreases for constant n levels in every percentage (out of 5%); P_{DC} decreases as the percentage decreases for constant n and k levels (out of 5%).

For polytomously scored scale (SAS), in contrast to dichotomously scored scale, P_{DC} increases as n decreases for constant k levels in percentage 27; but, P_{DC} decreases as k decreases for constant n levels in every percentage. P_{DC} decreases as the percentage decreases for all constant n and k levels.

For each two scales, PDC has bigger values in 27% groups.

b) Relation of P_{DC} with split-half coefficient and Spearman's rank order correlation

Spearman's rank order correlation coefficient applied to two equivalent halves has been corrected with Spearman-Brown formulae for two times, and it has been used as the evidence of "the ordinal validity" (ρ_{ov}). Moreover, PDC has been corrected for 46% for two times as used Equation 2 and has been compared to the conventional split-half coefficient (ρ_{oe}) and Spearman's rank order correlation as the ordinal validity.

For HS, both r_{oe} and r_{ov} decrease as k decreases, but P_{DC} increases. On the other hand, all of the coefficients have decreased when n has decreased. P_{DC} has a value between ρ_{oe} and ρ_{ov} for original k ; but, it has bigger values as k decreases for each of two n groups.

For SAS, all of the coefficients have decreased as k decreases, but P_{DC} has increased as the other indices decrease when n has decreased.

c) Comparison of P_{DC} to P_o and κ

P_{DC} has been compared to P_o and k for only HS (dichotomic). The indices have been ordered as $P_o > P_{DC} > k$.

Conclusion

It has been found that 27% was more appropriate cutoff area than the others for P_{DC} . This result is consistent with Fan (1954) and Brennan (1972). But, P_{DC} has a different sensitiveness for dichotomously and polytomously scored scales in k and n manipulations. Increase in P_{DC} depending on decrements in the percentage and k draws attention. So, it can be suggested that the variance restricts in small k and % on the low groups when the scale is scored dichotomously. In this case, the individuals have been sequenced more consistently in these groups and so it may be suggested that the value of P_{DC} has been inflated by interacting decrease of % and k . On the other hand, P_{DC} has differently moved on n manipulations for dichotomously and polytomously scored scales: It decreases for the first one, as it increases for the second one. In order to find a satisfactory minimum n number, the distribution function of P_{DC} must be studied.

Having the values of P_{DC} between ρ_{oe} and ρ_{ov} can be interpreted as P_{DC} having reliability information of the scales. On the other hand, increase of P_{DC} , as the variance restricts, has indicated that it has information of the classification consistency, by noticing also P_{DC} 's relation to P_o and κ . Moreover, it can be suggested that Spearman's rank order correlation (by corrected) can be used as the ordinal validity to the results.

Finally, an instrument can have validity of measurement, but it isn't enough; at the same time, it must have validity for decisions: Does it classify and/or order the individuals? P_{DC} seems as an appropriate index to reach this aim. But the distribution of P_{DC} needs to be investigated and further research needs to be made.