

GÜVENİRLİĞİN DEĞERLENDİRİLMESİNDE GENELLENEBİLİRLİK KURAMININ KULLANILMASI: ENDÜSTRİ VE ÖRGÜT PSİKOLOJİSİNDE BİR UYGULAMA

Aiilla YELBOĞA

Ö Z

Bu araştırmada endüstri psikolojisinde yapılan bir ölçme işleminde genellenebilirlik kuramının kullanılabilirliği değerlendirilmiştir. Genellenebilirlik Kuramı kullanılarak, potansiyel hata kaynakları tanımlanabilir ve değerlendirilebilir. Genellenebilirlik kuramı, bütün hata kaynaklarını birlikte değerlendiren ve ölçmenin güvenirliliğini belirlenmesini sağlayan güçlü bir araçtır. G kuramında, Genellenebilirlik (G) ve Karar (K) çalışmaları olarak iki tür çalışma bulunmaktadır. G çalışmaları, ölçme hatalarını belirli değişkenlik kaynaklarına ayırmak için düzenlenirken, K çalışmasında, G çalışmasından elde edilen bilgileri kullanarak belli bir amaçla yapılan ölçmedeki hataları en aza indirmenin yolları aranır. Bu çalışmada farklı ve çok kaynaklı bir ölçme durumunda, G ve Phi katsayılarının elde edilmesi gösterilmiş, G katsayısı 0,867 ve Phi katsayısı 0,793 olarak bulunmuştur.

Anahtar Kelimeler: Genellenebilirlik Kuramı, Güvenirlilik, Genellenebilirlik

The Assessment of Reliability With Generalizability Theory: An Application in Industrial and Organizational Psychology

ABSTRACT

The objective of this study was to examine the potential use of the generalizability theory for assessing reliability of measurement in industry psychology. Using generalizability theory, potential sources of measurement error can be recognized and estimated. Generalizability theory is a powerful tool for estimating the magnitude of multiple sources of measurement error and for assessing the reliability of measurements. Generalizability theory considers two types of studies: generalizability (G) studies and decision (D) studies. G studies are desing to estimate as many sources of error as possible, while D studies obtain measurements for a particular purpose. The use of generalizability theory by obtaining G coefficients for norm referenced assessment and Phi coefficients for

criterion referenced assessment has been demonstrated through an application in this study, as a measurement case possessing various and multiple sources of error. In this application, the G coefficient for measurement is 0,867 and the Phi coefficient 0,793.

Key Words: *Generalizability Theory, Reliability, Generalizability*

Toplumdaki Ölçme çalışmalarında amaç, ölçülen özellik bakımından elde edilen puanın olabildiğince gerçek puana yakın olmasının sağlanmasıdır (Baykul, 2000). Gerçek puanlara yakın ölçme sonuçları ise ölçme ile elde edilen puanlara karışan hatanın az olması ya da hatasızlığı ile olanaklıdır ki bu da güvenilirlik olarak adlandırılır. Çoğu zaman ölçme sonuçlarına dayalı olarak önemli kararların verilmesi söz konusudur. Kararların isabetliliği yapılan işlemin karar vermek için uygun olup olmadığı ya da ölçme sonuçlarına karışan hatalardan arındırılarak bu sonuçların kullanılması ile olanaklıdır. Bu nedenle de ölçme işlemi ve sonuçlarına karışan hatanın araştırılması, yani güvenilirliğinin test edilmesi gereklidir.

Güvenirlik tanımı, klasik test kuramının da etkisiyle genellikle, paralel testler arası korelasyon olarak tanımlanmaktadır (Crocker ve Algina, 1986; Gulliksen, 1967; Lord ve Novick, 1968). Anastasi (1988) ise güvenilirliği aynı testi farklı zamanlarda alan bireylerin puanlarının tutarlılığı şeklinde tanımlamıştır. Güvenirliği, bir bireyin kazandığı puanların kendi kendisiyle tutarlılığı şeklinde tanımlayanlar da vardır (Ghiselli, Campbell ve Zedeck, 1981). Kısaca ölçme aracı neyi ölçüyorsa onu kararlı bir şekilde ölçmelidir, buna ölçme aracının güvenilirliği denir (Özçelik, 1989).

Fiziksel ölçmelerden farklı olarak, psikolojik ölçme işlemlerinde, genellikle bireyin tepkileri gözleendiğinden ölçme sonuçlarına mutlaka hata karışır. Puanlarda değişime neden olan tüm etkenler tam olarak açıklanamaz ve tanımlanamaz, zaten açıklanabilseydi güvenilirlik kavramına gerek olmazdı (Ghiselli, Campbell ve Zedeck, 1981).

Klasik test kuramında güvenilirlik aşağıdaki gibi formüle edilmiştir (Urbina, 2004).

$$X_{gözlenen} = X_{gerçek} + X_{hata} \quad (1)$$

Yukarıda verilen formülde, hata puanları azaldığı oranda gözlenen puanın gerçek puana yaklaşacağı görülmektedir. Ölçmelere karışan hatalar nedeniyle güvenilirliğin tam olarak bilinemeyeceği, bazı yöntemler kullanılarak tahmin

edilebileceği ifade edilebilir. Ölçme sonuçlarının ne derece güvenilir olduğunu “Güvenirlilik katsayısı” belirlemektedir. Güvenirlilik katsayısı, gerçek puanlar varyansının gözlenen puanlar varyansına oranı olarak açıklanır (Crocker ve Algina, 1986). Güvenirlilik tahmininde izlenen yöntem ne olursa olsun, güvenirlilik tahmini sonucunda 0,00 ile 1,00 arasında bir korelasyon elde edilir. Korelasyonun 1,00’e yakın olması ölçme aracının güvenirliliğinin yüksek olduğu, 0,00’a yakın olması da ölçme aracının güvenirliliğinin düşük olduğu anlamına gelir. Ölçme aracının güvenirliliğinin yüksek olması puanlara karışan hatanın az olduğu, güvenirliliğin düşük olması da puanlara karışan hatanın fazla olduğunu gösterir (Özçelik, 1989).

Ölçmelerde hata, ölçülmek istenmeyen miktarın ölçme sonuçlarına karışmasıdır. Hatalar; sabit, sistematik ve rastgele (seçkisiz) hatalar olarak incelenebilir (Crocker ve Allgina, 1986; Cronbach, 1984; Guilford, 1954; Gulhksen, 1967; Lord ve Novick, 1968).

Güvenirlilik tahmin yöntemleri, ilgili oldukları hata kaynaklarına göre farklı adlar alır ve farklı anlamlar taşır. Bu nedenle de güvenirliliğin çeşitli türleri vardır. Genel olarak ölçmede güvenirlilik kavramı; ölçme araç ya da araçlarından elde edilen puanların tutarlılığı olarak tanımlanır. Bir ölçme aracının iki ya da daha çok uygulamasından (test-tekrar test) elde edilen sonuçların tutarlılığı kararlılık olarak adlandırılır ve hata kaynağı iki uygulama arasındaki zaman olarak değerlendirilir. Eşdeğer formlar veya paralel formlar güvenirliliği aynı ölçme için iki paralel formdan elde edilen puanların tutarlılığıdır ve hata kaynağı olarak ölçme aracı dikkate alınır. İç tutarlılık anlamındaki güvenirlilik, ölçme aracında yer alan madde puanlarının bir biri ile tutarlılığıdır ve hata kaynağı olarak ölçme aracında yer alan maddeler dikkate alınır. Değerlendiriciler arası güvenirlilik ise birden çok değerlendiricinin ölçülen aynı nitelik için verdikleri puanların tutarlılığı olarak tanımlanır ve hata kaynağı değerlendiriciler olarak değerlendirilir (Yelboğa, 2007).

Değerlendiriciler (puanlayıcıların, yargılayıcıların) arası uyum ya da güvenirliliğin belirlenmesi gereken birçok ölçme durumu vardır. Aslında çok sayıda değerlendirici (puanlayıcı, yargılayıcı) tarafından yapılan puanlamalar ya da değerlendirmeler özellikle psikoloji ve eğitim olmak üzere sosyal bilimlerin birçok alanında kullanılmakla birlikte, sosyal bilimlerin dışında da yaygın bir kullanım alanına sahiptir.

Değerlendiriciler (puanlayıcıların, yargılayıcıların) arası güvenirliliğin belirlenmesinde kullanılabilen geleneksel istatistiksel tekniklerden; basit yüzde,

Pearson Çarpım-Moment Korelasyon katsayısı, bazı durumlarda ortalamaların karşılaştırılması (t testi veya varyans analizi) gibi teknikler sayılabilir (Goodwin, 2001).

Basit yüzde, iki değerlendiricinin performans, davranış vb. birçok durumda verdikleri puanların ya da aynı değerlendiricinin aynı davranışı iki kez puanlaması durumunda değerlendiricilerin verdikleri puanların uyumunun basit yüzdesi, tutarlılık anlamında kullanılabilir. Basit yüzde tekniğinin en belirgin avantajı hesaplanması ve anlaşılmasının oldukça kolay olması yanında sınıflama, sıralama, aralık ve oran ölçeği değerindeki tüm veriler üzerinde kullanılabilir olmasıdır (Goodwin, 2001). Madde veya görev için uyum yüzdesinin hesaplanması değerlendiriciler arasındaki uyumsuzluk noktalarının tam olarak belirlenmesini sağlayabilir. Böyle bir belirleme değerlendiricilerin puanlama davranışlarına ilişkin olarak eğitilmesi ya da tekrar eğitilmesi için kullanılabilir olması nedeniyle önemli olabilir.

Basit korelasyon ise, bir grup bireye ya da objeye ilişkin olarak iki değerlendiricinin verdikleri puan dağılımları arasındaki ilişkidir. Bu ilişki, değerlendiricilerin verdikleri puanların bir biri ile tutarlılığının bir ölçüsü olarak hesaplanabilir. Bu tür bir ilişki hesaplaması ve güvenilirlik katsayısı olarak değerlendirilmesi klasik test kuramının test – tekrar test yöntemiyle güvenilirlik hesaplamasına benzer. İki değişken dağılımı arasındaki ilişkinin hesaplanmasında kullanılan bir çok korelasyon tekniği vardır. Ancak bunlar arasında en çok bilinen ve kullanılan korelasyon tekniği, aralık ve oran ölçeğindeki değişkenlere uygulanabilen Pearson Momentler Çarpımı Korelasyon katsayısıdır. Değerlendiriciler arası güvenilirliğin belirlenmesi için Pearson Momentler Çarpımı Korelasyon katsayısının hesaplanması, iki değerlendiricinin bireyler, objelere vs. ölçülen özellik boyutunda verdikleri puanların tutarlılığı olarak tanımlanır. Güvenirlik hesaplamada, güvenilirlik katsayısı olarak kullanılan bu korelasyon katsayısı iki değerlendiricinin puanlarının doğrusal ilişkisini gösterir (Baykul, 1999). Pearson Momentler Çarpımı Korelasyon katsayısının güvenilirlik kestirmek amacıyla kullanılması durumunda, değerlendiricilerin verdikleri puanların ortalamalarının eşleştirilmiş gruplar için t testi ya da tekrarlı ölçümler için ANOVA kullanılarak karşılaştırılması ve hesaplanan korelasyon katsayısıyla birlikte yorumlanması daha uygun olabilir.

Pearson Momentler Çarpımı Korelasyon katsayısının, değerlendiricilerin arasındaki güvenilirliğin belirlenmesindeki dezavantajını ortadan kaldırmak için

uygulanabilecek bir diğer yol varyans analizine dayalı olarak hesaplanan gruplar-İçi korelasyonun hesaplanmasıdır. Varyans analizi yoluyla hesaplanabilen bu güvenilirlik katsayısı “sıfır” ile “bir” arasında değerler alır. Bu katsayı varyans analizi literatüründe “*intraclass correlation coefficient*” olarak ifade edilen grup-İçi korelasyon katsayısıdır. (Baykul, 1999). Bu korelasyon katsayısı bir tek hata kaynağını dikkate alarak hesaplanma yapılmasına olanak sağlar. Oysa daha önce de değinildiği gibi birçok ölçme durumunda birden çok değişkenlik kaynağından gelen hatalar ölçme sonuçlarına karışabilmektedir. Grup-İçi korelasyon katsayısından hareketle birçok değişkenlik kaynağından gelen hataların birlikte değerlendirilmesini sağlayacak biçimde geliştirilmesi yoluyla güvenilirlik hesaplanmasına ilişkin bir teknik olan Genellenebilirlik Kuramı (G, Generalizability Theory) ilk defa Cronbach, Rajaratnam ve Gleser (1963) tarafından ortaya atılmıştır. G Kuramı davranış ölçümlerinde güvenilirliğin değerlendirilmesini, güvenilir gözlemlerin tasarımı, araştırılmasını ve kavramlaştırılmasını sağlayan istatistiksel bir kuramdır (Brennan, 2001; Cronbach, Rajaratnam ve Gleser 1963; Cronbach, Gleser, Nanda, ve Rajaratnam, 1972; Shavelson ve Webb, 1991).

G kuramında değerlendirici, madde, ölçme aracı, zaman gibi ölçmeye ilişkin her bir özellik (facet) olarak tanımlanmaktadır (Brennan, 2001). G kuramı; maddeleri, zamanı, değerlendiricileri ve benzeri hata kaynaklarını, değişkenlik kaynağı olarak adlandırır. Bu değişkenlik kaynaklarının düzeyleri de koşullar veya seviyeler (levels) olarak adlandırılırlar. Değişkenlik kaynakları ve koşul veya seviye ifadeleri varyans analizi literatüründeki faktör ve düzey kavramlarına karşılık gelmektedir (Brennan, 2001; Crocker ve Algina, 1986, Shavelson ve Webb, 1991). Facet’ler faktör veya değişkenlik kaynağı olarak da adlandırılabilir. G kuramında kullanılan bir kavram olan “değişkenlik kaynağı” (facet) deneysel desenlerin literatüründeki faktör kavramına benzer. Gözlemlerin evreni, değişkenlik kaynaklarının bütün olası kombinasyon düzeyleri olarak tanımlanır. Genellenebilirlik kuramında, ölçme hatalarının farklı değişkenlik kaynaklarından kestirilmesi, gruplararası korelasyon katsayısının hesaplanmasında olduğu gibi varyans analizi yaklaşımının genişletilmiş bir halidir. G kuramının amacı, ölçme konusu olan bireyler ya da objelerin gözlenen puanlarının evren puanlarına (Klasik test kuramındaki gerçek puan kavramına karşılık gelir) doğru olarak genellenmesini sağlamaktır (Atılgan, 2004; Atılgan ve Tezbaşaran, 2005).

G kuramında, genellenebilirlik (G, generalizability) katsayısı olarak adlandırılan bir katsayı hesaplanır. Bu katsayı Klasik test kuramındaki güvenilirlik katsayısına benzetmekle birlikte güvenilirlik kavramını yeniden yorumlamaz. G kuramı ayrıca güvenilirlik ve geçerlik arasındaki geleneksel farkın güvenilir gözlemler düzenlemekle nasıl ortadan kaldırılabileceğini de gösterir. Geçerlilik türleri içinde yapı geçerliği, gözlenebilen ölçmeler üzerine temellenmiş örtük bir yapı hakkında vardama yapılmasıyla ilgilidir. G kuramında bir evren, onun değişkenlik kaynakları ve gözlemlerin koşulları geçerlik kuramının geleneksel alanında açıklanan yapıdan tanımlanır. G kuramı gözlemlerin örtülü yapısını (yani kabul edilebilir gözlemlerin evreni) hakkındaki vardamaları doğrulukla göstermeyi sağlarsa, gözlemleri güvenilir olarak tanımlar (Brennan, 2001; Crocker ve Algina, 1986; Shavelson ve Webb, 1991).

G kuramında, Genellenebilirlik (G, Generalizability) ve Karar (K, Decision) çalışmaları olarak iki tür çalışma bulunmaktadır.

G çalışmaları, ölçme hatalarını belirli değişkenlik kaynaklarına ayırmak için düzenlenmektedir. G çalışması düzenlemede öncelikle potansiyel değişkenlik kaynaklarının ve ölçme örnekleminin genelleneceği evrenin belirlenmesi gereklidir. G çalışmalarında potansiyel hata kaynaklarına bağlı olarak farklı desenler bulunmaktadır. Bu desenlerde değişkenlik kaynaklarının seçilmesinde tesadüfi ve sabit etki, verilerin düzenlenmesinde ise çapraz ve yuvalanmış olması durumu söz konusudur. Değişkenlik kaynaklarının özelliklerine ve sayısına bağlı olarak farklı varyans analizi modeli bulunmaktadır. G kuramının desenleri de bu varyans modellere bağlı olarak belirlenir (Brennan, 2001; Crocker ve Algina, 1986; Shavelson ve Webb, 1991).

G kuramındaki değişkenlik kaynakları içerisinde bulunan tesadüfi değişkenlik, ölçme işlemlerinin koşullarının tesadüfi olarak örneklenmesi demektir. Tesadüfi değişkenlik kısaca değerlendiricilerin, puanlama yapması gereken bir ölçme durumunda, geniş bir değerlendirici evreninden tesadüfi olarak çekilen bir değerlendirici örnekleminin puanlama yapmasıdır. Bunun yanında genellenebilmenin evreni sonsuz büyüklükte ve değiştirilebilir özellikte olduğu varsayıldığında, ölçme işlemlerinin koşulları tesadüfi olarak çekilmemişse bile değişkenlik kaynağı tesadüfi olarak kabul edilir. Örneklemin evrenden küçük olduğu ve örneklemin evrenden tesadüfi olarak çekildiği ya da değiştirilebilir olduğu durumlarda değişkenlik kaynağı tesadüfi değişkenlik kaynağı olarak adlandırılır (Brennan, 2001; Shavelson ve Webb, 1991).

Diğer bir değişkenlik kaynağı da sabit değişkenlik kaynağıdır. Sabit değişkenlik kaynağı özellikle seçilmiş olan ve bunun dışında genellenmesiyle ilgilenilmeyen veya çalışılan evrenin küçük olması nedeniyle evren üzerinde çalışılması durumlarında kullanılır (Brennan, 2001; Crocker ve Algina, 1986; Shavelson ve Webb, 1991).

Verilerin çaprazlanmış ya da yuvalanmış olmaları da G kuramındaki desenlerin belirlenmesinde göz önünde tutulmaktadır. Verilerin çaprazlanmış olması, bir değişkenlik kaynağının bütün koşullarının diğer bir değişkenlik kaynağının tüm koşullarında gözlenmesini ifade eder ve iki kaynak arasında " x " simgesi konularak gösterilir. Verilerin yuvalanmış olması ise, bir değişkenlik kaynağının bütün koşulları diğer bir değişkenlik kaynağının sadece bazı koşullarınca gözlemlenmesidir ve iki kaynak arasında ":" işareti konularak gösterilir (Brennan, 2001; Shavelson ve Webb, 1991). Bu açıklamalara örnek olarak, bireylerin performanslarının değerlendiriciler tarafından puanlanması söz konusu olduğunda, her değerlendirici her bireyi gözlemleyip puanlıyorsa diğer bir deyişle verilerin çaprazlanmış olması durumunda, birey b, madde m ve değerlendirici d ile gösterildiğinde, verilerin çaprazlanmış olması durumu simgesel olarak $b \times m \times d$ şeklinde ifade edilir. Benzer şekilde bireylerin her birinin performansı farklı değerlendiriciler tarafından gözlemleniyorsa $m \times (b:d)$ ile şeklinde ifade edilir (Atılğan, 2004).

G kuramının $b \times m \times d$ deseninde tanımlanmış geniş bir evrende yer alan bütün maddeleri ve değerlendiricilerin genellenmesinde, belli bir bireyin (b), belli bir maddeden (m) ve değerlendiriciden (d) elde edilen puanları;

$$\begin{aligned}
 X_{bmd} = & \\
 \mu & \text{ Genel ortalama} \\
 \mu_b - \mu & \text{ Birey etkisi} \\
 + \mu_m - \mu & \text{ Madde etkisi} \\
 + \mu_d - \mu & \text{ Değerlendirici etkisi (2)} \\
 + \mu_{bm} - \mu_b - \mu_m + \mu & \text{ Birey x madde etkisi} \\
 + \mu_{bd} - \mu_b - \mu_d + \mu & \text{ Birey x değerlendirici etkisi} \\
 + \mu_{md} - \mu_m - \mu_d + \mu & \text{ Madde x değerlendirici etkisi} \\
 + \mu_{bmd} - \mu_b - \mu_m - \mu_d + \mu_{bm} + \mu_{bd} + \mu_{md} - \mu & \text{ Kalan}
 \end{aligned}$$

eşitliği ile yazılır. Bu eşitlikte E (expected value) sembolü beklenen değer olarak ifade edildiğinde $\mu = E_d E_m E_b X_{bmd}$ ve $\mu_b = E_d E_m X_{bmd}$ şeklinde yazılır. Tesadüfi etki modelinde; genel ortalamanın beklenen değeri, her bir etkinin dağılımının ortalamasının sıfır ve varyans bileşeni olarak adlandırılan bir varyansının olduğu varsayılır (Brennan, 2001; Shavelson ve Webb, 1991).

Gözlenen puanların varyansı;

$$\sigma^2(X_{bmd}) = \sigma_b^2 + \sigma_m^2 + \sigma_d^2 + \sigma_{bm}^2 + \sigma_{bd}^2 + \sigma_{md}^2 + \sigma_{bmd,e}^2 \quad (3)$$

eşitliği yukarıdaki etkilerin varyanslarının toplamı olarak tanımlanır. (3) numaralı eşitlikte yer alan her bir varyans bileşeni, varyans analizi kullanılarak çizelge 1'deki verilen eşitlikler yardımıyla kestirilir. Varyans bileşenlerinin göreceli büyüklükleri, davranış ölçmesini etkileyen potansiyel hata kaynakları hakkında bilgi sağlar (Brennan, 2001; Crocker ve Algina, 1986).

G kuramında, genellenebilirlik (G) çalışmasından farklı olarak karar (K) çalışması olarak adlandırılan diğer bir çalışma vardır. K çalışmasında, G çalışmasında elde edilen bilgileri kullanarak belli bir amaçla yapılan bir ölçmedeki hataları en aza indirmenin yolları aranır. G çalışmasında yapılan bir ölçmenin güvenilirliği değerlendirilirken, karar çalışması genellenebilirlik çalışmasından hareketle karar üzerinde temellenen verileri toplamak üzere düzenlenir (Brennan, 2001; Crocker ve Algina, 1986; Shavelson ve Webb, 1991).

G Kuramı davranış ölçümlerinde bağıl (relative ya da norm-reference) ve mutlak (absolute ya da criterion reference) değerlendirme olmak üzere iki tür değerlendirmenin söz konusu olduğunu dikkate alır (Brennan, 2001; Goodwin, 2001). Bağıl bir değerlendirme, ölçme sonucunda elde edilen puanların gruba bağlı olarak çıkarılmış olan bir norma göre yapılırken, mutlak değerlendirme ise gruba bağlı olmadan önceden belirlenmiş mutlak bir ölçüte göre yapılır. Bu nedenle de G kuramında ölçme hataları, bağıl kararlar için ölçme hataları ve mutlak kararlar için ölçme hataları olarak değerlendirilir.

TÜRKİYE PERSONEL YÖNETİMİ DERNEĞİ

Tablo 1. Varyans bileşenleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Kestirilen Varyans Bileşenleri
Birey (b)	KT_b	$n_b - 1$	$KO_b = KT_b / sd_b$	$\sigma_b^2 = \frac{KO_b - \sigma_{bmd,e}^2 - n_d \sigma_{bm}^2 - n_m \sigma_{bd}^2}{n_m n_d}$
Değerlendirici (d)	KT_d	$n_d - 1$	$KO_d = KT_d / sd_d$	$\sigma_d^2 = \frac{KO_d - \sigma_{bmd,e}^2 - n_b \sigma_{dm}^2 - n_m \sigma_{bd}^2}{n_b n_m}$
Madde (m)	KT_m	$n_m - 1$	$KO_m = KT_m / sd_m$	$\sigma_m^2 = \frac{KO_m - \sigma_{bmd,e}^2 - n_d \sigma_{bm}^2 - n_b \sigma_{dm}^2}{n_b n_d}$
b x d	KT_{bd}	$(n_b - 1)(n_d - 1)$	$KO_{bd} = KT_{bd} / sd_{bd}$	$\sigma_{bd}^2 = \frac{KO_{bd} - \sigma_{bmd,e}^2}{n_m}$
b x m	KT_{bm}	$(n_b - 1)(n_m - 1)$	$KO_{bm} = KT_{bm} / sd_{bm}$	$\sigma_{bm}^2 = \frac{KO_{bm} - \sigma_{bmd,e}^2}{n_d}$
d x m	KT_{dm}	$(n_d - 1)(n_m - 1)$	$KO_{dm} = KT_{dm} / sd_{dm}$	$\sigma_{dm}^2 = \frac{KO_{dm} - \sigma_{bmd,e}^2}{n_b}$
b x m x d, e	$KT_{bmd,e}$	$(n_b - 1)(n_m - 1)(n_d - 1)$	$KO_{bmd,e} = KT_{bmd,e} / sd_{bmd,e}$	$\sigma_{bmd,e}^2 = KO_{bmd,e}$

Bu doğrultuda G kuramıyla hesaplanan G (genellenebilirlik) katsayısı bağıl ve mutlak değerlendirmeler için ayrı ayrı hesaplanabilir (Brennan 2001; Shavelson ve Webb, 1991; Shavelson ve Webb, 2003). İki değişkenli analiz için b x m x d deseninde yedi adet ana ve birbiriyle etkileşimde olan varyans bileşeni ($\sigma_b^2 + \sigma_m^2 + \sigma_d^2 + \sigma_{bm}^2 + \sigma_{bd}^2 + \sigma_{md}^2 + \sigma_{bmd,e}^2$) bir tek G çalışması ile kestirilebilir. K çalışmasında genellenebilirlik (generalizability) katsayısı (G ya da $E\rho^2$) ve güvenilirlik (dependability) katsayısı (Φ ya da Phi) olmak üzere iki farklı güvenilirlik katsayısı farklı ölçme senaryoları için hesaplanabilir (Brennan, 2001; Crocker ve Algina, 1986; Shavelson ve Webb, 1991; Shavelson ve Webb, 2003).

Bağıl hata terimi δ ile gösterilmek üzere bağıl hata varyansı;

$$\sigma_\delta^2 = \frac{\sigma_{bm}^2}{n_m} + \frac{\sigma_{bd}^2}{n_d} + \frac{\sigma_{bmd}^2}{n_m n_d} \quad (4)$$

eşitliği ile gösterilir. σ_b^2 bireylerin puanlarının evren değerinin varyansı olarak klasik test kuramındaki gerçek puanların varyansına karşılık geldiğinden genellenebilirlik katsayısı;

$$G = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\delta^2} \quad (5)$$

olarak yazılır. Bu eşitlikte, bağıl hata varyansı yerine konularak,

$$G = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_{bm}^2}{n_m} + \frac{\sigma_{bd}^2}{n_d} + \frac{\sigma_{bmd}^2}{n_m n_d}} \quad (6)$$

eşitliği ile bağıl hata değerlendirme için genellenebilirlik (G) katsayısı elde edilir (Brennan, 2001; Crocker ve Algina, 1986; Shavelson ve Webb, 1991; Shavelson ve Webb, 2003).

Mutlak hata terimi Δ ile gösterilmek üzere mutlak hata varyansı;

$$\sigma_\Delta^2 = \frac{\sigma_m^2}{n_m} + \frac{\sigma_d^2}{n_d} + \frac{\sigma_{bm}^2}{n_m} + \frac{\sigma_{bd}^2}{n_d} + \frac{\sigma_{md}^2}{n_m n_d} + \frac{\sigma_{bmd}^2}{n_m n_d} \quad (7)$$

eşitliği ile gösterilir. Bağlı hata varyansı eşitliğinde δ ile gösterilen bağlı hata terimi yerine Δ ile gösterilen mutlak hata terimi yazılırsa,

$$\Phi = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_m^2}{n_m} + \frac{\sigma_d^2}{n_d} + \frac{\sigma_{bm}^2}{n_m} + \frac{\sigma_{bd}^2}{n_d} + \frac{\sigma_{md}^2}{n_m n_d} + \frac{\sigma_{bmd}^2}{n_m n_d}} \quad (8)$$

eşitliği elde edilir. Bu eşitlik güvenilirlik (dependability) indeksi olarak adlandırılır (Brennan, 2001; Crocker ve Algina, 1986; Shavelson ve Webb, 1991; Shavelson ve Webb, 2003).

G ve Φ katsayılarında yer alan hata varyanslarını azaltabilmek için alternatif K çalışmaları düzenlenebilir (Yelboğa, 2007). Klasik test kuramında Spearman-Brown tahmin katsayısında ya da klasik test kuramında ortalamanın standart hatasında olduğu gibi n_m ve n_d ile gösterilen madde ve değerlendirici sayıları yerine değişik kombinasyonlarda değerlendirici ve madde sayıları yazılarak hata varyanslarının azaltılması ve daha yüksek G ve Φ katsayılarına ulaşmanın alternatifleri araştırılabilir. Bu tür bir araştırmaya G kuramında alternatif K çalışmaları adı verilir (Brennan, 2001; Crocker ve Algina, 1986; Shavelson ve Webb, 1991; Shavelson ve Webb, 2003).

Amaç

Bu araştırmada endüstri psikolojisinde yapılan bir ölçme işleminde genellenebilirlik kuramının kullanılabilirliği değerlendirilmiştir. Genellenebilirlik kuramı, bütün hata kaynaklarını birlikte değerlendiren ve ölçmenin güvenilirliğinin belirlenmesini sağlayan güçlü bir araçtır. Genellenebilirlik Kuramı kullanılarak, potansiyel hata kaynakları tanımlanabilir ve değerlendirilebilir. Bu çalışma ile Genellenebilirlik Kuramı kullanılarak G ve Phi katsayılarının hesaplanarak çok hata kaynaklı bir ölçme durumunda güvenilirliğin tek bir analizle gösterilmesi amaçlanmıştır.

Yöntem

Genellenebilirlik Kuramını inceleyen tarama türündeki bu araştırma için aşağıda örneklem grubu, veri toplama aracı ve işlem bölümleri verilmiştir.

Örneklem

Araştırma örneklemini, İstanbul ilinde faaliyet gösteren Üretim firmasında Üretim Şefi kadrosuna terfi ettirilmek üzere teknik yeterlilik sınavına alınan 10 aday mühendis personel oluşturmaktadır.

G kuramına göre her bir değişkenlik kaynağının düzeylerini oluşturan değerlendirici ve görevler, değerlendirici ve görev evrenini oluşturduğundan, sınav formunu değerlendirici olabilecek niteliklere sahip tüm yöneticiler ve bu sınav formunda kullanılacak görevler ise değerlendirici ve görev evrenini oluşturmaktadır. Bu evrenlere genelleme yapmak için sınav komisyonunda yer alan değerlendiriciler ve kullanılan görevler ise değerlendirici ve görev örneklemini oluşturmaktadır.

Veri Toplama Aracı

Araştırma'da İstanbul ilinde faaliyet gösteren Üretim firmasında Üretim Şefi kadrosuna terfi ettirilmek üzere 2008 yılında yapılan teknik yeterlilik sınavı verileri kullanılmıştır.

Teknik yeterlilik sınavında 10 adayın her biri; Üretim firmasının çeşitli bölümlerinde yönetici olarak görev yapan üç kişilik bir komisyon (değerlendirici örneklemini) tarafından on görev (görev örneklemini) üzerinden birbirinden bağımsız olarak 1 – 5 puan arasında teknik yeterlilik formu ile puanlanmıştır. Teknik yeterlilik formundaki görevlerin her biri adayların görev yaptığı üretim bölümü iş analizinden yararlanılarak hazırlanmıştır. Sınav esnasında kullanılan form ilgili firma tarafından halen kullanılmaktadır. Formun geçerliliği ve gizliliğini korumak amacıyla formda yer alan görevler bu makalede paylaşılmamaktadır. Uygulama sonucundaki puanlamalar Tablo 2'de verilmiştir.

Tablo 2. Uygulama verileri

Birey	1.Değerlendirici										2.Değerlendirici										3.Değerlendirici										
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	
1	4	3	2	2	4	4	3	2	2	4	4	3	2	2	2	4	3	2	2	2	2	3	3	3	3	2	3	3	3	3	2
2	4	2	3	4	3	4	2	3	4	3	3	3	1	2	1	3	3	1	2	1	3	3	2	3	2	3	3	2	3	2	2
3	1	1	0	1	1	1	1	0	1	1	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	4	0	0	3	3	4	0	0	3	0	4	3	1	2	0	4	3	1	2	1	3	0	0	3	1	3	0	0	3	
5	3	3	2	1	3	3	3	2	1	3	2	2	0	0	2	2	2	0	0	2	2	2	0	2	1	2	2	0	2	1	
6	4	2	0	0	2	4	2	0	0	2	1	2	1	1	1	2	1	1	1	1	2	4	0	1	0	2	4	0	1	0	
7	3	4	1	0	4	3	4	1	0	4	1	4	1	0	2	1	4	1	0	2	2	3	1	0	4	2	3	1	0	4	
8	2	3	3	3	4	2	3	3	3	4	2	2	3	3	4	2	2	3	3	4	3	4	4	3	3	3	4	4	3	3	
9	4	3	4	3	4	4	3	4	3	4	2	1	2	2	4	2	1	2	2	4	2	2	3	1	3	2	2	3	1	3	
10	3	4	1	3	4	3	4	1	3	4	3	4	0	3	4	3	4	0	3	4	1	2	0	2	4	1	2	0	2	4	

İşlem

Araştırmada, aday (a), görev (g) ve değerlendirici (d) değişkenlik kaynağını ifade etmektedir. Sınavdaki değişkenlik kaynağının bütün koşulları diğer bir değişkenlik kaynağının bütün koşullarını gözlediği için (a x g x d) G Kuramı çaprazlanmış deseni kullanılmıştır. Araştırma deseninde; adaylar, görevler, değerlendiriciler olmak üzere üç ana etki, aday-görev, aday-değerlendirici, görev-değerlendirici olmak üzere üç ortak etki ve kalan etki olmak üzere yedi varyans bileşeni hesaplanacaktır.

G kuramı içerisinde G çalışmasıyla varyans bileşenlerinin kestirilmesi ve K çalışmasıyla (ve alternatif K çalışması) G ve Phi katsayılarının hesaplanmasında GENOVA (Crick ve Brennan, 1983) programı kullanılmıştır.

Bulgular

Çalışmanın amacı doğrultusunda öncelikle Genellenebilirlik kuramı G çalışmasına ilişkin analizler daha sonra K çalışmasına ilişkin analizler yapılarak elde edilen bulgular verilmiştir.

Tablo 3.de verilen varyans bileşenleri ve toplam varyans içindeki değişimden hareket ederek varyansların bağıl büyüklüklerine bağlı olarak ana ve ortak etkiler aşağıdaki şekilde yorumlanabilir.

Tablo 3. G Çalışması ile kestirilen varyans bileşenleri (ANOVA Tablosu)

Varyans Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Kestirilen Varyans Bileşeni	Varyans Yüzdesi
b (aday)	175,52	9	19,502	0,563	29,25
g (görev)	80,453	9	8,939	0,233	12,10
d (değerlendirici)	24,826	2	12,413	0,108	5,61
b x g	141,413	81	1,745	0,412	21,40
b x d	24,24	18	1,346	0,083	4,31
g x d	12,506	18	0,694	0,018	0,94
b x g x d, e	82,426	162	0,508	0,508	26,39

Bireyler (b) ana etkisi için kestirilen varyans bileşeni (0,563) toplam varyans içinde en yüksek varyans payına sahiptir. Bireyler için kestirilen varyans bileşeni Klasik test kuramındaki gerçek puan varyansına karşılık gelen evren puanı varyansıdır. Bu sebepten dolayı bu varyansın büyük olması gerekir. Bu durumda bireylerin ölçülen özellik bakımından ayrılabildiklerini ifade etmektedir. Görev (g) ana etkisi için kestirilen varyans bileşeni (0,233) toplam varyansın % 12,10'nu açıklamaktadır. Görev ana etkisi için kestirilen varyans bileşeninin toplam varyansı açıklama oranı bazı görevlerin diğerlerine göre kolay veya zorluk bakımından birbirlerinden farklılaştıklarını göstermektedir. Değerlendirici (d) ana etkisi için kestirilen varyans bileşeni (0,108)'dir ve toplam varyansın % 5,61'ni açıklamaktadır. Değerlendirici (d) ana etkisi için kestirilen varyans bileşeninin yüzdesinin düşük olması, değerlendiricilerin bütün adaylar için yaptıkları puanlamaların bütün bireyler için eşit uzaklıkla (katılık/cömertlik) puanlama yaptıkları ifade edilebilir. Birey x görev (b x g) ortak etkisi (0,412) toplam varyansın % 21,40'mı açıklamaktadır. Bu durum, bireylerin bir görevden diğer göreve bağlı konumlarının değiştiğini göstermektedir. Birey x değerlendirici (b x d) ortak etkisi (0,083) toplam varyansın % 4,31'ini açıklamaktadır. Birey x değerlendirici varyans yüzdesinin sıfıra yakın olması, değerlendiricilerin hiçbiri, bireylerden hiçbirini diğer değerlendiricilere göre daha katı/cömert puanlamadıklarını göstermektedir. Görev x değerlendirici (g x d) ortak etkisi (0,018) toplam varyansın % 0,94'ünü açıklamaktadır. Bu varyans yüzdesi de değerlendiricilerin puanlamalarının bir görevden diğerine puanlamalarında bireyleri kararlı olarak puanladıklarını göstermektedir. Birey x görev x değerlendirici (b x g x d) ortak etkisinin (0,508) toplam varyansı açıklama oranının (% 26,39) ikinci en büyük varyans olduğu

görülmektedir. Artık varyansın büyük olması; birey x görev x değerlendirici ortak etkisinden kaynaklanan farkların ve/veya tesadüfi hataların çokluğunu göstermektedir.

G çalışmasıyla ana ve ortak etkilere dayalı olarak elde edilen varyans büyüklüklerine göre bağıl değerlendirmeler için G ve mutlak değerlendirmeler için Phi katsayısı aşağıdaki gibi hesaplanabilir.

Tablo 3'de kestirilen varyans bileşenlerini kullanarak (4) numaralı eşitlik yardımıyla bağıl hata varyansı;

$$\sigma_g^2 = \frac{0,412}{10} + \frac{0,083}{3} + \frac{0,508}{30} = 0,0858$$

olarak bulunur. Bağıl hata varyansı kullanılarak G katsayısı (5) numaralı eşitlik yardımıyla;

$$G = \frac{0,563}{0,563 + 0,0858} = 0,867 \text{ olarak kestirilir.}$$

Tablo 3'den kestirilen varyans bileşenlerini kullanarak mutlak hata varyansı (7) numaralı eşitlik yardımıyla;

$$\sigma_{\Delta}^2 = \frac{0,233}{10} + \frac{0,108}{3} + \frac{0,412}{10} + \frac{0,083}{3} + \frac{0,018}{30} + \frac{0,508}{30} = 0,1457$$

olarak hesaplanır. Mutlak hata varyansı (8) numaralı eşitlik kullanılarak Phi katsayısı;

$$\Phi = \frac{0,563}{0,563 + 0,1457} = 0,793 \text{ olarak kestirilir.}$$

Yukarıda verilen bağıl hata ve mutlak hata eşitliklerinde bulunan görev sayısı ve değerlendirici sayıları değiştirilerek ölçme amacına uygun olarak G ve Phi katsayıları hesaplanabilir. Dolayısıyla farklı görev ve değerlendirici sayılarıyla alternatif K çalışmaları sonucu elde edilen bağıl ve mutlak hata varyansları ile G ve Phi katsayıları kestirilir.

Tablo 4. Alternatif K Çalışması ile kestirilen G ve Phi katsayıları

Değerlendirici Sayısı	Görev Sayısı	Alternatif K Çalışmaları	
n_d	n_g	G	Phi
2	9	0,829	0,740
2	10	0,838	0,750
2	11	0,846	0,759
3	9	0,858	0,783
3	10	0,867	0,793
3	11	0,874	0,802
4	9	0,874	0,807
4	10	0,882	0,817
4	11	0,889	0,825

G Katsayma ilişkin bulgular aşağıdaki şekilde verilebilir; Orijinal değerlendirici (3) ve görev (10) sayıları için Genellenebilirlik Kuramı K çalışması ile kestirilen G katsayısı 0,867 olarak hesaplanmıştır. Alternatif K çalışmasında değerlendirici sayısı sabit tutularak (3) görev sayısının bir azaltılması (9) durumunda G katsayısı 0,858, görev sayısının bir artırılması (11) durumunda ise G katsayısı 0,874 olarak hesaplanmıştır. Değerlendirici sayısının sabit tutulması koşulu altında, görev sayısının azaltılması ile kestirilen G katsayısındaki azalma 0,009 olurken, görev sayısının artırılması durumunda kestirilen G katsayısındaki artış 0,007 olmuştur.

Aynı biçimde görev sayısı sabit tutularak (10) değerlendirici sayısının bir azaltılması (2) durumunda G katsayısı 0,838, değerlendirici sayısının bir artırılması (4) durumunda ise G katsayısı 0,882 olarak hesaplanmıştır. Görev sayısının sabit tutulması koşulu altında, değerlendirici sayısının azaltılması ile kestirilen G katsayısındaki azalma 0,029 olurken, değerlendirici sayısının artırılması durumunda kestirilen G katsayısındaki artış 0,015 olmuştur.

Görüldüğü gibi her ne kadar değerlendirici sayısının artırılması durumunda G katsayısı görev sayısının artırılması durumundakinden büyük olsa da, G katsayısındaki artış çok büyük değildir.

Phi katsayma ilişkin bulgular incelendiğinde; Orijinal değerlendirici (3) ve görev (10) sayıları için Genellenebilirlik Kuramı K çalışması ile kestirilen Phi katsayısı 0,793 olarak hesaplanmıştır. Alternatif K çalışmasında değerlendirici sayısı sabit tutularak (3) görev sayısının bir azaltılması (9)

durumunda Phi katsayısı 0,783, görev sayısının bir arttırılması (11) durumunda ise Phi katsayısı 0,802 olarak hesaplanmıştır. Değerlendirici sayısının sabit tutulması koşulu altında, görev sayısının azaltılması ile kestirilen Phi katsayısındaki azalma 0,01 olurken, görev sayısının arttırılması durumunda kestirilen Phi katsayısındaki artış 0,009 olmuştur.

Aynı biçimde görev sayısı sabit tutularak (10) değerlendirici sayısının bir azaltılması (2) durumunda Phi katsayısı 0,750, değerlendirici sayısının bir arttırılması (4) durumunda ise Phi katsayısı 0,817 olarak hesaplanmıştır. Görev sayısının sabit tutulması koşulu altında, değerlendirici sayısının azaltılması ile kestirilen Phi katsayısındaki azalma 0,043 olurken, değerlendirici sayısının arttırılması durumunda kestirilen Phi katsayısındaki artış 0,024 olmuştur.

Görüldüğü gibi burada da her ne kadar değerlendirici sayısının arttırılması durumunda Phi katsayısı görev sayısının arttırılması durumundakinden büyük olsa da, Phi katsayısındaki artış çok büyük değildir.

Tartışma

Klasik test kuramının sayılılarına dayanan güvenilirlik kestirimleri güçlü teknikler olmasına karşın, tüm güvenilirlik ya da güvenilirlik sorunlarını yeterince kapsayamaz. Klasik test kuramının işlemleri bir bireyin iki ayrı ölçümüne dayandığından bazı durumlarda bu güvenilirlik kestirimleri yetersiz kalır. Örneğin, bireylerin iş performanslarının belli bir puan üzerinden dört değerlendirici tarafından puanlandığı durumda klasik test kuramının güvenilirlik kestirimleri kullanılamaz. Diğer bir örnek, üç klinik psikoloğun iki farklı oturumda danışanların ilgilenilen özelliklerini derecelendirmeleri verilebilir. Böyle bir durumda klasik test kuramının işlemleri doğrudan uygulanamaz. İstatistik konularında hatırlanacağı üzere değişim kaynaklarını belirlemek varyans analizi teknikleriyle mümkün olmaktadır.

Ölçme sonuçlarına karışabilecek hataların kaynaklarına yönelik güvenilirlik hesaplama yöntemleri, ölçme kuramlarının başlıca çalışma alanlarıdır. Klasik test kuramı ve madde tepki kuramının birçok modeli tek bir analizle, tek bir değişkenlik kaynağından gelen hatayı dikkate alır. Oysa birçok ölçme durumunda potansiyel hata ya da değişkenlik kaynağı birden çoktur. Birden çok değişkenlik kaynağından gelen hatanın birlikte değerlendirilmesi ölçmenin uygulama alanının psikoloji ve eğitim alanlarında genişlemenin yanında diğer alanlara da giderek yayılması sonucunda artan bir gereklilik

olmuştur. Sonuç olarak değişkenlik kaynaklarından gelen potansiyel hataların birlikte değerlendirilmesi ve ortak bir güvenilirlik katsayısının bir analizle hesaplanmasının daha anlamlı olacağı düşünülmektedir.

Genellenebilirlik kuramı, davranış ölçmede güvenirliliğin değerlendirilmesini, güvenilir gözlemlerin tasarlanmasını ve incelenmesini sağlayan varyans analizine (ANOVA) dayalı istatistiksel bir kuramıdır. Bir ölçme durumu için olası hata kaynaklarının tümünü tek bir analizle kapsamlı olarak bağıl ve mutlak değerlendirmeler için ayrı ayrı kestirebilmektedir.

Bu araştırma kapsamında Genellenebilirlik kuramının personel değerlendirmede kullanılması örnek vaka olarak sunulmuştur. Özellikle psikoloji ve eğitim alanlarında araştırmacılar güvenirliliğin değerlendirilmesi çalışmalarında Genellenebilirlik kuramının, Klasik test kuramına göre daha avantajlı ve ekonomik analiz yöntemi ortaya koyduğunu göz önünde bulundurmalarıdır. Bundan dolayı gerek güçlü bir istatistiksel temele sahip olması gerekse de psikometrik çalışmalarda kullanım kolaylığı getirmesi açısından, Genellenebilirlik kuramının psikoloji ve eğitim çalışmalarında güvenilirlik belirlemede klasik test kuramının yerine kullanılabilmesi ifade edilebilir.

Kaynakça

- Anastasi, A. (1988). *Psychological Testing*, New York: Macmillan Publishing Co.Inc.
- Atılğan, H. (2004). *Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin Bir Araştırma*. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi. Ankara.
- Atılğan, H. ve Tezbaşaran A.A. (2005). Genellenebilirlik Kuramı Alternatif Karar Çalışmaları ile Senaryolar ve Gerçek Durumlar İçin Elde Edilen G ve Phi katsayılarının Tutarlılığının İncelenmesi. *Eğitim Araştırmaları*, 5(18), 28-40.
- Baykuş, Y. (1999). *İstatistik: Metodlar ve Uygulamalar*. Anı yayıncılık. Ankara.
- Baykuş, Y. (2000). *Eğitimde ve Psikolojide Ölçme*. Ankara: Cem Web Ofset
- Brennan, R. L. (2001). *Generalizability Theory*. Springer-Verlag, New York.
- Crick, J. E. & Brennan, R. L. (1983). *Manual For GENOVA: A Generalized Analysis of Variance System*, The American College Testing Program, Iowa.
- Crocker, L.M. ve Algina, L. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.
- Cronbach, L.J. (1984). *Essential of Psychological Testing (4th Edition)*. New York: Harper and Row Publish

- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16, 137-163.
- Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.
- Ghiselli, E.E., Campbell, J.P. ve Zedeck, S. (1981). *Measurement Theory For The Behavioral Sciences*. San Francisco: W.H. Freeman and Co.
- Guilford, J.P. (1954). *Psychometrics Methods*. 2nd Edition. New York: McGraw-Hill Book Co.
- Gulliksen, H. (1967). *Theory of Mental Tests*. 6th Edition. John Willey and Sons, New York.
- Goodwin, L. D. (2001). Interrater agreement and reliability, *Measurement in Physical Education and Exercises Science*, 5(1), 13-14
- Lord, F.M. & Novick, M.R. (1968). *Statistical Theories Of Mental Test Scores*. New Jersey: Addison-Wesley.Co.
- Ozçelik, D.A. (1989). *Test Hazırlama Kılavuzu*, OSYM Yayınları. Ankara.
- Shavelson, R.J & Webb, M.N. (1991). *Generalizability Theory: A Prime*. SAGE Publication, Inc., California.
- Shavelson, R.J. & Webb, M.N. (2003). "Generalizability theory" *Encyclopedia of Social Measurement*, ed. Kempf-Leonard, Kimberly. Academic Pres, San Diego.
- Urbina, S. (2004). *Essentials of Psychological Testing*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Yelboğa, A. (2007). *Klasik Test ve Genellenebilirlik Kuramına Göre Güvenirliğin Birliş Performansı Ölçeği Üzerinde İncelenmesi*, Yayınlanmamış Doktora Tezi, Ankara Üniversitesi, Ankara.

Yazışma Adresi:

Dr. Atilla Yelboğa

Ziya Gökalp Cad. No:64

Kurtuluş/ANKARA

Tel: 0532 7922608

e-posta: ayelboga@hrm.gen.tr