

Ölçme Araçlarında İki Yaşamsal Kavram: Geçerlik ve Güvenirlik

Dr. Servet AKER, Dr. Cihad DÜNDAR, Dr. Yıldız PEKŞEN

Ondokuz Mayıs Üniversitesi Tıp Fakültesi Halk Sağlığı Anabilim Dalı, SAMSUN

- ✓ Geçerlik ve güvenilirlik; insanlar hakkında önemli kararlar verebilmek amacıyla geliştirilen ölçme araçlarının olmazsa olmaz, yaşamsal iki niteliğidir. Araştırmacıların, ölçme aracı seçiminde ve kullanımında çok titiz, dikkatli olmaları kaçınılmaz bir sorumluluktur. Çünkü; bu özellikleri sınanmadan kullanılan ölçme araçları geri döndürülemez sonuçlara yol açabilir. Bu derlemenin amacı; araştırmacılara bu iki kavram hakkında genel bir bilgi vermek ve dikkatlerini bu konulara çekebilmektir.

Anahtar kelimeler: Ölçme aracı, geçerlik, güvenilirlik

- ✓ **Two Vitaly Concept of Measurement instruments: Validity and Reliability**
Reliability and validity are important characteristics of measurement instruments, which are developed for making critical decisions about individuals. Researchers need to be extremely careful in choosing/ using measurement instruments, because those were not examined carefully may cause irreversible results. The purpose of this paper was to inform the researchers about these two concepts and attract their attention on this subject.

Key words: Measurement instruments, reliability, validity

Ölçme; herhangi bir objenin belirli bir niteliğini-özelliğini, belirli kurallara göre, sayarak, sınıflandırarak, derecelendirerek ya da birimlerle sayısal olarak ifade etme süreci olarak tanımlanmaktadır⁽¹⁾.

Ölçmenin, günlük yaşamda ve bilimsel çalışmalarda yeri yadsınamaz. Hemen hemen tüm bilim dallarının bir kuramsal bir de deneysel yönü vardır ve bu iki yönün ilişkisi ölçüm ile kurulur. Bilim dallarında; kuramsal çalışmalarla ortaya atılan kavramsal yapının, deneysel çalışmalarla uygun koşullarda gözlemlenip, sembolleştirilmesi gerekir. Ölçme araçlarıyla sembolleştirilebilen kavramlar genelleştirilebilir, bu da bilim dallarındaki ilerlemelere yön ve hız verir.

Ölçme araçları geliştirmede olduğu kadar ülkemiz koşullarına uyarlama çalışmalarında da sorgulanması zorunlu olan iki yaşamsal kavram: geçerlik ve güvenirliliktir.

Her ölçme aracı daima belirli bir amaç için,

belirli koşullar altında ve belirli bir gruba uygulanmak üzere geliştirilir⁽²⁾. Örneğin; ilköğretim 7. sınıf öğrencilerinin matematik bilgisini ölçmek için geliştirilen bir ölçme aracı, aynı öğrencilerin Türkçe bilgisini ölçmek için kullanılamayacağı gibi, ilköğretim 4. sınıf öğrencilerinin matematik bilgisini ölçmek için de kullanılamaz. Dolayısıyla, herhangi bir ölçme aracı geliştirildiği amaca hizmet etmelidir. Ölçme aracının amaca hizmet etmesi, onun ilgili özelliği doğrulukla ölçmesiyle yakından ilişkilidir. Doğru bir ölçüm yapamayan ya da doğru ölçüm yapıp, kullanılmama amacına hizmet etmeyen bir ölçme aracının kullanılması uygun değildir⁽³⁾.

Bu durum ölçme araçlarının güvenilirliğinin ve geçerliğinin birlikte ele alınmasını gerekli kılar. Çünkü; geçerlik ve güvenilirlik kavramları birbirleriyle ilintili ve iç içe kavramlardır. Güvenirlik; bir ölçme aracının hatalardan arınık olarak ölçme yapabilme yeteneği, geçerlik

ise; ölçme aracının kullanılma amacına hizmet etme derecesi olarak tanımlanabilir. Bu kavramlar şu şekilde karikatürize edilebilir. A, B ve C olarak adlandırılan 3 kişinin randevularına gelme özellikleri; A randevularına bazen erken, bazen zamanında, bazen de geç, B sürekli tam 10 dakika sonra veya önce ve C ise daima tam zamanında geliyor ise 3 kişiden (ölçek olarak düşünülebilir) A kişisi ne güvenilir ne de geçerli, B kişisi güvenilir ama geçerli değil, C kişisi ise hem güvenilir hem de geçerlidir⁽⁴⁾.

Diğer nitelikleri sınanmış olsa da, eğer testlerin güvenilirliği ve geçerliği düşük düzeyde ise kullanılması sakıncalıdır. Bunlar sağlanmadığı zaman, uygulamada bireyler hakkında karar vermek (seçme, tanı, sağaltım, yönlendirme vb.) ya da araştırma bulgularını geçerli kabul etmek, geriye dönülmesi olanaksız, zararlı sonuçlar yaratabilir⁽⁵⁾.

Güvenirlik (Reliability)

Güvenirlik ölçme araçlarının önemli teknik özelliklerinden biridir. Kaynaklar incelendiğinde güvenilirlikle ilgili tanımların birbirine eşdeğer ya da birbirini tamamlayan anlatımlar olduğu gözlenir.

Güvenirliğin; "bir ölçme aracının hatalardan arınık olarak ölçme yapabilme yeteneği"⁽³⁾, "bir ölçme aracında bütün soruların birbiri ile tutarlılığı"⁽⁵⁾, "ele alınan oluşumu ölçmede türdeşliğini ortaya koyan kavram"⁽⁶⁾, "ölçme aracının duyarlı, birbiriyle tutarlı ve kararlı ölçme sonuçları verebilmesi", "aynı değişkenin bağımsız ölçümleri arasındaki kararlılık", "ölçülme istenen belli bir değişkenin, sürekli olarak aynı sembolleri alması", "aynı süreçlerin izlenmesi ve aynı ölçütlerin kullanılması ile aynı sonuçların alınması"⁽⁷⁾, "ölçmenin rastlantısal yanılılardan arınık olması"⁽⁸⁾, "ölçme aracının ölçtüğü özelliği ya da özellikleri ne derecede kararlılıkta ölçmekte olduğunun göstergesi"⁽⁹⁾, "paralel testler arası korelasyon"⁽¹⁰⁾, "aynı testleri farklı durumlarda alan bireylerin puanlarının tutarlılığı"⁽¹¹⁾, "bir bireyin kazandığı puanların kendi kendisiyle tutarlılığı"⁽¹²⁾, "belirli bir durumda araçtan elde edilen sonucun yeniden üretilebilirliği"⁽¹³⁾ ya da "ölçüm

aracının yinelenebilir sonuç verme yeteneği" şeklinde çeşitli tanımları vardır. Ancak bu tür tanımlamalar güvenilirliğin sadece bir türünü yansıtmaktadır⁽²⁾. Oysa güvenilirlik saptamanın bir çok yolu vardır, bu bakımdan böylesi tanımlar işlemsel (işe vuruk, operational) olarak görülmelidir⁽³⁾.

Daha açık bir anlatımla; bir ölçme aracıyla her hangi bir grupta ya da bireyde ölçüm yapıldığında, ölçme aracının her uygulandığında birbirine benzeyen, kararlı sonuçlar alınması beklenir. Eğer ölçme aracı aynı koşullarda aynı bireylere uygulandığında farklı sonuçlara ulaşıyorsa ölçme aracının güvenilirlik derecesinin düşük olduğu söylenebilir. Sonuçta güvenilirlik kavramına; değişmezlik, yeterlilik, kestirim, eşdeğerlik ve tutarlılığın sağlanması gibi sıfatları yüklemek olasıdır.

Kavramsal ve istatistiksel olarak güvenilirlik ise; "bir ölçme aracındaki gerçek farklılıkların, toplam farklılığa oranıdır"⁽¹⁴⁾.

Klasik test teorisinin güvenilirlik kavramı Spearman'ın basit eşitlikleri üzerine temellenir ve $X_{gözlenen} = X_{gerçek} + X_{hata}$ şeklinde formülize edilir⁽³⁾.

Burada, hata puanlarının azaltıldığı oranda gözlenen puanın gerçek puana yaklaşacağı açıktır. Ancak her zaman ara değişkenler şu ya da bu şekilde ölçümlere karışacağı için, ölçümlerin hatalardan tümüyle arınık olması mümkün değildir. Puanlardaki değişime yol açan tüm etkenler asla tam olarak açıklanamaz; eğer öyle olsaydı, güvenilirlik teorisine de gerek kalmazdı. Bu nedenle güvenilirlik tam olarak belirlenemez; ancak kestirimden söz edilebilir⁽⁹⁾.

Hatayı, ölçülme istenmeyen miktarın, ölçüm sonuçlarına karışması olarak tanımlamak mümkündür. Bir araştırmada 3 tip ölçüm hatası olabilir.

a) Sabit hatalar, her bir ölçmede aynı miktar ve yönde olan hatalardır. Bu tür hatalar ölçek iyi kalibre edilmediğinde oluşur. Örneğin; bir tartı aletinin kişinin gerçek ağırlığından daha az/çok göstermesi gibi.

b) Seri hatalar, belirli bir yönde ilerleyici ya da gerileyici miktarlarda ortaya çıkan hatalardır. Örneğin yaşlılık ve unutkanlık ilerleyici bir

şekilde puanlarda azalmalara, öğrenme ve olgunlaşma gibi etmenler ise ilerleyici bir şekilde puanlarda artışlara yol açar. Seri hatalar bireylerin ilgili özelliğindeki gerçek değişimlerdir.

Sabit ve seri hatalar sistematik hatalar olarak değerlendirilirler.

c) Rastlantısal (seçkisiz, sistematik olmayan) hatalar, değişen kişisel koşullar, çevre koşulları ya da ölçüm aracındaki değişiklikler nedeniyle oluşabilen hatalardır. Bu tür ölçüm hataları güvenilirlik için daha önemlidir.

Rastlantısal hatalar ne yönde ve miktarda olduğu bilinmeyen hatalardır. Hataların bu özelliğinden hareketle “gerçek puan teorisi” ya da “klasik test teorisi” bir takım sayılılar geliştirmiştir. Bu sayılıları dikkate alan klasik test teorisi, paralel testlerle yapılan sonsuz sayıda tekrarlı ölçümde gözlenen puanların ortalamasının gerçek puana eşit olacağını öngörür.

Ölçme Araçlarının Güvenirliğini Sınama Yöntemleri;

Ölçme araçlarının güvenirliliklerini sınımanın birçok yolu vardır. Güvenirliği sınımada, bunlardan biri kullanılabilir gibi, birkaçı da aynı anda kullanılabilir. Mümkünse ölçme aracının güvenirliliği için birden çok yola başvurmak uygun olacaktır. Ancak bir ölçme aracının güvenirlilik kestirme yollarını sınırlayan faktörler vardır. Bunları 4 başlık altında toplamak mümkündür.

1. Testin niteliği ve biçimi (testin hız veya güç testi olması, puanlama biçimi, farklı boyutları ölçen alt-testlerden oluşup oluşmadığı, madde güçlük düzeyleri, yordama amaçlı kullanılıp kullanılmaması, karmaşık bir psikolojik yapıyı ölçüp ölçmemesi)

2. Pratik olanaklar (uygulama zamanı, çoğaltma ve geliştirme masrafları)

3. Daha önce ölçülmüş olmanın bireylerin ilgili özelliğini değiştirip değiştirmediği (öğrenme, olgunlaşma gibi faktörler)

4. Hata kaynakları

Araştırmacı bu faktörleri göz önüne alarak hangi güvenirlilik yöntem ya da yöntemlerini uygulayacağına karar vermelidir. Birçok güvenirlilik ölçütünden söz edilmekle birlikte, bunlar temelde üç başlık altında toplanmaktadır.

I. Değişmezlik (Stability) – Kararlılık:

a) Test – tekrar test yöntemi:

Test-tekrar test güvenirliliğini, “bir ölçme aracının uygulamadan uygulamaya tutarlı sonuçlar verebilme, zamana göre değişmezlik gösterebilme gücü” şeklinde tanımlamak mümkündür⁽¹⁵⁾.

Bir ölçme aracının aynı bireylere, aynı koşullarda, ancak farklı zamanlarda uygulanması esasına dayanır. Bireylerin; genel bilişsel yetenekler ve kişilik özellikleri gibi, devamlılık-kararlılık gösteren özelliklerinin ölçüldüğü, ölçme araçlarının güvenirliliğinin sınanmasında kullanılır. Uygulamalar sonucunda elde edilen korelasyon katsayısına devamlılık ya da kararlılık katsayısı da denir. Ancak ölçülecek özellik; bilgi, tutum, ruh hali ve fiziksel durum gibi, sürekli değişiklik gösterebilecek bir özellik ise kullanılması doğru değildir⁽⁷⁾.

Bu yöntemin; paralel test yönteminden farklı olarak eşdeğer bir ölçüğe gereksinim duymama ve uygulanan her iki test de aynı olduğu için kapsam geçerliği sıkıntısına yol açmama gibi iki temel avantajı bulunur.

Test-tekrar test yönteminde dikkat edilmesi gereken özellik, iki uygulama arasında bırakılması gereken zaman aralığıdır. Bırakılan zaman aralığının çok kısa olması, yeniden anımsamayı kolaylaştıracağından, güvenirliliğin yapay olarak yüksek çıkmasına (taşımaya etkisi) neden olabilir. Zaman aralığı uzun tutulduğunda ise, iki ölçme için “aynı koşullar” ’ın sağlanması olanaksız olabileceğinden ölçülen özellikte bazı değişimlerin oluşması sonucu güvenirlilik ölçütünün yorumu güçleşir. Ayrıca bireyin, testin içeriği ile ilgili olarak başka kaynaklardan öğrendikleri, ikinci uygulamadaki puanını etkileyebileceği gibi bireyin ilk uygulamada test üzerinde pratik yapmış olması ikinci uygulamadaki puanlarının artmasına da neden olabilir.

Olası sakıncalar dikkate alındığında iki uygulama arasındaki süre, ölçülecek özelliğe göre değişmekle birlikte, genellikle 2-3 ile 4-6 hafta arasında olmalıdır⁽¹⁶⁾.

Olası sınırlılıklarına karşın bir testin zamana göre değişmezliğinin en önemli ölçütü olan

test-tekrar test yöntemi, en sık kullanılan ve önerilen güvenilirlik göstergesidir.

Geliştirilen ya da uyarlanan ölçeğin test-tekrar test güvenilirliğini bulmak için iki uygulamadan elde edilen puanlar arasındaki korelasyon hesaplanır. Ölçek puanları sürekli bir değişken ve eşit aralıklı ölçek nitelikleri taşıdığı zaman güvenirliliğin bulunmasında Pearson Momentler Çarpımı Korelasyon eşitliği kullanılmalıdır⁽¹⁴⁻¹⁶⁾. Ancak burada korelasyon katsayısı örneklem büyüklüğünden etkilenir. Örneklem büyüklüğünün en az 30 olması gerekir, daha küçük gruplarda korelasyon katsayısı değişkenliklere açıktır⁽¹⁶⁾. Eğer değişkenler sıralama ölçeğinden elde edilmiş ya da örneklem büyüklüğü 30'un altında ise Spearman'ın Sıralama Farkı Korelasyon Katsayısı kullanılmalıdır⁽¹⁶⁾. Korelasyon katsayısı doyurucu olsa bile, her iki testin ortalamaları ve standart sapmaları incelenmelidir. Değişken gerçekten kararlı ise, her iki ölçeğin ortalama ve standart sapma değerleri birbirine yakın olacaktır.

b) Paralel (Alternatif-Eşdeğer) formlar yöntemi:

Birbirine eşdeğer formların, aynı bireylere, aynı koşullarda aynı gün ya da farklı günlerde uygulanması esasına dayanır⁽³⁾. Genellikle ölçek geliştirilirken kullanılan bir yöntemdir⁽⁷⁾. Ancak daha önce aynı amaçlar için geliştirilmiş bir ölçeğin (referans test) varlığında kullanılabilir. Uygulanan iki formdan elde edilen ölçümler arasında korelasyon katsayısı Pearson Momentler Çarpımı formülüyle hesaplanır⁽¹⁶⁾. Bulunan katsayı eşdeğerlik katsayısıdır ve ölçeğin paralel formlar güvenirliliğini verir. Ancak sadece korelasyon katsayısı belirleyici değildir. Ayrıca iki ölçeğin eşdeğer olması için; ölçeklerin ortalamalarının, standart sapmalarının, varyanslarının da eşit olması gerekir^(3,16).

Ölçeğin iki paralel formu yoksa bu güvenilirlik katsayısı hesaplanamaz. Bunun için her iki formun madde sayısı, niteliği, ölçekleme tekniği, faktör yapısının da denk olması gerekir. İki formun eşdeğer olabilmesinin ölçütlerini yerine getirebilmek araştırmacı için güç olduğu gibi uygulama, geliştirme zamanı

ve masraflar artacağından, uygulamada pek tercih edilen bir güvenilirlik sınaama yöntemi değildir^(3,14-16).

II. Ölçümcü Güvenirliliği (Bağımsız gözlemciler arası ve gözlemci içi uyum= Inter-rater and intra rater consistency):

Verilerin gözleme dayandığı, birden çok sayıda gözlemcinin, önceden eğitilerek ve birbirinden bağımsız, aynı durumu, aynı zamanda, aynı ölçme aracı ile ölçmeye çalıştıkları durumlarda, gözlemciler arasındaki uyumu saptamak amacıyla kullanılan güvenilirlik ölçütüdür. Gözlemcilerin ölçüm sonuçları ortalamaları ve standart sapmaları birbirine ne kadar yakınsa gözlemciler arası güvenilirlik de o kadar yüksektir^(14,15).

Birden fazla ölçümcü arasında %70 ve daha yüksek tutarlılık, güvenilirlik sınaaması için uygundur. Ancak, tutarlılık yüzdesinin, şansa bağlı tutarlılığın yüksek olması nedeniyle yüksek çıkabileceği dikkate alınmalıdır. Bunun için yalnızca şansa bağlı olmayan (rastlantısal olmayan uyum) beklenen uyumu gösteren Cohen Kappa istatistiği kullanılır. Kappa değeri 0 ise gözlemciler arasında hiç uyum yoktur, 1 ise uyum mükemmeldir, -1 ise bir gözlemcinin "ak" dediğine diğeri "kara" demektedir⁽¹⁷⁾.

Bağımsız gözlemciler arası uyumu hesaplamak için korelasyon, t-testi ya da ikiden fazla gözlemcinin olduğu durumlarda Kendall'ın Konkordans Sayısı (Kendall'ın Uyuşum)⁽¹⁶⁾ ve Cronbach alfa kullanılır

İki ya da daha fazla gözlemin, aynı gözlemci tarafından yapılarak puanlanması söz konusu olduğunda, gözlemci içi uyumu saptamak amacıyla güvenilirlik sınaaması yapılabilir. Örneğin, anne-çocuk etkileşimine ilişkin bir video kasetinin aynı kişi tarafından farklı zamanlarda izlenmesi ve puanlanması ya da bir radyoloji uzmanının aynı röntgen filmlerini farklı zamanlarda değerlendirmesi gibi. Burada da güvenilirlik sınaaması için en çok kullanılan istatistik yöntem ölçümler arası tutarlılığın yüzdesidir. Ayrıca, gene iki ya da fazla ölçümcünün kendi içlerindeki uyumu göstermek

için Cronbach Alfa ve şansa bağlı tutarlılığın etkisini azaltmak için ise Kappa istatistiği kullanılabilir^(14,15).

III. İç Tutarlılık (Internal Consistency):

Bir ölçme aracının birden çok uygulanması, hem ölçme aracının hem de bireylerin ölçülen özelliğinin niteliği, zaman, ekonomi gibi nedenlerle olanaklı olmayabilir. Bu güçlükler nedeniyle ölçme araçlarının tek kez uygulanmasıyla güvenilirlik kestirimi yöntemleri geliştirilmiştir. Ölçek bir gruba tek kez uygulanır ve ölçeğin kendi kendisiyle tutarlığına bakılır. Bu yüzden bu yönteme “iç tutarlık”, elde edilen güvenilirlik katsayısına da “iç tutarlık katsayısı” denir. Ölçek güvenilirliği yöntemleri içinde en sık kullanılanıdır⁽¹⁶⁾.

İç tutarlılığın temel dayanağı; “her ölçme aracının belli bir amacı gerçekleştirmek, bir bütünü oluşturmak üzere, birbirinden deneysel olarak bağımsız ünitelerden oluştuğu ve bunların bir bütün içinde, bilinen ve birbirine eşit ağırlıklara sahip olduğu” varsayımıdır⁽¹⁵⁾.

Farklı iç tutarlılık güvenilirlik saptama yöntemleri vardır. Bir ölçme aracının iki yarıya bölünmesi yöntemi bunlardan biridir. Ölçeğin her iki yarısının birbirine paralel olduğu varsayımı üzerine kuruludur; yani her iki yarının ortalama ve varyanslarının eşit olduğunu kabul eder. Sonuçta elde edilen güvenilirlik katsayısına da “eşdeğer iki yarı güvenirligi” denir. Eğer ölçek tek boyuttan oluşuyorsa ölçeğin tümü için uygulanabileceği gibi, ölçeğin alt boyutları varsa her alt boyut için kendi içinde bir bütün olarak kabul edilip alt boyutlar için de uygulanabilir.

Yarıya bölme yöntemi motivasyon, yorgunluk ya da zaman içerisinde performansı etkileyecek diğer psikolojik etkenler söz konusu olduğunda ve özellikle de uzun testlerde kullanılır. Aynı fizik, mental ve çevresel etkenler söz konusu olduğundan test-tekrar test ve paralel formlara göre daha güçlüdür⁽⁷⁾. Bu yöntemde; ölçek bir gruba tek kez uygulanır ve puanlamadan önce, ölçek maddeleri tek-çift olarak iki eşit yarıya ayrılır. İki yarıdan elde edilen ölçümler arasında koşullar yerine getiriliyorsa

Pearson Momentler Çarpımı Korelasyon katsayısı hesaplanır. Bu katsayı yarı testin korelasyon katsayısıdır. Ölçeğin bütününe ilişkin güvenilirlik katsayısını elde etmek için Spearman-Brown tarafından geliştirilen bir eşitlikten yararlanılmaktadır. Ayrıca ölçeğin iki yarısına ait puanlarından iç tutarlık katsayısı hesaplamada Stanley, Cronbach alfa, Rulon, Flanagan, Mossier, Horst gibi teknikler de kullanılabilir⁽¹⁶⁾. Testin iki yarısına ilişkin varyans eşit ya da çok yakın değilse güvenilirlik katsayısı Cronbach alfa formülü ile hesaplanmalıdır^(15,18).

İç tutarlılık hesaplamada; madde kovaryansına dayanan güvenilirlik hesaplama yöntemleri de vardır. Burada ölçeğin maddeleri birçok testin bileşkesi gibi görülür; yani her madde diğer maddelerin her biriyle paralelmiş gibi değerlendirilir. Bu yöntemler uygulama ve hesaplamadaki kolaylıkları nedeniyle sıklıkla kullanılırlar^(3,16).

Homojen bir yapıyı ölçtüğü varsayılan ve benzer maddelerden oluşan ölçme araçlarının tek uygulamayla güvenilirliğinin belirlenmesi, o ölçme aracının iç tutarlılığı hakkında bilgi verir⁽¹⁹⁾. Burada; ölçeğin her bir maddesinin aynı ortalama ve varyansa sahip olduğu varsayımından hareket edilir. En çok kullanılan formüller; Kuder Richardson 20 (KR-20) ve Cronbach Coefficient Alfa teknikleridir. Kuder Richardson 20 ve Cronbach Alfa formülleri birbirlerinden türetilmiş olmasına karşın, hangisinin kullanılacağı, madde puanlarının ölçeklenme biçimine bağlıdır. Madde puanları var-yok, evet-hayır, doğru-yanlış veya 1-0 şeklinde süresiz ise KR - 20; Likert tipi ölçeklerde olduğu gibi sürekli ise Cronbach Alfa hesaplanması gerekir. Ayrıca Kuder-Richardson 21 (KR-21) eşitliği ise, ölçekteki her sorunun güçlük derecesinin aynı olduğu varsayımında kullanılabilir. Uygulamada bu varsayım nadiren gerçekleştiği için KR-21 çok fazla kullanılmaz^(3,4,16).

SPSS programındaki “Reliability-alpha” seçeneği Cronbach Alfa katsayısını hesapladığı için bu işlem, 1-0 şeklinde puanlanan maddeli testlere uygulanmamalıdır^(4,19).

Birbiriyle yüksek ilişki gösteren maddelerden oluşan ölçeklerin alfa (α) katsayısı yüksek olur. Cronbach Alfa katsayısı ölçek içinde bulunan maddelerin iç tutarlılığının, homojenliğinin ölçüsüdür. Ölçeğin alfa katsayısı ne kadar yüksek olursa bu ölçekte bulunan maddelerin o ölçüde birbirleriyle tutarlı ve aynı özelliğin öğelerini yordayan maddelerden oluştuğu varsayılır. Ancak, alfa katsayısı güvenirlilik katsayısını vermez, yalnızca güvenirliliğin bundan çok daha yüksek olacağına işaret eder. Ölçeğin güvenirlilik katsayısı alfa katsayısından daha yüksektir⁽¹⁾. Likert tipi bir ölçekte yeterli olabilecek alfa katsayısı, olabildiğince 1'e yakın olmalıdır⁽¹⁴⁾.

Ayrıca; ölçek parçaları (madde) ile ölçeğin bütünü arasındaki ilişkinin bulunması yoluyla tutarlılığı bulmada madde-toplam ve madde-kalan korelasyon katsayıları teknikleri de kullanılır. Madde istatistikleri ya da madde güvenirliliği olarak bilinen bu yöntemde, her bir test maddesinin varyansı, toplam test puanının varyansı ile karşılaştırılarak arasındaki ilişkiye bakılır. Bu ilişki Pearson Momentler Çarpımı korelasyonunun düzeltilmiş formülü ile hesaplanır. Eğer ölçek yanıtları var-yok, 0-1 gibi kategorik iki seçenekli ise hesaplamada bi-serial; seçenekler sürekli derecelendirmeli ya da likert tipi ise point-bi-serial teknikleri uygulanır⁽¹⁸⁾.

Ölçekteki maddeler, eşit ağırlıkta ve bağımsız üniteler şeklinde ise, her madde ile toplam değerler arasındaki korelasyon katsayısının yüksek olması beklenir. Eğer bir maddenin toplam puanla olan korelasyonu düşük ise, bu o maddenin testteki diğer maddelerden farklı bir niteliği ölçtüğünü gösterir^(1,15).

Madde toplam puan korelasyon katsayısının hangi ölçütün altına düşünce güvenirliliğinin yetersiz sayılacağı konusunda belirli bir standart olmamakla birlikte, çoğu araştırmacı 0.20 alt seviyesini kullanmaktadır^(6,16). Ancak; katsayının 0.30'un üstünde⁽²⁰⁾, hatta 0.50'nin üstünde⁽¹⁵⁾ olması gerektiğini belirten araştırmacılar da bulunmaktadır.

Güvenirlilik Katsayısı Nasıl Yorumlanmalıdır?

Herhangi bir ölçme aracının güvenirliliğini belirleme çalışmasında sayılan ölçütlerden

hangisinin kullanılması gerektiğine kolayca verilen bir yanıt yoktur. Aynı durumda kullanılacak çeşitli güvenirlilik katsayıları olabilir. Araştırmacı; araştırma problemi, kullanılacak ölçek, toplanacak veri türü ve yanıtlardaki olası objektifliğe göre birden fazla, çeşitli güvenirlilik sınamalarını gerçekleştirmelidir.

Güvenirlilik, mutlaka görgül yollarla saptanır ve sayısal bir değerle ifade edilir. Bu sayısal değer, genellikle bir korelasyon katsayısıdır.

Korelasyon katsayısı iki değişken arasındaki ilişkinin derecesi ve yönü hakkında bilgi verir. Hesaplanan korelasyon katsayısı -1 ile +1 arasında değer alabilmektedir. İki değişken arasındaki korelasyon -1 ise değişkenler arasında ters yönde, +1 ise aynı yönde çok güçlü bir ilişki olduğunu göstermektedir. Korelasyon katsayısının 0 olması ise ilişki olmadığını belirtir⁽²¹⁾. Güvenirlilik katsayısı temelde bir korelasyon katsayısı olmasına karşın ondan farklı olarak sadece 0-1 arasında değer alabilir. Bu özellik güvenirlilik katsayısının istatistiksel tanımından kaynaklanmaktadır⁽³⁾.

Bir ölçmenin güvenirlilik katsayısı değerlendirilirken, bu değer olabildiğince yüksek olması arzu edilmektedir. Değer 1'e yaklaştıkça güvenirliliğin yüksek olduğu kabul edilir. Bir ölçme aracının güvenirlilik katsayısı 0.70 ise bu ölçeği yanıtlayanlar arasındaki değişkenliğin %70'inin ölçülen özellikle ilgili gerçek değişkenliğe, geri kalan %30'unun rastlantısal hatalara bağlı olduğunu gösterir. Güvenirlilik katsayısı değeri bireyin yaklaşık gerçek puanını, bu değer birden farkı ise, bireyin ölçme aracından aldığı hata puanını gösterir. Yani güvenirlilik katsayısı arttıkça ölçme aracının hata oranı azalmaktadır.

Güvenirlilik katsayısı; hangi düzeyde olduğunda, yeterli kabul edilebileceği sorusunun yanıtı, ölçme aracının yapısı ve hangi amaçla kullanılacağına bağlıdır. Çoğu araştırmacı 0.70 ve üzeri değerleri yeterli kabul etmekle birlikte^(3,16,22) 0.85-0.95 aralığını doyurucu seviye olarak kabul edenler de bulunmaktadır⁽²³⁾.

Ancak; bireylerin geleceği hakkında önemli kararlar verilmesine yönelik ölçmelerde gü-

venirliğin 0.95 ve üzerinde olmasının arzu edildiği belirtilmektedir^(1,20).

Fizyolojik ölçümlerde 0.90 ve üzeri, tutum ölçeklerinde 0.70 kabul edilebilir düzeylerdir. Gözlemciler arası uyum için korelasyon katsayısı en az 0.80 olmalıdır. Ayrıca, yeni geliştirilmiş ölçekler için test-tekrar test yönteminde korelasyon katsayısı 0.70 kabul edilebildiği halde, önceden geliştirilen ve çalışmalarda sınananlar için 0.80 olması gerektiği vurgulanmaktadır⁽¹⁴⁾. Ölçeklerde hesaplanan iç tutarlılık katsayısı içinse kabul edilebilen alt sınır 0.70'dir^(3,16).

Güvenirlik araştırmalarının sonucunun, basitçe "X testinin güvenilirliği 0.90'dır." şeklinde rapor etmek yerine "Samsun ili orta sınıf devlet okulu 8. sınıf öğrencilerinden oluşan 300 kişilik örneklem grubuna ölçme aracının A ve B eşdeğer formları ayrı günlerde uygulanmış ve iki form arasında korelasyon bulunmuştur. Paralel test güvenilirliği, %95 güven aralığı ile 0.90 olarak tahmin edilmiştir" ifadesi daha doğrudur⁽²⁴⁾.

Geçerlik (Validity)

Bir ölçme aracının güvenilirliğiyle birlikte önemli teknik özelliklerinden biri de geçerliğidir. Geçerlik; güvenilirliğe oranla çok daha karmaşık bir kavramdır ve bilimsel olduğu kadar felsefi bir sorundur.

Geçerliği; "bir ölçme aracının neyi ölçtüğü ve bu işi ne kadar iyi yaptığının değerlendirilmesi"⁽¹⁶⁾ şeklinde tanımlayabileceğimiz gibi, "bir ölçme aracının ölçülmek üzere hazırlandığı amacı, değişkeni ölçme derecesi" "bir ölçeğin "neyi", ne denli "isabetli/doğru" olarak ölçtüğüyle ilgili bir kavram"^(1,14,15,25), "test geliştiricinin, test puanlarından çıkarılacak vardamaların tiplerini desteklemek için kanıt toplama süreci"⁽²⁶⁾ ya da "test puanlarından ya da diğer değerlendirme biçimlerinden yapılan vardamaların uygunluğu"⁽³⁾ şeklinde de tanımlamak mümkündür.

Bu kavram, ölçme aracı puanlarına dayalı özel çıkarsamaların uygunluğuna, anlamlılığına ve kullanılabilirliğine işaret etmektedir⁽²⁴⁾.

Geçerlikle ilgili tanımlama ve açıklamaların

ortak özelliği; "ölçme aracının başka bir özelliği değil sadece ölçmek istediği özelliği ölçmesi" ve "ölçtüğü özelliği tam ve doğru bir biçimde ölçmesi" olduğudur.

Bir ölçme aracı başka yönleriyle ne kadar üstün yetenekli olursa olsun, örneğin ölçtüğü özelliği ne kadar az hata ile ölçmekte olursa olsun istenilen özelliği, başka özelliklerle karıştırmadan ölçen bir ölçme aracı olmadıkça işe yaramaz. Böyle bir ölçme aracını kullanmak bilimsel anlamda sakıncaları da beraberinde getirir.

Ayrıca; bir ölçme aracının geçerli sayılabilmesinin ilk koşulu güvenilirlik olmasına karşın, güvenilirlik hiçbir zaman geçerliği garantileyemez. Yüksek bir güvenilirlik katsayısı ölçme sonuçlarının tutarlılığını gösterir, ancak ölçme aracının geçerliğinin güvencesi değildir. Öte yandan güvenilir olsa da geçerli olmayan bir ölçme aracı, uygulamada önemli değildir⁽³⁾.

Bu nedenle bir ölçme aracının güvenilirliği ile geçerliği, üzerinde birlikte durulması gereken konulardır. Geçerlik bir ölçüm aracı için yapılması zorunlu olan ancak, ölçümün her zaman ve her durum için geçerli olduğunu söylemeyi engelleyen ve asla sonu olmayan bir süreçtir ve ölçeğin her kullanımında yeniden sınanması demektir^(14,15).

Ölçme Araçlarının Geçerliğini Sınama Yöntemleri;

Geleneksel olarak, geçerlik türleri, kapsamlı bağlantılı, ölçüle bağlantılı ve yapıyla bağlantılı geçerlik olarak sınıflandırılmaktadır⁽²⁴⁾.

Ancak geçerlik kategorilerinin adlarını kullanmak, değişik geçerlik türlerinin bulunduğunu ya da belirli bir geçerlik türünün, her özel çıkarsama ya da test kullanımı için en iyisi olduğunu göstermez. Bu kategoriler arasında kesin ayrımlar yapmak mümkün değildir. Örneğin ölçüt ya da kapsam geçerliği verileri aynı zamanda yapı geçerliğiyle ilişkilidir⁽²⁴⁾.

Bunlardan başka, belli hastalık ve fonksiyon bozukluklarının erken tanısı amaçlı tarama çalışmalarında seçicilik (specifity) ve duyarlılık (sensitivity) ile prediktif (yordayıcı) değer sınamaları önem kazanmaktadır. Bu sına-

malar, ileride bahsedilecek olan hemzaman/eşzaman geçerliliği içinde de anılabilmektedir⁽⁷⁾.

I. Kapsam Geçerliliği (Content validity):

Kapsam geçerliliği; ölçme aracı içindeki maddeler veya soruların ölçme aracının ölçmeyi amaçladığı konuları dengeli bir şekilde temsil etme derecesidir⁽¹⁶⁾. Bu kavram, ölçme aracının ölçüm amacı ile ilgisiz olan faktörlerin etkisinden arınık olmasını anlatır. İçerik, intrinsic ya da domain geçerlik olarak da anılır⁽⁷⁾. Örneğin, kaba motor yeteneğini ölçen bir araç, ince motor yeteneğini değerlendiren maddeleri içermemeli, ya da hastanın kaygı düzeyi vb.den etkilenmemelidir. Daha açık bir anlatımla, kapsam geçerliliğinin amacı; ölçüğün bütünü ve alt boyutlarının ölçülmek istenen alanı ölçüp ölçmediğini ve ölçülecek alan dışında farklı kavramları barındırıp barındırmadığının değerlendirilmesidir⁽⁷⁾.

Kapsam geçerliliği belirlemede, ölçme aracını geliştiren kişinin tümüyle kendisinin yapacağı değerlendirme yanıltıcı olur. Bu nedenle; geçerlik, alan uzmanlarının işbirliğini gerektirir ve uzmanlara danışılarak saptanır. Burada sözü edilen uzman kişi, ölçüğün hem hazırlandığı bilim alanını iyi bilen ve hem de ölçek sorusu hazırlama teknik ve yöntemlerini bilen kişilerdir. Ölçme aracı uzman eleştiri ve önerilerine göre tekrar şekillendirilir⁽¹⁶⁾.

Kapsam geçerliliği çalışmaları; uzman yargılarına dayanan, mantıksal ve görgül bazı işlemlerin de kullanılabileceği işlemler bütünüdür. Uzmanlar, ölçme aracını inceledikten sonra "Belirtke tablosu" adı verilen değerlendirme tablosu hazırlarlar. Uzmanların kararlarını özetleyen sayısal indekslere bakılarak yargıda bulunulur. Uzmanların çoğunluğunun aynı fikirde olması bir göstergedir^(3,16).

Kapsam geçerliliğini saptama, özellikle ölçek geliştirme çalışmalarında yapılması gereken bir aşamadır. Ancak herhangi bir dilde geliştirilen bir ölçme aracını Türkçe'ye uyarlamak isteyen araştırmacı da bu mantıksal ölçütü kullanmalıdır⁽⁷⁾. Ayrıca, başarı testleri ve personel seçiminde kullanılan meslek testleri için kap-

sam geçerliliğinin uygun olduğu belirtilmektedir⁽¹⁰⁾.

Akılda tutulması gereken bir diğer durum ise; bazı güvenirlilik katsayılarının, ölçme aracının kapsam geçerliliği hakkında önemli bir kanıt oluşturabileceğidir. Örneğin tek boyutlu bir ölçme aracının iç tutarlık katsayısının yüksek çıkması, kapsamdaki maddelerin aynı davranış alanını ölçtüğüne dolaylı bir kanıt oluşturur⁽³⁾.

II. Ölçüt Bağımlı Geçerlik (Criterion-related validity):

Ölçüt bağımlı geçerliliği; bir ölçme aracının diğer bir ölçme aracıyla elde edilen sonuçları verme yeteneği olarak tanımlanabilir⁽¹⁴⁾ ve bir ölçme aracının ölçmeyi amaçladığı özelliği ne kadar başarıyla yordadığı sorusuna yanıt verir⁽¹¹⁾. En objektif, en pratik geçerlik sınamasıdır⁽⁷⁾. Uygun korelasyon tekniğiyle bulunan ilişkinin derecesi "geçerlik katsayısı" olarak adlandırılır. Tüm geçerlik sınamaya yolları arasında tek geçerlik katsayısı veren yöntem ölçüt-bağımlı geçerlik yöntemidir. Ayrıca geçerlik katsayısının istatistiksel olarak test edilebilme olanağı da vardır⁽³⁾.

Eşzaman (benzer ölçekler, halihazır geçerlik) ve yordama geçerliliği olmak üzere iki türü vardır. Her ikisinde de, geliştirilen ölçme aracından bireylerin aldığı puanlar bir dış ölçütle karşılaştırılarak ilişki düzeyine bakılır. İkisi arasındaki fark; eşzaman geçerliğinde ölçek puanları o anda var olan bir ölçütle karşılaştırılır, oysa ki yordama geçerliğinde puanlar gelecekte gözlenecek, ölçülecek sonuçlarla karşılaştırılarak korelasyonu hesaplanır⁽¹⁶⁾. "Ahmet başarılı bir öğrenci mi?" sorusu hemzaman geçerliğine, "Ahmet'in başarı olasılığı nedir" sorusu ise yordama geçerliğine örnek olarak verilebilir⁽¹⁸⁾.

a) Kestirim (yordama) geçerliliği (Predictive validity)

Eğer bireylerin ölçek puanı, kestirimde bulunmak, ilerde ne olacağına karar vermek gibi bireylerin ilerdeki performansını yordamak için kullanılacaksa ya da ölçütle ilgili puanlar ilerde elde edilmek zorunda ise ölçüğün yordama geçerliğine bakmak gerekir⁽³⁾. Bu bir an-

lamda, yapılan ölçme ile ölçülmeye çalışılan şeyin gerçek hayattaki yansımalarının karşılaştırılmasındaki uyumu gösteren uygulama geçerliğidir⁽¹⁵⁾.

Örneğin; üniversite seçme sınavında alınan puanın bireylerin akademik başarısını yordama gücüne sahip olduğu düşünülüyorsa, seçme sınavının ölçütü bireylerin gösterdiği akademik başarı olacaktır. Seçme sınavı sonuçları ile bireylerin üniversiteye kabulü sonrası akademik başarılarının karşılaştırılıp, korelasyon teknikleri ile değerlendirilmesi bize seçme sınavının yordama geçerliği hakkında bilgi verecektir⁽³⁾.

Ölçüt-bağımlı geçerlikte geçerlik katsayısı; ölçütün uygunluğu ve zaman, örneklem büyüklüğü, ölçütün karıştırıcı değişkenlerle kontaminasyonu, doğal aşınma gibi faktörler etkileyebilmektedir⁽³⁾. Özellikle seçilecek ölçüt çok önemlidir ve sonucu direkt etkilemektedir. Bu nedenle geçerlik katsayısı, kullanılan ölçütün ışığında değerlendirilmeli ve ölçütün kolaylıkla saptanamadığı durumlarda bu geçerlik sınaması uygulanmamalıdır⁽³⁻⁷⁾. Bu durumda daha karmaşık olan yapı geçerliğini sınamak gerekir⁽⁷⁾. Ayrıca, bu geçerlik sınaması örneklem özelliklerine en çok bağımlı olan geçerlik sınaması yöntemidir. Örneklem seçimine dikkat edilmeli ve örneklem büyüklüğü 200'ün üzerinde olmalıdır⁽¹⁶⁾.

Yordama geçerliği genellikle eğitim ve işe yerleştirme için ölçme aracı seçiminde kullanılır^(3,16).

b) Hemzaman/eşzaman geçerliği (Concurrent validity)

Ölçme aracı uygulandığında, ölçme aracının ölçtüğü ilgili özelliğe ilişkin bir ölçüt varsa, aynı bireylerin ölçme puanlarıyla ölçüt puanları arasındaki ilişkinin derecesi ölçeğin ölçüt bağımlı geçerliğinin kanıtı olarak kabul edilir. Bu geçerlik türüne "zamandaş geçerlik" ya da "teşhisin kullanışlılığı" da denmektedir⁽³⁾.

Bu geçerlik sınamasında; bir ölçme aracından elde edilen sonuçlar ölçüm anında var olan bir diğer ölçütle karşılaştırılır. Benzer ölçek geçerliği olarak da bilinen bu yöntemde, daha önceden geçerliği saptanmış olan bir ölç-

me aracına (referans yöntem) gereksinim vardır. Yeni uyarlanan ölçme aracının geçerliğini bulmak için yeni ölçme aracı (ya da yöntem) ve geçerliği yüksek olduğu bilinen önceki ölçme aracı (ya da yöntem) birlikte aynı gruba uygulanır, bireylerin yeni ve eski ölçme aracından aldıkları sonuçlar arasındaki korelasyon hesaplanır ve bu korelasyon katsayısının yüksek olması beklenir⁽¹⁾. Bu yöntem; özellikle yeni geliştirilen tanı, tarama ya da laboratuvar yönteminin ne kadar doğru sonuca ulaştırdığı sorusuna yanıt aramak için yapılır. Burada iki kavram dikkati çeker. Bunlar duyarlılık (sensitivity) ve seçicilik (selectivity-specificity)'tir. Duyarlılığı; ölçüm aracının gerçekten hasta olanları, hasta olarak algılayabilme gücü, seçiciliği; ise gerçekten sağlam olanları, sağlam olarak algılayabilme gücü olarak tanımlamak olasıdır^(21,27).

III. Yapı Geçerliği (Construct validity):

Bazı karmaşık psikolojik özellikler, tek boyutta açıklanamaz, bu nedenle belli bir kuramsal yaklaşımla ve kavramsal çerçevede birer yapı olarak tanımlanırlar. Zeka, sosyal uyum gibi doğrudan gözlenemeyen soyut kavramlarda; ölçme aracı puanlarıyla, ilişkili olduğu yapının ortaya konması ya da ölçek puanlarının ilişkili olduğu düşünülen yapının varlığına ilişkin kanıtları, ölçme aracının ölçme amacı doğrultusunda çalıştığını göstermek açısından önemlidir. Bu nedenle, yapı geçerliği çalışması, bilim adamının öngördüğü bu yapıyı ölçmek için geliştirdiği ölçme aracının, gerçekten öngörülen yapıyı ölçüp ölçmediğini anlamak için yapılır. Başka bir ifade ile yapı geçerliği, bir ölçme aracının ve ondan elde edilen puanın ne anlama geldiğini araştırma sürecidir. Messick ise yapı geçerliğini; "yapı geçerliği, geçerliğin tümü olmayabilir, fakat şüphesiz geçerliğin yüreğidir" şeklinde ifade etmiştir^(3,16).

Yapı geçerliği çalışmaları, ölçme aracının ölçtüğü faktörler incelenerek ya da geçerliği araştırılan ölçme aracının diğer ölçek ve ölçülerle olan ilişkisini araştırarak gerçekleştirilir. Her seferinde ölçme aracıyla ilgili yeni bir par-

ça bilgi elde edilerek, yığılmalı bir şekilde ölçüğün yapısı ve puanın anlamı hakkında bilgiler elde edilir⁽¹⁾.

Bir ölçme aracının yapı geçerliğini sınamada en çok kullanılan iki yaklaşım; faktör analizi ve bilinen grup ile karşılaştırmadır⁽¹⁵⁾. Bunların dışında diğer yapı geçerliği değerlendirme yolları hipotez sınaması (mantıksal analiz) ve çok değişkenli-çok yöntemli matriks'dir⁽¹⁴⁾.

Faktör analizi; ölçme aracındaki maddelerin farklı faktörler altında toplanıp-toplanamayacağını değerlendirmek üzere yapılan bir işlemdir⁽¹⁵⁾. Ölçülen yapıda birbiriyle yüksek korelasyon gösteren maddeler birer faktör altında kümelendirilir⁽³⁾. Faktör analizi, **açıklayıcı (exploratory)** ya da **doğrulayıcı/hipotez destekleyici (confirmatory)** olabilir. Açıklayıcı faktör analizinde; bilinmeyen bir kuramsal yapıyı ölçmek için oluşturulan ölçme aracından elde edilen sonuçlara dayanarak, söz konusu yapının nasıl olduğu açıklanmaya çalışılır. Doğrulayıcı faktör analizinde ise; varolan bir kuramsal yapıya dayanarak geliştirilen ölçme aracından elde edilen verilere dayanarak söz konusu kuramsal yapının doğrulanıp doğrulanmadığı test edilmeye çalışılır⁽³⁾. Ölçek uyarlamalarında daha çok, ölçekteki maddelerin yapısı hakkında var olan bir hipotezi sınadığı için doğrulayıcı faktör analizi kullanılır. Son yıllarda bu teknikler, yeni kuramlar geliştirmede ve varolan kuramların geçerliğini test etmede yaygın olarak kullanılmaktadır.

Bilinen grup karşılaştırmasında; araştırmacı, yapının dayandığı kurama bağlı olarak, bir testten anlamlı olarak farklı puan alabilecek grupları belirler, ölçme aracını her iki gruba uygular ve gruplar arası farka bakar. Karşıt gruplar yöntemi adı da verilen bu yöntemde, aynı testin karşıt iki grup arasındaki farkın yapı geçerliğinin dolaylı kanıtı olarak kabul edilir. Örneğin, test şizofreniyi ölçmek için geliştirilmişse, aynı testin şizofren olmayanlara uygulanması durumunda iki grup arasındaki fark yapının dolaylı kanıtı olur. Ancak, bu işlem çoğu durumda yeterli görülmez.

Hipotez sınaması/mantıksal (logical) analizde; ölçme aracının, ilgili yapıya ilişkin gelişt-

rilmiş diğer ölçüm araçlarıyla ilişkisine bakılır. Bilinen grup karşılaştırmasına benzer bir yöntemdir. Burada araştırmacı, ilgili kaynaklar doğrultusunda, önceden aralarında ilişki olacağı varsayımını kurduğu ilişkilerin yönünü ve düzeyini korelasyon analiziyle değerlendirerek test eder⁽¹⁴⁾. Ölçüt bağımlı geçerlikten farkı, korelasyonun daha düşük olmasının beklenmesidir⁽³⁾.

Çok özellikli-çok yöntemli matriks: Bir ölçme aracının ölçümlerinin aynı ya da farklı yapılardaki diğer testler ile nasıl ilişkili olduğunun değerlendirilmesi yoluyla da yapı geçerliği sınanabilir. Bu yöntemde, bir testin ölçtüğü yanında ölçmediği değişkenleri belirlemek de önemlidir. Ölçümlerin her birinden alınan puanlar bir korelasyon matriksinde gösterilir. Katılımcıların bir oturumda dört ya da fazla testi doldurmaya gönüllü olması gerekir, bu nedenle güç bir geçerlik sınamaya yöntemidir^(7,14).

Geçerlik Katsayısı Nasıl Yorumlanmalıdır?

Geçerlik katsayısı, klasik yollarla elde edildiğinde ölçme aracı ile ölçüt arasındaki korelasyondur. Bu korelasyon katsayısı, test ya da ölçütün sürekli ya da süreksiz değişken olmasına göre değişik biçimlerde hesaplanabilir. Geçerlik ya hep ya da hiç biçiminde olmaz^(3,28). Bir ölçme aracı geçerli ya da geçersiz değil, az ya da çok geçerli olabilir⁽³⁾. Çünkü geçerlik katsayısının ne kadar olması gerektiğinin nesnel ve kesin bir yanıtı yoktur. Geçerlik katsayısının değeri onun sadece sayısal büyüklüğüne göre yargılanmaz; bu değer, testin kullanılacağı durumlara göre farklılaşır⁽²⁸⁾. Test geçerliğini kanıtlamada karşılaşılan güçlüklerin çoğu, ölçülecek değişkenin açık-seçik ve genellikle kabul edilebilir bir tanımı olmayışı ile ölçütün yapısından kaynaklanır. Bu nedenle her geçerlik katsayısı, kullanılan ölçütün ışığında değerlendirilmelidir. Ölçütün güvenilir olmamasının geçerlik katsayısını düşürmesinin yanı sıra, ölçütün hangi değişkenlerden oluştuğu ve bu değişkenlerle testin ölçtüğü değişkenler arasında bir ilişkinin olup olmaması da önem taşır.

Geçerlik sonucunun rapor edilmesinde "bu testin geçerliği" şeklinde ayrıntı belirtmeyen

cümlecikler kullanmak uygun değildir. Hiçbir ölçme aracı, bütün amaçlar ya da durumlar için geçerli değildir. Bir ölçme aracının, belirli kararlar için yanlış kullanılma olasılığı varsa bu tür kullanımlar için özel uyarılar belirtilmelidir⁽²⁴⁾.

Sonuç olarak;

“Güvenilir bir ölçme aracı geçerli olabilir ya da olmaz ancak güvenilir olmayan bir ölçme aracı hiçbir zaman geçerli olamaz”^(3,4,14,16,28) ilkesi ve “ölçme araçları bilimin özü gereği asla değişmez-sorgulanamaz değildirler”⁽⁴⁾ gerçeği unutulmamalıdır. Aradan geçen zamanda ölçme aracının güvenilirliği ve geçerliliği değişebileceği için ölçme aracını kullananlar, ölçme aracı ne kadar geçerli ve güvenilir olursa olsun değişik zaman ve gruplarda ölçme aracının geçerliliğini ve güvenilirliğini tekrar irdelemelidir⁽³⁾.

Geliş Tarihi : 07.12.2004

Yayına kabul tarihi : 11.08.2006

Yazışma adresi:

Dr. Servet AKER

Ondokuz Mayıs Üniversitesi, Tıp Fakültesi,

Halk Sağlığı Anabilim Dalı,

55139 Kurupelit, SAMSUN

KAYNAKLAR

- Özgüven İE. Psikolojik Testler. (1.Baskı). Ankara, Psikolojik Danışmanlık ve Rehberlik Eğitim Merkezi Yayınları, 1994.
- Arcı H. Psikometri Üzerine (Söyleşi: Tavat B.). 3P Dergisi 1994; 2: 78-81.
- Erkuş A. Psikometri Üzerine Yazılar. (1. Baskı). Ankara, Türk Psikologlar Derneği Yayınları, 2003.
- Erkuş A. Sık Kullanılan Bazı Psikolojik Ölçeklerin Güvenirliklerinin İrdelenmesi. Türk Psikoloji Yazıları 2000; 3: 3-17.
- Öner N. Güvenirliği ve/veya Geçerliliği Sınanmış Psikolojik Testler. Türk Psikoloji Dergisi: Özel Sayı 1994; 9: 9-18.
- Özdamar K. Paket Programlar ile İstatistiksel Veri Analizi. (2. Baskı). Eskişehir, Kaan Yayınları, 1999.
- Gözüm S, Aksayan S. Kültürlerarası Ölçek Uyarlaması İçin Rehber. Hemşirelikte Araştırma Geliştirme Dergisi 2003; 5(1): 3-14.
- Turgut FM. Eğitimde Ölçme ve Değerlendirme Metotları. (8. Baskı). Ankara, Saydam Matbaacılık 1992; 3.
- Tekin H. Eğitimde Ölçme ve Değerlendirme. (14. Baskı). Ankara, Yargı Yayınları, 2000; 57.
- Crocker L, Algina J. Introduction to Classical and Modern Test Theory. Fort Worth: Holt, Rinehart and Winston Inc, New York 1986.
- Anastasi A. Psychological Testing. New York, Mac Millan Pub. Co. Inc., 1976.
- Ghiseli EE, Campbell JP, Zedeck S. Measurement Theory for the Behavioral Sciences. San Francisco, WH Freeman and Co, 1981.
- Magnusson D. Test Theory. Massachusetts, Addison-Wesley Pub.Co. Reading, 1967.
- Peirce AG. Measurement Instruments, Principles and Practice of Nursing Research. St. Louis, Mosby-Year Book Inc., 1995.
- Karasar N. Bilimsel Araştırma Yöntemi. (7. Baskı). Ankara, Sim Yayınları, 1995.
- Tavşancıl E. Tutumların Ölçülmesi ve SPSS ile Veri Analizi. (1. Baskı). Ankara, Nobel Yayınları, 2002.
- Hayran M, Özdemir O. Bilgisayar İstatistik ve Tıp. Ankara, Hekimler Yayın Birliği Yayınları, 1995; 372.
- Öner N. Türkiye’de Kullanılan Psikolojik Testler: Bir Başvuru Kaynağı. (3. Baskı). İstanbul, Boğaziçi Üniversitesi Yayınları, 1997.
- Erkuş A. İstatistik Paket Programlarını Doğru Kullanabiliyor muyuz?. Türk Psikoloji Bülteni, 1999; (12): 14-17.
- Öner N. Kültürlerarası Ölçek Uyarlamasında Bir Yöntembilim Modeli. Türk Psikoloji Dergisi, 1987; 6: 80-83.
- Aksakoğlu G. Sağlıkta Araştırma Teknikleri ve Analiz Yöntemleri. (1. Baskı). İzmir, Dokuz Eylül Üniversitesi Yayınları, 2001.
- Bruning JL, Kintz BL. İstatistik. (Çev. Dönmez, A). (1. Baskı). Ankara, Gündoğan Yayınları, 1993; 279.
- Polit DF, Hungler B, Bernadette P. Essentials of Nursing Research: Methods, Appraisal and Utilization, (4th ed.) Philadelphia. Lippincott Williams & Wilkins 1997.
- American Educational Research Association, The American Psychological and the National Council on Measurement in Education. Eğitimde ve Psikolojide Ölçme Standartları. (Çev. Hovardaoğlu, S. &Sezgin, N.), Ankara, ÖSYM-Türk Psikologlar Derneği Yayınları, 1998; 11-30.
- Tezbaşaran A. Likert Tipi Ölçek Geliştirme. (1. Baskı). Ankara, Türk Psikologlar Derneği Yayınları, 1996.
- Cronbach LJ. Essentials of Psychological Testing. (3rd ed.) New York, Harper and Row Pub., 1970.
- Dawson B, Trapp R. Basic & Clinical Biostatistics, (3rd ed). Singapore: Lange Medical Books/McGraw-Hill 2001: 262-281.
- Ercan İ, Kan i. Ölçeklerde Güvenirlik ve Geçerlik. Uludağ Üniversitesi Tıp Fakültesi Dergisi, 2004; 30: 211-216.