

DERLEME

Ölçeklerde Güvenirlik ve Geçerlik

İlker ERCAN, İsmet KAN

Uludağ Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Bursa.

ÖZET

Somut veya soyut özelliklerin ölçülmesinde kullanılacak ölçeğin standart bir ölçme aracı olması gerekmektedir. Ölçeğin standardize olabilmesi ve sonrasında uygun bilgiler üretme yeteneğine sahip olması için ölçüm değerlerinin kararlılığının bir göstergesi olan "güvenirlik" ve ölçmeyi amaçladığı özelliği doğru ölçebilme derecesinin göstergesi olan "geçerlik" olarak nitelendirilen iki temel özelliğe sahip olması istenir. Somut özelliklerin ölçülmesinde güvenilir ve geçerli bir ölçekten yararlanılması bir sorun olmamakla birlikte soyut özelliklerin ölçümünde bu bir sorun haline gelmektedir.

Anahtar Kelimeler: Ölçek. Güvenirlik. Geçerlik.

Reliability and Validity in The Scales

ABSTRACT

The scale that will be used in order to measure concrete and abstract characteristics must be a standard measurement tool. The scale must have two basic properties to become a standard scale and then in order to have capability of producing suitable knowledge; reliability, indicator of consistency of measurement values and validity, indicator of the degree of correctly measuring the characteristic that aims to measure. It doesn't matter the scale's being reliable and valid in case of measuring the concrete characteristics, but it is an important problem in case of measuring the abstract characteristics.

Key Words: Scale. Reliability. Validity.

Ölçme, "İstatistik birimlerinin ilgilenilen özelliğe sahip olma derecesinin, belirli kurallara uyarak, sembolle ve özellikle sayı ile eşleştirilmesidir"^{1,2}. Üzerinde çalıştığımız özelliği ölçebiliyor ve sayısal olarak ifade edebiliyorsak, onları bilimsel bir şekilde değerlendirebilir ve açıklayabiliriz. Ölçülemeyen ya da sayısal olarak ifade edilemeyen özellik için ancak betimleme yapılabilir^{3,4}.

İstatistik birimlerinin somut özelliklerine belirli kurallara doğrultusunda sayılar ve semboller atanmasındaki netlik, olayların veya bireylerin soyut özelliklerine sayılar ve semboller atanmasındaki netlikle eşit şartlarda değildir. Bu nedenle sosyal bilimlerde, fen ve sağlık bilimlerine göre ölçme yapmak daha güç olmaktadır^{1,3,5}.

Bilimsel araştırmalar ancak veriler aracılığıyla yapılabilir. Verilerin, değişkenin standart ölçme teknikleri ile saptanan sayılardan oluşmasına özen göstermek

gerekir. Değişkenin boyutunu ölçmede kullanılacak ölçme aracının standart bir ölçme aracı olması verilerin istatistiksel özelliklere sahip veriler olmasını sağlayacaktır⁶. Ölçeğin standart bir ölçme aracı olabilmesi için de bazı özellikler taşınması gerekir.

Ölçeğin Taşınması Gereken Özellikler

Ölçme, belirli bir amaç için yapılır. Amaç, ölçme konusu olan özellik bakımından bireyler, olaylar yada nesnelere hakkında değerlendirme yapmak ve elde edilen değerlendirme sonuçlarına dayanarak belli kararlar vermektir. Verilen kararların doğruluğu ve uygunluğu kararların dayandığı değerlendirme sonuçlarına, dolayısıyla değerlendirmede kullanılacak olan ölçüm sonuçlarına ve ölçütün uygun olmasına bağlıdır. Bunun içinde ölçü aracının standardize olması istenir.

Ölçek kalitesi standardize edilene kadar, maddeleri analiz edilir ve tekrar gözden geçirilir. Standardize edilen ölçeğin yönetimi, puanlaması ve yorumlaması dikkatli bir şekilde açıkça belirtilmelidir. Bu şekilde standardize edilen ölçeklere objektif (nesnel) ölçekler denir⁷.

Ölçeğin standardize olabilmesi ve sonrasında uygun bilgiler üretme yeteneğine sahip olması için "güvenirlik" ve "geçerlik" olarak nitelendirilen iki özelliğe sahip olması istenir.

Geliş Tarihi: 28.06.2004

Kabul Tarihi: 26.10.2004

Dr. İlker ERCAN
Uludağ Üniversitesi Tıp Fakültesi,
Biyoistatistik Anabilim Dalı
Görükle Nilüfer/BURSA
Tel: 0 224 442 82 00 (21028)
GSM: 0 505 378 19 03
e-mail: ercan@uludag.edu.tr

1. Güvenirlilik

Ölçeğin taşınması gereken özelliklerden birisi olan güvenirlilik, bir ölçme aracıyla aynı koşullarda tekrarlanan ölçümlerde elde edilen ölçüm değerlerinin kararlılığının bir göstergesidir^{1,7,8,9}. Ölçümlerin kararlılığının yorumlanmasında, iki farklı yaklaşım vardır. Birinci yaklaşım, birbirini izleyen ölçmelerde, bireyin grup içindeki sırasının değişmezliği; ikinci yaklaşım ise birbirini izleyen ölçümlerde ölçme hatalarının büyüklüğü, yani ölçeğin tekrarlı ölçümlerde aynı denekten yaklaşık olarak aynı ölçüm değerini elde etmesi ve dolayısıyla ölçmenin standart hatasının düşük olması ile ilgilidir^{10,11}.

Güvenirlilik sadece ölçme aracına ait bir özellik değildir, ölçme aracı ve aracın sonuçlarına ilişkin bir özelliktir⁹.

Ölçekle sağlanan bilgilerin kararlı özellik taşıdığına, yani hatadan arındırılmış olduğuna ve aynı amaçla yapılacak ikinci bir ölçümde aynı sonuçların elde edileceğine güven duyulması gerekir^{1,12}. Güvenilir olmayan bir ölçek kullanışsızdır⁷.

Somut özelliklerin ölçülmesinde kullanılan ölçeklerle yapılan tekrarlı ölçümlerde her seferinde birbirine çok yakın sonuçlar alınır. Soyut özelliklerin ölçülmesinde kullanılan ölçeklerde ise aynı sonuçların alınması çok zordur. Dolayısıyla somut özelliklerin ölçülmesinde kullanılan ölçekler daha güvenilirdir⁴.

Birimlerin soyut özelliklerinin ölçülmesinde kullanılan ölçekler diğerlerinden önemli bazı ayrılıklar göstermektedir. Bu tür ölçekler, birimleri ölçmek üzere, belirlenmiş olan standart aralıkları birbirine eşit olsa bile başlangıç noktaları keyfi olması nedeniyle miktar anlamı olan ölçümler verememektedir. Bu nedenle, ölçeğin ölçtüğü özelliğin gerçek değerine yakın ölçümler verme gücünün dikkatle incelenmesine ve dolayısıyla bu incelemede analitik olmaya yani bir ölçme aracıyla elde edilen ölçümlerdeki değişkenliği aşağıdaki şekilde ayırmaya gerek vardır¹³:

- i. ölçülmek istenen özellikteki gerçek farklardan kaynaklanan değişkenlik
- ii. ölçülmek istenmeyen özelliklerin, ölçülmek istenen özelliği etkilemesi sonucunda gerçek farklarından kaynaklanan değişkenlik,
- iii. i ve ii'nin etkileşimden kaynaklanan değişkenlik
- iv. Ölçme hatasından kaynaklanan değişkenlik

Güvenilir bir ölçek hatasız ölçme yapan ölçektir. Hatasız bir ölçme olması mümkün olmadığından, ölçmenin güvenirliliğini arttırmak hatayı en aza indirmekle mümkün olabilecektir. Hatayı en aza indirmenin temel şartı, hata kaynaklarını belirleyip kontrol altına almaya çalışmaktır^{7,9,14}. Hatanın kaynağı ölçme aracından veya ölçme aracı dışındaki unsurlardan kaynaklanabilir. Ölçme aracıyla ilgili hata kaynakla-

rını azaltmak için ölçme aracının bazı yöntemlerle güvenirliliğinin araştırılması gerekir¹⁵.

Ölçme sonucunu tamamıyla hatalı kabul etmek mümkün olmadığı gibi, tamamıyla hatasız kabul etmek de mümkün değildir. Doğa bilimlerinde bile somut ve kararlı özelliklerin ölçülmesinde bir miktar hata vardır. O halde herhangi bir ölçme sürecinden elde edilen ölçüm değerinde bir miktar gerçek ve bir miktar da ölçme hatası vardır. Bu durum ölçümün gerçek ve hata olmak üzere iki öğeden oluştuğunu göstermektedir^{1,9,13}.

Gözlenen ölçüm değerinde hatanın katkısının artması, ölçülmek istenen özellik bakımından birimler arası gerçek farklardan gelmekte olan değişkenliğin azalmasına yol açacak ve güvenirliliğin azalmasına neden olacaktır¹³. Gözlenen ölçüm değerindeki hata katkısının azalması ise ölçeğin birimler arasındaki bazı gerçek farklılıkları yansıtmamasını sağlayacak ve güvenirliliğin artmasını sağlayacaktır^{9,16}.

Gözlenen ölçüm değerini (x_i), iki öğesi olan gerçek değer (t_i) ve ölçüm hatası (e_i) unsurlarını da dikkate alarak aşağıdaki şekilde ifade edebiliriz^{7,17,18}.

$$x_i = t_i + e_i \quad (1.1)$$

Uygulamada, sadece gözlem değeri (x_i) olan ölçüm değeri bilinebilir. Gerçek değer (t_i) ve hata değeri (e_i) hakkında bir bilgi elde edilemez. Hataların tesadüfi olarak dağıldığı varsayımıyla, hatalar tesadüfi olarak pozitif ve negatif yönde gelişir ve birbirlerini elemesi sayesinde ölçüm hataları ortalaması sıfır olur ($E(e_i) = 0$). Bu varsayımdan hareketle ölçmenin varyansı şöyle formüle edilebilir^{7,17}.

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2 \quad (1.2)$$

σ_x^2 : Gözlenen değerlerin varyansı,

σ_t^2 : Gerçek değerlerin varyansı,

σ_e^2 : Ölçüm hatalarının varyansı.

Bu bilgilerden faydalanarak bir ölçme aracının ne derece güvenilir olduğunu, gerçek değerlerin varyansını gözlenen değerler varyansına oranlayarak elde edeceğimiz güvenirlilik katsayısı ile anlayabiliriz¹⁷⁻¹⁹.

$$\rho_x = \frac{\sigma_t^2}{\sigma_x^2} \quad (1.3)$$

Güvenirlilik katsayısı, hatanın değil ölçmedeki hatasızlığın bir ölçüsüdür ve bireysel bir istatistik değil, bir grup istatistiğidir⁹.

İlgili özelliklerin dışında var olan etmenler nedeniyle ortaya çıkan, ölçüm değerlerindeki değişkenlik olan, ölçmenin standart hatasını son iki denklemden yararlanarak elde edebiliriz^{9,14,18,19}.

Ölçeklerde Güvenirlik ve Geçerlilik

$$\sigma_t^2 = \sigma_x^2 \rho_x \quad \sigma_e: \text{Ölçmenin standart hatası}$$

$$\sigma_e^2 = \sigma_x^2 - \sigma_t^2 \quad \sigma_x: \text{Gözlenen değerlerin standart sapması}$$

$$\sigma_e^2 = \sigma_x^2 - \sigma_x^2 \rho_x \quad \sigma_t: \text{Gerçek değerlerin standart sapması}$$

$$\sigma_e^2 = \sigma_x^2 (1 - \rho_x) \quad \rho_x: \text{Güvenirlik katsayısı}$$

$$\sigma_e = \sigma_x \sqrt{1 - \rho_x} \quad (1.4)$$

Gözlenen değerlerin standart sapması arttıkça veya güvenilirlik azaldıkça ölçmenin standart hatasında artış olacaktır.

σ_t^2 / σ_x^2 oranıyla güvenilirlik katsayısının hesabı, gerçek değerlerin uygulamada bilinmesinin mümkün olmaması nedeniyle teoride kalmaktadır. Bu nedenle güvenilirlik katsayısını dolaylı yoldan, farklı durumlara yönelik olarak, hesaplamak için yöntemler geliştirilmiştir^{7,17}.

1.1. Norm-Referans Güvenirliği (Norm-Referenced Test)

Ölçeğin güvenirligi farklı yollarla incelenebilir. Ölçeğin güvenirligi bir ölçek bir kez uygulanarak, bir ölçek iki kez uygulanarak veya iki eşdeğer ölçek bir kez uygulanarak incelenebilir. Bir ölçeğin bir kez uygulanması durumunda iç tutarlılık güvenirliligi incelenir. Güvenirlik katsayısı sayısal olarak 0 ile 1 arasında değişmektedir.

1.1.1. Formun Tekrarı Yöntemi (Test-Retest Method)

Formun tekrarı yöntemi, bir ölçme aracının aynı denek grubuna aynı koşullarda, önemli derecede hatırlamaları önleyecek kadar uzun, fakat ölçülecek özellikte önemli değişimler olmasına izin vermeyecek kadar kısa bir zaman aralığında iki kez uygulanmasıdır¹³. İki uygulamadan elde edilen ölçüm değerleri korelasyon katsayısı ölçeğin güvenirlilik katsayısıdır^{9,14,20}.

1.1.2. Eşdeğer (Paralel) Formlar Yöntemi (Parallel-Forms Method, Equivalent-Forms Method, Alternative-Form Method)

Paralel formlar yönteminde aynı davranış kalıplarını temsil edebilecek farklı maddeler örneklenecek iki eşdeğer form oluşturulmaktadır⁷. İki formun eşdeğer olabilmesi için, formların kapsamalarının, yapısının, zorluk derecesinin, talimatların, puanlamanın, madde sayısının ve yorumlamanın aynı olması gerekir^{7,9}. Eşdeğer iki form aralıksız olarak aynı anda ya da aralıklı olarak farklı iki zamanda uygulanır. Formlar arasındaki korelasyon hesaplanır ve güvenirlilik katsayısı olarak yorumlanır⁹. Eşdeğer formların uygulanmasında, aradaki zaman aralığının artması kararlılığı

olumsuz yönde etkileyecekse, formlar deneklerin sıkılmalarını ve yorulmalarını engelleyecek kadar ara süre verilerek ard arda uygulanmalıdır^{11,14}.

1.1.3. İç Tutarlılık Yöntemleri (Methods of Interval Consistency)

Bir ölçeğin bir kez uygulanmasıyla güvenirlilik tahmini yapılıyorsa, diğer güvenirlilik tahmini yöntemlerine göre, güvenirlilik tahmininde meydana gelebilecek hata daha az olacaktır.

1.1.3.1. Yarıya Bölme Yöntemi (Split-Half Method)

Yöntem, formu iki eş parçaya bölerek, iki yarının deneklere aynı anda uygulanması sonrası, deneklerin yarılarından aldıkları puanlar arasındaki korelasyon ile güvenirlilik tahmini yapılmasını sağlar^{1,12,19}.

1.1.3.2. Kuder-Richardson Güvenirlilik Katsayıları

İçsel tutarlılığın güvenirliliğini belirlemek için en sık kullanılan yöntemlerden birisi de Kuder-Richardson yaklaşımıdır^{14,17}. Yöntem tüm maddelerin birbirleriyle ve ölçeğin tamamıyla iç tutarlılığını tahmin etme amacı üzerine kuruludur⁷. Bu nedenle yöntem, ölçekteki tüm maddelerin aynı değişkeni ölçtüğü varsayımına dayanır^{9,10}. Yöntemin uygulanmasında veri seti, ölçekteki maddelerden alınan cevaplar istenilen özelliği taşıyorsa "1" puan, istenilen özelliği taşıyamıyorsa veya boş bırakılmışsa "0" puan verilerek oluşturulur. Bu yöntemle, iç tutarlılığa yönelik güvenirlilik kestiriminde bulunmada belirli kriterler dikkate alınarak Kuder-Richardson 20 veya 21 formüllerinden uygun olanı kullanılır.

1.1.3.3. Cronbach Alfa Güvenirlilik Katsayısı

Cronbach (1951) tarafından geliştirilen alfa katsayısı yöntemi, maddeler doğru-yanlış olacak şekilde puanlanmadığında, 1-3, 1-4, 1-5 gibi puanlandığında, kullanılması uygun olan bir iç tutarlılık tahmin yöntemidir⁹. Cronbach alfa katsayısı, ölçekte yer alan k maddenin varyansları toplamının genel varyansa oranlanması ile bulunan bir ağırlıklı standart değişim ortalamasıdır⁶.

1.1.3.4. Teta Güvenirlilik Katsayısı

Temel bileşenler analizi üzerinde temellenen bir güvenirlilik katsayısıdır¹.

1.1.3.5. Omega Güvenirlilik Katsayısı

Doğrusal bağıntılar için güvenirlilik tahminlerinde kullanılacak diğer bir güvenirlilik katsayısı Heise ve Bohrnstedt (1970) tarafından önerilen omega katsayısıdır. Omega katsayısı faktör analiz modeli üzerinde temellenmiştir¹.

1.1.3.6. Guttman Güvenirlilik Katsayıları

Louis Guttman (1945) güvenirlilik katsayısının alt sınırlarını veren altı katsayı tanımlamıştır. Guttman, formun kısımları arasındaki doğal ilişki hakkında varsayımda bulunmadan, katsayıların güvenirliliğin alt sınırını verdiğini ispatlamıştır¹⁸.

1.2. Kriter-Referans Güvenirliği-Livingston Formülü

Kriter-referans güvenirligi (criterion-referenced test), önceden belirlenmiş bir kritere göre hesaplanır. Bu durum her zaman belli bir kriterin bulunamaması nedeniyle kısıtlayıcı bir unsur olmaktadır²¹.

2. Geçerlik

Uygun bir yöntemle, ölçme aracının güvenirligi saptansa bile, güvenirligin ölçme aracının kararlılığı ile ilgili olmasından dolayı, yapılan işlem "Kullanılan ölçüm aracıyla neyi ölçmek istiyoruz?", "Maddelerimiz, amaç doğrultusunda ölçmek istediğimizi doğru olarak ölçebilir mi?" sorularına cevap veremez^{4,12}. Bu nedenle, davranışsal özellikleri ve bunlardan da özellikle bilişsel ve duyuşsal yönü baskın olanları ölçerken kullanılan ölçme aracının, ölçmek istediğimiz özelliğe yönelik ölçme dereceleri araştırılmalıdır¹³. Bu araştırma, ölçme aracının geçerliğı ile ilgilidir ve şu gerekçeye dayanır: Doğrudan ölçme yapılabilen somut özelliklerde, amaca hizmet gücü yüksek olan ölçme araçlarıyla çalışılmaktadır^{12,13}. Ölçülmek istenen özellik soyutlaştıkça amaca hizmet gücü olmayan yada bu gücü sınırlı olan ölçeklerle çalışma tehlikesi artar¹³. Ölçümlerin amaca hizmet edebilmesi, ölçme aracının ölçülmek istenen değişkeninin ölçüsü olabilecek ölçümler vermesine bağlıdır⁹.

Yukarıdaki ifadeler sonucunda geçerliğin tanımını şöyle yapabiliriz: Geçerlik, bir ölçme aracının ölçmeyi amaçladığı özelliğı, başka herhangi bir özellikle karıştırmadan, doğru ölçebilme derecesidir¹⁰. Geçerliğin davranışsal özelliklerin ölçülmesinde önemli bir yeri vardır. Davranışsal özelliklerin ölçülmesinde kullanılan ölçeklerin hiç birinin geçerliğı tam değildir. Fakat, ölçeklerin geçerliğı yeterli sayılabilecek doğrulukta ölçümler verecek düzeye getirilebilir¹³.

Bir ölçme aracının geçerliğı, standart koşullarda o ölçme aracı ile elde edilecek ölçümlerdeki değişkenliğin ne kadarının, incelenen bireylerin ölçülen özelliğe sahip oluş dereceleri arasındaki gerçek farklardan gelmekte olduğunu gösterir. Bir ölçme aracının geçerliğinin belirlenmesinde bu düşünceden yararlanır. Ölçülen özelliğe sahip oluş derecesi bakımından bireyler arası gerçek farklardan meydana gelen değişkenliğin, ölçümlerde meydana gelen toplam değişkenliğe oranı hesaplanır¹³.

Bir ölçme aracı ile standart koşullarda elde edilmiş ölçümlerde gözlenen toplam değişkenliğin, kişilerin ilgilenilen özelliğe sahip oluş dereceleri arasındaki gerçek farklardan gelen değişkenlikten daha büyük olması beklenir. Bunun tersi olanaksızdır. Bireylerin bir özelliğe sahip oluş dereceleri arasındaki gerçek farklardan gelen değişkenlik, olsa olsa, aynı ölçümlerde gözlenen toplam değişkenlik kadar olabilir. Bu son durumda kullanılan ölçeğin geçerliğı tam demektir ki, bu bir idealdir. Ölçümler-

deki değişkenliğin tümünün bireylerin ilgili özelliğe sahip oluş dereceleri arasındaki gerçek farklardan gelmekte olması demek, ölçeğin başka hiçbir özelliğin etkisinden etkilenmemiş olması ve ölçme hatasının tam olarak sıfıra indirgenmiş bulunması demektir. Ölçme hatasının tam olarak sıfıra indirgenmesi imkansızdır. Çünkü, istenmeyen özellikler ölçümleri hiç etkilememiş olsalar bile, ölçümlerde gözlenen değişkenlik ölçme hatası yüzünden artacak ve bu değişkenliğin miktarı, gerçek farklardan gelen değişkenlikten daha az olacaktır. Ölçümlerde gözlenen toplam değişkenlik en azından bu iki kaynaktan gelen katkıyı da kapsamak durumundadır. Bunlar sonucunda, mümkün ve olası etkileşimler dikkate alındığında davranışsal özelliklerin ölçülmesinde kullanılan araçların neden tam geçerli olamayacağı daha iyi anlaşılır¹³.

Bir ölçeğin geçerliğı, bir derece problemdir, yani ölçeğin geçerliğı vardır veya yoktur şeklinde düşünülemez^{7,9}. Bu nedenle ölçeğin geçerliğini özel bir amaç dışında yüksek, orta ve düşük olarak da nitelendirmek gerekir. Geçerlik, bir ölçme aracının kullanıldığı amaca hizmet etme derecesi olduğundan, ölçme araçlarından elde edilecek ölçümler hangi amaçla kullanılacaksa ölçme geçerliğı de o amaca bağlı olarak değişecektir⁹. Bir ölçeğin geçerliğı sadece kendisiyle de belirlenemez. O ölçeğin, kullanım amacına, uygulandığı gruba, uygulama ve puanlama biçimine de bağlıdır⁹.

Geçerliğin yüksek olması, büyük ölçüde, ölçülmek istenen değişkenin ifade edilebilmesine bağlıdır. Bu yönü ile doğrudan ölçmelerde geçerlik daha yüksek, dolaylı ölçmelerde ise değişkenin yeterince tanımlanamaması ve kriterlerinin yeterince duyarlı olmaması nedeniyle, geçerlik daha düşük olabilmektedir. Ölçeğin geçerlik düzeyi onun geçerlik katsayısının hesaplanmasıyla anlaşılır. Geçerlik katsayısı, ölçekten elde edilen değerlerle ölçeğin kullanım amacına göre belirlenen kriter ya da kriterler takımı arasındaki ilişki katsayısıdır ve -1.00 ile +1.00 arasında değerler alır. İlişki katsayısı ne kadar yüksekse ölçek amaca o kadar yüksek hizmet ediyor demektir^{1,9,10,12,22}.

Geçerlik katsayısının düşük bulunması, sadece ölçekten elde edilen değerlerle kriter değerleri arasındaki ilişkinin zayıflığından kaynaklanmaz, aynı zamanda elde edilen değerlerin güvenirliklerinin tam olmayışından da kaynaklanabilir. Bu nedenle geçerlik katsayıları güvenirlilik katsayılarıyla birlikte yorumlanır⁹.

Bir ölçmenin geçerli sayılabilmesinin ilk koşulu onun güvenilir olmasıdır²². Güvenirlilik, geçerlik için gerekli koşul olmasına rağmen, yeterli koşul değildir. Güvenilir bir ölçek her zaman geçerli olmayabilir. Hatta bazen, ölçeği güvenilir yapma amacı, ölçeği geçerli kılma amacıyla çatışabilir¹⁰. Bu nedenle geçerliğı yüksek olan ölçme aracının bir dereceye kadar güvenirligi de yüksektir. Fakat güvenirligin yüksek olması aracın, geçerliğinin de yüksek olacağı hakkında tam bir bilgi vermez¹³. Ölçüm sonuçlarının ne kadarıyla

Ölçeklerde Güvenirlik ve Geçerlilik

ölçüm hatasını yansıttığı sorusuna yanıt verebilmek için ölçeğin aracının geçerliğinin saptanmasına gerek vardır¹².

2.1. Kapsam Geçerliği (Content Validity)

Kapsam geçerliği, bir bütün olarak ölçeğin ve ölçekteki her bir maddenin amaca ne derece hizmet ettiği-dir¹⁰.

Bir ölçeğin kapsam geçerliği mantıki yoldan ve istatistiki yoldan olmak üzere iki türlü incelenebilir.

i) Mantıki yoldan ölçeğin kapsam geçerliğini araştırmak: Bu yaklaşımla ölçeği uygulamadan, ölçeğin geçerliği tahmin edilmektedir. Ölçüm konusu kavramsal olarak tanımlanır. Tanımlanan kavram doğrudan ölçüye vurulamayacağı için ölçekteki her maddenin ve bunların dağılımının ölçüm konusunu örnekleyip örneklemeyeceği araştırılır^{9,10,12}.

ii) İstatistiki yoldan ölçeğin kapsam geçerliğini araştırmak: İlgilenilen alanda daha önceden geliştirilmiş olan ve ilgilenilen alanın geçerli ölçüsü olarak kabul edilen standart bir ölçek ve yeni geliştirilen ölçek aynı anda bireylere uygulanır ve bireylerin her iki ölçekten aldıkları puanlara göre ilişki katsayısı hesaplanır. Hesaplanan ilişki katsayısı kapsam geçerliği belirlenmeye çalışılan ölçeğin geçerlik katsayısı olarak nitelendirilir. Bu işlem, ölçüt alınan ölçeğin kapsamının geçerli olduğu varsayımına dayanır. Bu nedenle, varılan sonuç temeldeki bu varsayımın sağlamlığı oranında geçerli olacaktır^{9,10}.

2.2. Kriter Geçerliği (Ölçütsel Geçerlik, Criterion Validity)

Bu geçerlik türü, ölçeğin etkinliğini belirlemek amacıyla, ölçekten elde edilen puanlarla belirlenen kriter arasında, gelecekteki veya o andaki, ilişkiyi ince-ler^{11,14,17}.

Kriter geçerliğin, kestirimsel geçerlik (predictive validity) ve uyum geçerliliği (concurrent validity) olmak üzere iki alt grubu vardır.

2.2.1. Kestirimsel Geçerlik (Yordama Geçerliği, Predictive Validity)

Bir ölçeğin kestirimsel geçerliği, o ölçekten elde edilen kestirimsel puan ile ölçülmek istenen özellikleri ölçtüğü bilinen kriter arasındaki korelasyonun hesaplanmasıyla elde edilir⁹.

2.2.2. Uyum Geçerliği (Concurrent Validity)

Eşzamanlı olarak geliştirilen ölçekten elde edilen puanlarla, belirlenen kriter arasındaki korelasyon uyum geçerliği olarak değerlendirilir¹.

2.3. Yapı Geçerliği (Construct Validity)

Yapı birbirleriyle ilgili olduğu düşünülen belli öğelerin ya da öğeler arasındaki ilişkilerin oluşturduğu bir örüntüdür. Bir ölçeğin yapı geçerliğini belirleme süreci bir ölçüde, bilimsel kuram geliştirme süreciyle

aynıdır¹⁰. Yapı geçerliğini belirleme süreci aşağıdaki adımları kapsar⁹.

- i) Ölçek performansını ölçmede kullanılacak yapının tanımlanması (kavram-psikolojik yapıyı tanımlama).
- ii) Tanımlanmış yapının altında yatan teoriden ölçek performansı ile ilgili sınanabilir hipotezler çıkarmak (kavramı ya da yapıyı belirleyen ölçek sonuçlarına ilişkin hipotezleri teoriden çıkarma).
- iii) Çıkarılan hipotezleri sınamak için deneysel ve istatistiki çalışmalar yapma ve yorumlamalarda bulunmak.

2.4. Görünüş Geçerliği (Face Validity)

Ölçeğin aracının neyi ölçtüğünü değil de neyi ölçer göründüğünü belirtmektedir. Bir ölçeğin görünüş geçerliği, o ölçeğin ölçmek istediği özelliği ölçüyor gözükmesidir. Ölçeğin görünüş geçerliğinin bazı durumlarda yükseltilmesi bazı durumlarda ise gizlenmesi gerekir^{9,10}.

Tartışma

Somut özelliklerin ölçülmesinde, ölçülecek olan özellik genel olarak koşulsuz bir değişmezlik göstermekte, kimi durumda ise bu özellik insan denetimindeki bazı koşullarla değişmezlik kazanabilmektedir. Soyut özelliklerin ölçülmesinde ise genellikle koşulsuz olarak (kendiliğinden) istenen derecede bir kararlılık göstermez¹³. Bu nedenle soyut özelliklerin ölçülmesi amacıyla geliştirilecek ölçeklerin uygun güvenilirlik analiz yöntemleriyle önce analiz edilmesi ve sonra da geçerlik yöntemlerinden uygulanması mümkün olanlarla geçerliğinin incelenmesi gerekmektedir. Ancak güvenilirlik ve geçerliği test edilerek yeterli bulunan ölçekler ölçeği uygulayan için geçerli veri sağlayacaktır. Ölçeğin güvenilirliği yükseldikçe ilgilenilen özelliğe ait anakütle parametresine için minimum varyanslı tahminlerde bulunmamızı sağlayacaktır.

Ölçek geliştirenlerin güvenilirlik ve geçerlik çalışmalarını yapması ve ölçeği uygulayanların da uygulayacakları ölçeğin güvenilirlik ve geçerlik çalışmasının yapıp yapılmadığını sorgulaması gerekmektedir. Eğer ölçeğin güvenilirlik ve geçerliğinin incelendiği toplum ile sonradan ölçeğin uygulanması düşünülen toplumlar arasında önemli farklılıklar varsa ölçeğin güvenilirlik ve geçerliğinin tekrar incelenmesi gerekebilir.

Kaynaklar

1. Carmines EG, Zeller RA. Reliability and Validity Assessment. 5th printing. Beverly Hills: Sage Publications Inc.; 1982.
2. Çömlekçi N. Temel İstatistik İlke ve Teknikleri. Eskişehir: Bilim Teknik Yayınevi; 1989.

3. Armağan İ. Yöntembilim-2 Bilimsel Araştırma Yöntemleri. İzmir: Dokuz Eylül Üniversitesi Güzel Sanatlar Fak. Yayınları; 1983.
4. Gümüş B. Eğitimde Ölçme ve Değerlendirme. Ankara: Kalite Matbası; 1977.
5. Gürsakal N. Bilgisayar Uygulamalı İstatistik-I. Alfa Yayınları; 2001.
6. Özdamar K. Paket Programlarla İstatistiksel Veri Analizi-1. 4. Baskı. Eskişehir: Kaan Kitabevi, 2002.
7. Gay LR. Educational Evaluation and Measurement. 2nd edition. London: A Bell & Howell Company; 1985.
8. Carey LM. Measuring and Evaluating School Learning. London: Allyn and Bacon Inc.; 1988.
9. Öncü H. Eğitimde Ölçme ve Değerlendirme. Ankara: Matser Basım San. Ve Tic. Ltd. Şti.; 1994.
10. Tekin H. Eğitimde Ölçme ve Değerlendirme. Ankara: Mars Matbaası; 1977.
11. Thorndike RM, Cunningham GK, Thorndike RL, Hagen EP. Measurement and Evaluation in Psychology and Education. 5th edition. Macmillian Publishing Co; 1991.
12. Sencer M, Sencer Y. Toplumsal Araştırmalarda Yöntembilim. Ankara: Doğan Basımevi; 1978.
13. Özçelik DA. Okullarda Ölçme ve Değerlendirme. Ankara: ÜSYM-Eğitim Yayınları; 1981.
14. O'Connor R. Issues in the Measurement of Health-Related Quality of Life. Working paper 30, NHMRC National Centre for Health Program Evaluation, http://www.rodconnorassoc.com/issues_in_the_measurement_of_qua.htm, Melbourne; 1993.
15. Dawson B, Trapp RG. Basic & Clinical Biostatistics. 3rd edition. New York: Lange Medical Books/McGraw International Editions; 2001.
16. Küçükahmet L. Öğretimde Planlama ve Değerlendirme. 11. baskı. Ankara: Nobel Yayın Dağıtım; 2000.
17. Gürsakal N. Sosyal Bilimlerde Araştırma Yöntemleri. Bursa: Uludağ Üni. Güçlendirme Vakfı; 2001.
18. Traub RE. Reliability for the Social Sciences. London: Sage Publications; 1994.
19. Turgut F M. Eğitimde Ölçme ve Değerlendirme Metodları. 9 baskı. Ankara: Saydam Matbaacılık; 1993.
20. Bruning JR, Kintz BL. (Çev: Dönmez A) İstatistik. Ankara: Gündoğan Yayınları; 1983.
21. Alpar R. Spor Bilimlerinde Uygulamalı İstatistik. Ankara: Nobel Yayın Dağıtım; 2001.
22. Karasar N. Bilimsel Araştırma Yöntemleri. 10. baskı. Ankara: Nobel Yayın Dağıtım Ltd.Şti.; 2000.