

Ölçek Geliştirme Çalışmalarında Temel İlkeler

Basic Principles Of Scale Development

Araştırma Görevlisi Dr. Fatma Yeşim Karakoç¹, Prof. Dr. Levent Dönmez¹

¹Akdeniz Üniversitesi Tıp Fakültesi Halk Sağlığı Anabilim Dalı

ÖZET: Ölçeklerin geliştirilmesi, uyarlanması ve kullanılması aşamalarında bir çok kriter ve standarda uygun çalışılması gerekir. Aksi halde ölçeğin geçerlik ve güvenilirlik düzeyi düşük olur. Geçerlik, bir ölçme aracının ölçmeyi amaçladığı özelliği, başka herhangi bir özellikle karıştırmadan, doğru ölçebilme derecesidir ve kullanım amacına, uygulandığı gruba ve uygulama biçimine göre değişebilir. Güvenirlik ise aracın, ölçmek istediği değişkeni ne tutarlılıkla ölçtüğünün ya da ölçme sonuçlarının hatalardan arınmış

olmasının derecesidir. Sadece ölçme aracına ait bir özellik olmayıp, ölçme aracı ve aracın sonuçlarına ilişkin bir özelliktir. Geçerliği istenen düzeyde olmayan bir ölçekle, istatistiksel testlerin gücü düşer ve gruplar arası farkın anlamlı olarak tespit edilmesi engellenir ya da çok daha fazla sayıda deneğin araştırmaya dahil edilmesini zorunlu kılarak maliyet artışı ve zaman kaybına neden olabilir. Güvenirliği düşük bir test ise yapılan ölçümde “bias”a neden olarak, özellikle klinik uygulamada bazı hatalı kararlara zemin hazırlayabilir. Bu çalışmada araştırmalarda kullanılan ölçeklerin oluşturulması ile ilgili yapılması gerekenlere ve geçerlik/güvenirlik için gerekli koşullara değinilmiştir.

Anahtar Sözcükler:

Geçerlik, Güvenirlik,
Ölçek

Key Words: *Validity,
Reliability, Scale.*

ABSTRACT: *Stages of development, adaptation and usage of the scale need to be studied in accordance with the criteria and standards. Otherwise, the reliability and validity of the scale level will be low. Validity is a measurement tool for correct calculation without mixing any particular characteristics and can change by depending on the group administrated and the way of administration. Reliability of a measurement tool depends on the degree of consistency or the degree of how much the tool is free of errors. It is not only a characteristics of a measurement tool, but also a characteristic for the results of the tool. If the validity of a scale is not at the desired level, it decreases the power of statistical tests and determination of the difference between groups is significantly blocked or may result in increased costs and loss of time, in a much larger number of subjects to be included in the research by making it mandatory. A test with low reliability concludes in a measurement “bias” and leads to faulty decisions in clinical practises.*

In this study, the validity / reliability of the necessary conditions have been mentioned, regarding the creation of the scales used in research needs to be done.

GİRİŞ

Ölçekler, ölçmeye konu olan özelliklerin sınıflanması, sıralanması ya da miktar ve derecelerinin belirlenmesi için uyulması gereken kural ve kısıtlamaları belirleyen ölçme araçlarıdır (1). Ölçekler, ölçme işlemini kolaylaştırmanın yanı sıra, elde edilen sonuçların niteliğinin de belirlenmesini sağlar. Bilimsel gelişme ölçmeye dayanır, duyarlı ölçüm araçlarıyla yapılan ölçmeler de bu gelişmeyi artırır (2).

Bir ölçeğin geçerli ve güvenilir olması için, ölçeğin geliştirilmesi ve kullanılması aşamalarında bir çok kriter ve standarda uygun çalışılması ve yorum yapılması gerekmektedir. Aksi halde ölçeğin geçerlik ve güvenilirliği düşerek, ölçeğin ülke düzeyinde ve hatta uluslararası düzeyde kullanıldığı çalışmalarda bir takım hatalar ve taraf tutmalara zemin hazırlanmış olacaktır.

Bu çalışmada araştırmalarda kullanılan ölçeklerin oluşturulması ile ilgili yapılması gerekenlere ve geçerlik/güvenirlilik için gerekli koşullara değinilmiştir.

ÖLÇEK GELİŞTİRME SÜRECİ

Ölçek geliştirme çalışmasının öncesinde ilk yapılacak şey, “konu ile ilgili mevcut bir ölçek olup olmadığı”, eğer varsa “ne kadar duyarlı bir ölçek olduğunun” ve mevcut ölçeğin duyarlılığı yüksekse, “bir uyarlama çalışması yapmanın mı yoksa yeni bir ölçek oluşturmanın mı daha anlamlı olacağı”nın değerlendirilmesidir (2,3).

a) Var olan bir ölçeğin uyarlanması: Bir ölçeği sadece farklı bir dile çevirip kullanmak değil, o ölçekle ilgili geçerlik ve güvenilirlik çalışmalarının da yapılmasına, ölçeğin başka dil ve kültürlerle uyarlanması denir.

Yeni bir ölçek geliştirmek yerine uyarlama yapmayı tercih etmenin bazı avantajları vardır. Bunlar arasında en önemlisi daha az

zaman ve maliyet gerektirmesidir. Ayrıca ölçülecek özelliğin kültürler arası veya belirli bir ülkenin bölgeleri arasında karşılaştırılması amaçlanıyorsa; mevcut bir ölçeğin uyarlanması bu özelliğin farklı bir kültürde ölçülmesinin daha uygun bir yolu olacaktır. Bir diğer avantaj ise, ikinci bir kültürde ölçek geliştirmek için uzmanlık bilgisinin yetersiz olduğu durumlarda ortaya çıkmaktadır. Böyle bir yetersizlikte yeni ölçek geliştirmek yerine uyarlama yapmak daha mantıklı bir seçenektir ve asıl ölçek iyi bilinen bir ölçek ise, uyarlandığında duyulacak olan güven, yeni geliştirilecek ölçeğe göre daha fazla olacaktır (3).

Uyarlama çalışması yapmanın da bazı sıkıntıları vardır. Bunlardan en önemlisi, her iki dile de hakim olan ve ölçek uyarlama konusunda çalışabilecek uzman bulma sıkıntısıdır. Eğer belirtilen bu özelliklere sahip olan bir uzman grubu yoksa konunun özelliklerini karşılamak için farklı farklı uzmanlar belirlenecek ve bunların da beraber çalışması zamanlama açısından sıkıntılı olabilecektir. Diğer bir sıkıntı ise, uyarlama sonrası ölçeğin psikometrik özelliklerinin sağlanamaması sorunu olabilir. Eğer uyarlanacak ölçek yüksek bir güvenilirlik ve geçerliğe sahipse, uyarlama sonrası da bu değerlere yakın değerler vermesi beklenir.

Ancak sonuç bazen, aslında iyi psikometrik özelliklere sahip bir ölçeğin, uyarlanmış şeklinin daha düşük değerlerde psikometrik özelliklere sahip olması şeklinde olabilir.

Uyarlama çalışmasının belli bir plan ve sistem dahilinde yapılması ve karşılaştırılması muhtemel durumların dikkate alınması bu açılardan önem arz etmektedir (3).

Uyarlama aşamaları: Uyarlama çalışmasına karar verildiğinde, yapılacak olan ilk iş ölçek sahibinden izin almaktır. Ölçek sahibi bu izni verirken bazı bilgilere gereksinim duyacaktır. Örneğin; uyarlanan ölçeğin kullanım amacının

belirtilmesi istenebilir (araştırma veya pazarlama amaçlı uyarlama izni isteniyorsa, ölçeğin telif haklarına sahip olan kişi ya da kurum telif ücreti talep edebilir) veya uyarlama çalışması yapacak ekibin yeterliliğinin olup olmadığının belgelenmesi talep edilebilir (diploma, daha önceden yapılmış çalışmalar, referans mektubu gibi). Bu yüzden izin istenirken bu türden bilgi ve belgelerin hazırlanması uygun olacaktır.

Daha sonra ölçülecek olan yapının, karşılaştırılacak tüm kültürlerde aynı anlama gelip gelmediği sorgulanmalıdır. Eğer aynı yapı yoksa yapılacak düzeltmelerle ortak bir yapı oluşturulmalı, oluşturulamıyorsa çalışma durdurulmalıdır.

Çevirme aşamasında, her iki dilde de akıcı konuşan, çalışma yapılan kültürleri tanıyan, ölçülen yapı hakkında bir miktar bilgiye sahip çevirmenler tercih edilmelidir. Çeviri konusunda yapılacak uygulama bir dilden diğer dile çevirme ya da bir dilden diğerine çeviri yapıldıktan sonra tekrar asıl diline çevirme şeklinde olmalıdır (3). Psikometrik bir terime karşılık gelen Türkçe karşılık uygun değilse, yapılan işlemler de hatalı olabilir. Ülkemizde üretilen Türkçe dil karşılıkları kimi zaman terimin anlamını karşılamayarak yapılan işlemlerin hatalı olmasına yol açabilmektedir (4). Çeviri sonrası uyarlama yapılmalı, uyarlanmış hali gözden geçirilmeli ve gerekiyorsa değişiklikler yapılmalıdır. Uyarlama sonrası ölçek deneme grubuna uygulanmalı ve ölçekle ilgili yapılması gereken başka düzeltmeler olup olmadığına bakılmalıdır. Pilot grubun özellikleri ölçeğin esas uygulamasının yapılacağı grubun özelliklerini taşımalıdır.

Uyarlama sonrası ölçekte geçerlik ve güvenilirlik çalışması yapılmalıdır. Geçerlik için yapı geçerliğini sağlamaya yönelik olan faktör analizi yöntemi ya da deneysel veya ilişkisel çalışmalar (yordama veya uygunluk geçerlikleri

gibi) şeklinde yapılmalıdır.

Uyarlama tamamlandıktan sonra, testin kullanıcıları için süreç ve hazırlamayı içeren bir el kitabı hazırlanmalı ve uygulanması ve sonuçların yorumlanmasıyla ilgili bilgiler verilmelidir. Uyarlama yapan araştırmacılar ölçeğin sürekliliği için ölçeği takip etmeli ve yeniden geçerlik ve güvenilirlik çalışmaları yapmalıdır (3).

b) Yeni bir ölçek geliştirme: Yeni bir ölçek çalışması yapılacaksa ilk adım, konuyla ilgili literatür taraması yapmaktır. Bunu yaparken, ölçek konusu için hangi soruların/konuların değerlendirilmesi gerektiğine dikkat edilmesi gerekir (5).

Ölçüm şekli için format belirlenmeli ve buna göre madde/soru havuzu oluşturulmalıdır (5). Bu format, verilerin matematiksel özelliklerine göre, sıralama, sınıflama, eşit aralıklı ve oran ölçeği şeklinde ya da bazı özel ölçüm şekillerine göre (Thurstone tipi, Likert tipi, Osgood boyutsal ayırma ölçeği gibi) olabilir (2).

Maddeler tasarlanırken, maddelerin sade ve anlaşılır olmasına, bir maddenin birden fazla yargı ve düşünce ifadesine sahip olmamasına özen gösterilmelidir (6,7). Açılış soruları aranan bilginin türüne göre ve daha sonraki sorular üzerinde etkisini dikkate alarak düzenlenmelidir. Katılımcılar ankete yanıt vermeye bir dizi kişisel ve paylaşmakta çekineceği soru ile başlamak istemeyebilir. Bu nedenle, bu tür soruların ilerleyen bölümlerde yer alması daha uygun olur (7).

Sonraki aşama, oluşturulan madde havuzu için uzman yorumuna başvurmaktır. Uzmanlar soruların gerekliliği, açık ve özgüllüğünün değerlendirmesini yapar. Böylece bazı ifadelerin çıkarılması ya da değiştirilmesi önerisi getirebilirler. Ancak, tavsiyelerin kabul ya da reddedilmesi, ölçek hazırlayanın kendi iradesine bırakılmalıdır (5). Uzman yorumu ile kapsam ve

görünüm geçerliği değerlendirilmiş olur (8). Uzman görüşüne göre şekillendirilen ölçek, taslak örneklem grubuna uygulanır. Deneme uygulaması ölçeğin performansı için önemlidir. Örneklem hedef kitleyi temsil etmelidir (9). Taslak uygulaması sonrası, ölçeğin geçerlik ve güvenilirlik analizleri ile değerlendirilmesi yapılarak, son hali verilir (5).

Geçerlik ve Güvenirlik Kavramları

Geçerlik (Validity): Geçerlik, bir ölçme aracının ölçmeyi amaçladığı özelliği, başka herhangi bir özelliklerle karıştırmadan, doğru ölçebilme derecesidir (10). Geçerlik, bir ölçme aracının kullanıldığı amaca hizmet etme derecesi olduğundan, ölçme araçlarından elde edilecek ölçümler hangi amaçla kullanılacaksa ölçme geçerliği de o amaca ve hatta uygulandığı gruba ve uygulama biçimine göre de değişebilecektir (10). Bu yüzden geçerliği bozan etkenler, ölçek geliştirme ve uygulama süreci ile ilgilidir (2). Genel olarak, doğrudan ölçmelerde geçerlik daha yüksek, dolaylı ölçmelerde ise daha düşük olma yönündedir (10).

Ölçeklerin geçerlik düzeyi, 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 iyi hizmet ediyor demektir (10).

Geçerlik katsayısının düşük bulunması, sadece ölçekten elde edilen değerlerle kriter değerleri arasındaki ilişkinin zayıflığından kaynaklanmaz, aynı zamanda elde edilen değerlerin güvenilirliklerinin tam olmayışından da kaynaklanabilir.

Güvenilir bir ölçek her zaman geçerli olmayabilir. Bu nedenle geçerlik katsayıları güvenilirlik katsayılarıyla birlikte yorumlanmalıdır (10).

Geçerlik Türleri ve Ölçümünde Kullanılan Yöntemler

Araştırmacılar arasında geçerlik türleri için terminolojik ve tanımsal farklılıklar yer almaktadır. İyi bir geçerlik ölçütü olma niteliği ise çalışmanın niteliğine göre değişmekle birlikte (2) yorumsal geçerlik, yapı geçerliği ve ölçüte dayalı geçerlik olmak üzere 3 ana gruptan oluşur (11).

1.Yorumsal Geçerlik (Translation Validity)

Ölçek yapısının ne kadar işlevselliğe dönüştüğünü gösterir.

a) Yüzeysel Geçerlik (Görünüş Geçerliği)

(Face Validity): Ölçme aracındaki soruların net bir şekilde incelenen konu ile ilgili bilgiye yönelik olduğuna dair uzman görüşüdür (2). Ancak bazı konular için görünüş geçerliği göz ardı edilebilir. Örneğin; ölçme aracındaki maddelerin cevaplayanda farklı tepki oluşturabileceği konular için, soruların etik kuralları ihlal etmeyecek tarzda düzenlenmesi kaydı ile farklı şekilde sorulması gibi. Ancak görünüş geçerliğinin teknik bir geçerlik sayılmayacağını düşünenler de vardır (2).

b) İçerik (Kapsam) Geçerliği (Content Validity):

Mantık yolu ve istatistiksel yol şeklinde iki farklı uygulama ile değerlendirilebilir (10). Kapsam geçerliği bir bütün olarak ölçeğin ve ölçekteki her bir maddenin amaca ne derece hizmet ettiği. Kapsam geçerliği için konunun uzmanlarının görüşü alınır. Konuyla ilgili kapsam belirlenmesi bir yargılamayı gerektirdiğinden, farklı ölçütlere göre değerlendirme yapılmasının engellenmesi için uzmanlar ile ölçeği geliştiren kişi arasında ortak tanımların olması gerekir. Burada maddelerin sunum biçiminin yaptığı farklı tepkiler açısından da bir değerlendirme yapılmış olur (görünüş geçerliği) (2).

Uzmanların kapsam geçerliği için yaptığı değerlendirme çeşitli tekniklerle yapılabilir. Davis tekniğinde maddeler; “uygun”, “madde

hafifçe gözden geçirilmeli”, “madde ciddi olarak gözden geçirilmeli” ve “madde uygun değil” şeklinde dördü derecelendirilmektedir. Bu teknikte, maddelerin uygun oluşu ve maddenin hafifçe gözden geçirilmeli seçeneklerini işaretleyen uzmanların sayısı toplam uzman sayısına bölünerek maddeye ilişkin “kapsam geçerlik indeksi” elde edilir. Bu değer 0,80 olması kabul edilebilir bir düzey demektir (8).

Soyut kavramlarla ilgili ölçüm araçlarında, kavramla simgelenen davranışın içeriğinin kolayca ve açıklıkla belirlenememesi kapsam geçerliğinin sınanmasını zorlaştırabilir. Bu nedenle yetenek ve kişilik testleri için kapsam geçerliğinin uygun olmayacağı, yanlış yönlendirilebileceği, bu yüzden nihai geçerlik aşaması için ampirik doğrulama gerektiğini savunanlar vardır (2).

Diğer bir yol, aynı kapsamı ölçtüğü bilinen geçerlik ve güvenilirliği kanıtlanmış bir ölçme aracı ile korelasyonunun hesaplanmasıdır. Ancak bu konuda böylesi bir ölçek varken, yeni ölçek için farklı ya da üstün özellik taşıdığı düşünülüyorsa ekonomik bir yaklaşım olmaz (2).

2. Ölçüte Dayalı Geçerlik (Criterion-Related Validation)

Test puanlarının belirlenen bir veya birkaç dış ölçüt ile ilişkisini inceleyen bir tekniktir. Eşzaman ve yordama geçerliği olmak üzere iki türdür.

Eşzaman geçerliğinde (Concurrent validity), katılımcıların geliştirilmek istenen ölçekten aldıkları puanların hem aynı davranışı ölçen başka bir test, hem de ilişkili başka bir davranışı ölçen bir test ile ölçülen puanlarının korelasyonuna bakılır. Karşılaştırılacak ölçmelerin aynı ya da yakın zamanda yapılması istenir (12).

Yordama geçerliği (Predictive validity), sonradan alınacak referans sonucun önceden tahmin edilmesidir. Burada duyarlılık – seçicilik

(özgüllük) önemli iki kavramdır ve incelenen sonuçlar dikotom (örn: hasta/sağlam) sonuçlar olmalıdır. Bu yolla test sonuçlarının gelecekteki davranışı ne ölçüde yordadığı araştırılır. Üniversiteye girişe yönelik bir sınavın başarıyı ne derece yordadığını araştırmak için bu öğrencilerin üniversitedeki başarıları ile kıyas yapmak buna örnek verilebilir (2,12).

3. Yapı Geçerliği (Construct Validity)

Yeni testi karşılaştıracak bir kriter (referans) yoksa yapı geçerliği sınanmalıdır. Ölçekten elde edilen sonucun ve bu sonucun ne ile bağlantılı olduğunun açıklanmasını sağlar (2). Hazırlanan ölçek maddelerinin belirlenen özellikleri ne derece doğru ölçtüğü ile ilgilidir (2,12).

Yapı geçerliğini değerlendirme yöntemleri içinde farklı araştırmacılar farklı yöntemlere işaret etmektedir. Bunlardan genellikle sözü edilenler; faktör analizi, iç tutarlık analizi, hipotez testi (2,12), benzer ölçek geçerliği (convergent and discriminant validity), sağlama geçerliği, yapısal eşitlik modellemesi (2), çoklu özellik/çoklu yöntem matrixi (Multi-trait Multi-method Matrix-MTMM), şablonların eşleşmesi teorisi gibi (11) yöntemlerdir. Burada en yaygın kullanılan yöntemler olduğu için, benzer ölçek geçerliği, yapısal eşitlik modellemesi ve faktör analizinden bahsedilmesi uygun olacaktır.

a) Benzer Ölçek Geçerliği (Convergent and Discriminant Validity): Birleşim-ayrışım geçerliği, ölçeğin belirli bir alanı ilgilendiren boyut puanının aynı kavramı sorguladığı iddia edilen diğer benzer ölçeğin aynı boyutu ile ya da aynı şeyi gösteren diğer bazı parametrelerle yüksek korelasyon göstermesi varsayımına dayanır (13).

b) Yapısal Eşitlik Modellemesi (Structural Equation Modeling): Faktör analizi ölçeğin hangi yapıları ölçtüğünü belirlerken, yapısal eşitlik modellemesi ile ölçeğin içerdiği yapılar arasındaki ilişki ve ilişkinin yönü de

incelenabilir (2).

Yapı geçerliği çalışmaları sonucu araştırmacı beklediğinin tam tersi olumsuz bir sonuçla karşılaşır; ölçeğin kavramsal yapıyı ölçmediği, hipotezin dayandığı teorik ilişki ağının doğru olmadığı ya da deneysel tasarımın hipotezi ölçmede yetersiz kaldığı düşünülür. İlk iki seçenek için ölçeğin yapısal geçerlikten uzak olduğu söylenirken, sonuncusu için araştırma tasarımını iyileştirmeye yönelik önlemler alınması gerektiği sonucuna varılır (14).

c) Faktör Analizi (Factor Analysis):

Cevaplayıcıların, geliştirilmekte olan ölçme aracındaki maddelere verdiği tepkiler arasında belli bir düzen olup olmadığını ortaya koymak için kullanılan bir yapı geçerliği tekniğidir (2). Faktör analizi ile birçok değişken, birkaç başlık altında toplanabilir. Örneğin 50 maddelik bir ankette hekimlerin empati, kişisel başarı hissi ve iş doyumunu düzeyleri değerlendiriliyor olsun. Faktör analizi sayesinde bu düzeylerin her biri için faktör skoru elde edilir. Böylece 3'ten daha az veya birbirinden farklı faktörün olup olmadığı ortaya çıkar (15).

Ölçek geliştirme sürecinde, hazırlanan araç, araştırma evreninden yansız olarak seçilen örnekleme verilir ve maddelere verilen cevaplar puanlandırılarak faktör analizi uygulanır. Analiz sonuçlarına göre maddelerin araçtan çıkartılması ya da eklenmesi sonrası analiz tekrar edilir. Bu süreç, ölçülecek alanı ölçmede yeterli sayıda madde içeren uygun bir çözüme ulaşıncaya dek devam eder.

Faktör analizi tüm değişkenlerin ve bu değişkenlerin lineer kombinasyonlarının normal dağıldığını varsayar. Bu varsayım karşılanıyorsa çözümün değeri artar. Ayrıca örneklem büyüklüğünün, ilişkilerin güvenilir bir şekilde kestirilebilmesini sağlayacak büyüklükte olması önemlidir (16). Bu sayı hakkında ilişki güvenilirliği ve belirgin faktör sayısına göre farklı tanımlamalar yapılmaktadır. Genel bir

kural olarak ise, örneklem büyüklüğünün en az gözlenen değişken sayısının beş hatta on katı olması gerektiği ifade edilmektedir (2,16,17). Faktör analizinde değişken çiftleri arasındaki ilişki doğrusal olmadığında, analizin değeri azalır (16).

Güvenirlilik (Reliability): Bir ölçme aracının güvenirliliği, aracın ölçmek istediği değişkeni ne tutarlılıkla ölçtüğünün ya da ölçme sonuçlarının hatalardan arınmış olmasının derecesidir (2,18). Bir ölçme aracıyla aynı koşullarda tekrarlanan ölçümlerde elde edilen ölçüm değerlerinin kararlılığının bir göstergesidir. Güvenirlilik sadece ölçme aracına ait bir özellik değildir, ölçme aracı ve aracın sonuçlarına ilişkin bir özelliktir (10). Bu yüzden testin ya da ölçme aracının güvenirliliği yanlış bir tanımlama olup ölçüm güvenirliliği olarak adlandırma yapılması daha uygundur (19).

Güvenirlikte 3 kavramdan söz edilir:

1. Tutarlılık: Bir ölçme aracındaki maddelerin (soruların) testin bütünü ile uyumudur.

2. Kararlılık: Ölçme konusu olan özelliğin, aynı ölçme aracıyla farklı zamanlarda birkaç kez ölçümünde aynı sonuçların elde edilmesidir.

3. Duyarlılık: Ölçme aracı veya sonuçlarının biriminin büyüklüğü ile ilgilidir. Birim aralığı küçükse ölçme daha duyarlıdır (20).

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 güvenilirdir (10).

Güvenirlilik Katsayılarının Hesaplanması için Kullanılan Yöntemler

Güvenirlilik katsayısı hesaplama kullanılan yöntemler için farklı sınıflamalar yapılmaktadır (2,10,11). Test-retest ve eşdeğer formlar yöntemi, puanlama tutarlılığı ve iç tutarlılık katsayıları güvenirlikle ilgili değerlendirmede kullanılan katsayılardır (11).

Güvenirliliği düşük olan bir ölçmenin bilimsel bir değeri olmadığı gibi, güvenirliliğin yüksek olması da yapılan ölçmenin amaca uygunluğunun garantisi değildir. Yani güvenirlilik, zorunlu fakat yeterli bir koşul değildir (21).

Formun Tekrarı Yöntemi (Zamana Göre Değişmezlik (Süreklilik) (Test-retest)

Bir ölçme aracının, aynı denek grubuna, aynı koşullarda ve belli bir zaman aralığında iki kez uygulanmasıdır. İki uygulamadan elde edilen ölçüm değerleri korelasyon katsayısı, ölçeğin güvenirlilik katsayısıdır (10, 21).

Örneğin, bir öğretim üyesinin sınavda sorduğu açık uçlu sorularının puanlamasını yaptıktan sonra verdiği ilk puanlarla, belli bir aradan sonra aynı sınav kağıtlarını yeniden incelemesi sonucu verdiği puanlar arasında hesaplanacak korelasyon katsayısı, o öğretim üyesinin ölçmelerine olan güvenirliliğini yansıtır. Bu tür yaklaşımın en kritik yönü, iki ölçme arasında bırakılması gereken zaman aralığının iyi ayarlanabilmesidir. Zamanın çok kısa olması, yeniden hatırlamayı kolaylaştıracağından, yapay (suni) olarak yükselmiş bir güvenirlilik ölçütü çıkmasına; zamanın uzaması ise ölçülen özellikte bazı değişmelerin meydana gelebileceği durumlarda iki ölçme için “aynı koşullar”ın sağlanmasını olanaksızlaştırabileceğinden güvenirlilik ölçütünün yorumunun güçleşmesine neden olur. Yani ölçeğin güvenirliliği mi düşüktür yoksa bireylerin sahip olduğu özelliklerde mi bir değişiklik olmuştur ayrımı yapılamaz (2, 10, 21). Bu sürenin ölçülen davranış ve hedef kitleye göre değişmekle birlikte ortalama 4 haftalık sürenin yeterli

olduğu belirtilmektedir (12). Zamana göre değişmezlik ölçütü ile bulunacak güvenirlilik katsayısı için, veri türüne göre, uygun bir korelasyon çözümü yapılabilir (21). Ölçülen özellik sürekli değişken ve geliştirilen ölçek eşit aralık ya da oran ölçeği ise en güçlü, kuvvet yetkinliği en yüksek korelasyon tekniği olduğu için Pearson Momentler Çarpımı Korelasyon Katsayısı hesaplanır (2).

Korelasyon katsayıları iki değişken arasındaki ilişkinin derecesi ve yönü hakkında bilgi verir ve -1 ile +1 arasında değer alır. Katsayının +1 olması pozitif ve mükemmel ilişki varlığını gösterir. Bir ölçeğin kararlılık gösterdiğinin kabulü için en az 0,70 olması gerekir. Bireylerin geleceğini etkileyecek kararlarda kullanılacak ölçme araçlarında ise en az 0,95 olması istenir (2).

2. Eşdeğer (Paralel) Formlar Yöntemi (Parallel Forms Reliability)

Ölçek geliştirilirken, aynı zorluk derecesinde ve aynı içerikte soru veya maddelerden oluşturulmuşsa iki eşdeğer (paralel) form haline getirilebilir (2, 10). Ölçmelerin eşdeğer olabilmesi için aynı değişkeni ölçmesi yanı sıra ortalama ve standart sapmalarının eşit olması gerekir (2).

Eşdeğer iki form aralıksız olarak aynı anda ya da aralıklı olarak farklı iki zamanda uygulanır (2,10). Eşdeğer formların uygulanışında, aradaki zaman aralığının artması kararlılığı olumsuz yönde etkileyecekse, formlar deneklerin sıkılmalarını ve yorulmalarını engelleyecek kadar ara süre verilerek ardışık uygulanmalıdır (10).

Formlar arasındaki korelasyon hesaplanır ve güvenirlilik katsayısı olarak yorumlanır (2,10). Koşullar uyuyorsa kullanılan korelasyon katsayısı genelde Pearson Momentler Çarpımı Korelasyon Katsayısıdır. Dilsel eşdeğerliği belirlemek için yapılan uygulama da eşdeğer formlar yöntemidir (2).

3. İç Tutarlılık (Internal Consistency)

Belirli bir alanı ölçtüğü varsayılan soruların kendi aralarında ne kadar homojen olduğunun, soruların doğru adrese gidip gitmediğinin yani yalnızca istenen kavramı ölçüp ölçmediğinin iyi bir ölçütüdür. Sık başvurulan bir güvenilirlik ölçütüdür. İç tutarlılığın dayandığı temel görüş, her ölçme aracının, belli bir amacı gerçekleştirmek (bütünü oluşturmak) üzere birbirinden deneysel olarak bağımsız ünitelerden (örneğin, test maddelerinden, anket sorularından) oluştuğu ve bunların bütün içinde, bilinen ve birbirlerine eşit ağırlıklara sahip olduğu varsayımdır (21).

a) Yarıya Bölme Yöntemi (Split-Half Method): Yöntem, formu iki eş parçaya bölerek, iki yarının deneklere aynı anda uygulanması sonrası, deneklerin yanırlardan aldıkları puanlar arasındaki korelasyon (yarı testin korelasyon katsayısı), güvenilirlik tahmini yapılmasını sağlar (koşulları uyuyorsa yine Pearson Momentler Çarpımı Korelasyon Katsayısı ile) (10). Ölçeğin alt boyutları varsa her boyut kendi içinde bütün kabul edilerek uygulanabilir.

Ölçeğin tamamının güvenilirlik katsayısı için Spearman –Brown formülü uygulanır (2).

b) Kuder-Richardson (KR) Güvenirlik Katsayıları: Kuder-Richardson formülleri, testteki her maddenin aynı değişkeni ölçtüğü yani testin ölçtüğü şeyin homojen olduğu varsayımına dayanır. Testin iki yarısı yerine testteki tüm maddeler arasındaki tutarlılığın ölçümünü verir ve iç tutarlılık katsayısı olarak adlandırılır (10).

KR-21 madde analizi yapılmamış testlere uygulanır ve testteki maddelerin madde güçlükleri eşit farz edilir. Bu nedenle KR-21 formülü ile elde edilen katsayı güvenilirliğin alt sınırı olarak kabul edilir. Bir testin KR-20 veya KR-21 güvenilirliği yüksek bulunmuşsa, bu testteki maddelerin aynı yeterliği ölçtüğü (testin tek boyutlu olduğu) ve test puanlarının

tesadüfi hatalardan arınık olduğu sonucuna ulaşılabilir (22). Verilerin dikotom olması gerekir. Ölçekteki maddelerden alınan cevaplar istenilen özelliği taşıyorsa “1” puan, istenilen özelliği taşıyamıyorsa veya boş bırakılmışsa “0” puan verilerek oluşturulur (10).

c) Cronbach Alfa Güvenirlik Katsayısı: Cronbach alfa katsayısı, ölçekte yer alan k maddenin varyansları toplamının genel varyansa oranlanması ile bulunan bir ağırlıklı standart değişim ortalamasıdır (10).

Cronbach (1951) tarafından geliştirilen alfa katsayısı yönteminin, maddeler doğru-yanlış olacak şekilde puanlanmadığında, 1-3, 1-4, 1-5 gibi puanlandığında, kullanılması uygun olan bir iç tutarlılık tahmin yöntemi olduğunu savunanlar (10) yanı sıra ikili durumlarda da kullanılabileceğini savunan görüşler de vardır (19).

Hesaplanan katsayı için genel kabul en az 0,70 olmasıdır.

Değerlendirme aşağıdaki şekilde de yapılabilir:

→ölçek güvenilir değil

→ölçek düşük güvenilir

→ölçek oldukça güvenilir

→ölçek yüksek derecede güvenilir (2).

Ölçek yerine, doğrudan davranışların gözlenmesi durumunda da her gözlemin bir madde gibi düşünülmesi ile yine Cronbach alfa katsayısından faydalanarak güvenilirlik hesaplanabilir.

4. Puanlama Tutarlılığı

a) Gözlemciler Arası Tutarlılık (Inter-observer Consistency): Birden çok gözlemcinin, birbirinden bağımsız olarak, aynı durumu, aynı ölçüm aracı ile ölçmeye çalıştığı zaman uygulanan bir güvenilirlik ölçütüdür.

Bağımsız gözlemciler arası uyumu hesaplamak için korelasyon, t-testi, özel varyans çözümlenmeleri ve Cronbach alfa kullanılır. %70 ve daha yüksek tutarlılık, güvenilirlik sınaması için uygundur (11). Ancak tutarlılık yüzdesinin şansa bağlı olma durumu da söz konusu olabilir. Bu durumda kappa katsayısı kullanılır (2).

Kappa katsayısı= gözlenen tutarlılık-beklenen tutarlılık/1- beklenen tutarlılık
Katsayının “0” olması tam uyumsuzluğu gösterirken 1’e yaklaşması uyumun göstergesidir (-1 ise ters uyum) (11).

b) Gözlemciler İçi Tutarlılık (Intra-observer Consistency): İki veya daha fazla gözlemin aynı gözlemci tarafından yapılması söz konusu olduğunda ölçümler arası tutarlılık yüzdesi kullanılır (11). Geçerlik ve güvenilirlik sınamaya yönelik bahsedilen yöntemler Tablo 1’de sunulmuştur.

Tablo 1. Geçerlik ve güvenilirlik sınama yöntemleri ve kullanılan testler

Güvenirlik		Geçerlik			
		İstatistiksel test			İstatistiksel test
Ölçeğin kararlılığı	Test-retest	Pearson momentler çarpımı korelasyon katsayısı	Yorumsal geçerlik	Yüzeysel geçerlik	Uzman yorumu
	Paralel form			İçerik (kapsam) geçerliği	
Ölçümcü güvenilirliği	Gözlemciler arası tutarlılık	Korelasyon, t-testi, özel varyans çözümlenmeleri ve Cronbach alfa	Ölçüte dayalı geçerlik	Eşzaman geçerliği	Korelasyon
	Gözlemciler içi tutarlılık			Yordama geçerliği	
İç tutarlılık	Yarıya Bölme Yöntemi	Pearson momentler çarpımı korelasyon katsayısı	Yapı geçerliği	Benzer ölçek geçerliği	Korelasyon
	Kuder-Richardson Güvenirlik Katsayıları	Kuder-Richardson Güvenirlik Katsayıları		Yapısal eşitlik modellemesi	Korelasyon
	Cronbach Alfa Güvenirlik Katsayısı	Cronbach alfa katsayısı		Faktör analizi	Faktör analizi

Sonuç olarak, sağlık alanında kullanılan ölçeklerin geçerlik ve güvenilirlik düzeyi hem klinik uygulamaların hem de araştırma sonuçlarının yorumlanmasında hatalara veya taraf tutmalara yol açabilecek önemli bir faktördür. Geçerliliği istenen düzeyde olmayan bir ölçek, araştırmalarda ölçümlerin heterojen hale gelmesine neden olarak ya istatistiksel testlerin gücünü düşürmekte ve gruplar arası farkın anlamlı olarak tespit edilmesini engellemekte ya da çok daha fazla sayıda deneğin araştırmaya dahil edilmesini zorunlu kılarak maliyet artışı ve zaman kaybına neden olmaktadır. Güvenilirliği düşük bir test ise yapılan ölçümün “bias”a yol açmasına ve özellikle de klinik uygulamada bazı hatalı kararlara zemin hazırlamaktadır.

Dolayısıyla kullanılan ölçeklerin geçerlik ve güvenilirliğinin iyi bilinmesi ve hatta seçilen testlerin geçerlik güvenilirliği yüksek testler olmasına özen gösterilmesi, sağlık alanında önemli bir gerekliliktir.

Ölçek geliştirme çalışmaları bilgi birikimi, araştırma, emek ve zaman isteyen çalışmalardır. Tutarlı, hatadan mümkün olduğunca arındırılmış, geçerli veriler üretebilmek ve bu şekilde, çözüm üretebilmek adına doğru saptamalar yapabilmek için “daha duyarlı ölçek” kullanabilme olanaklarının sürekli olarak araştırılması bilime yapılan katkıyı artıracaktır.

KAYNAKLAR

1. Terlemez L, İstatistik, <http://home.anadolud.edu.tr/~lterlemez/istatistik-O%C3%96%C3%96P.pdf> adresinden 24 Aralık 2013 tarihinde ulaşılmıştır.

2. Tavşancıl E. Tutumların Ölçülmesi ve SPSS ile Veri Analizi. Ankara: Nobel yayınevi, 1.baskı, 2002.

3. Deniz Z. Psikometrik Ölçüm Aracı Uyarlama.

Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi. 2007; 40(1):1-16.

4. Erkuş A. Psikometrik Terimlerin Türkçe Karşılıklarının Anlamları İle Yapılan İşlemlerin Uyuşmazlığı. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi. 2010;1(2):72-7.

5. DeVellis, R. F. Scale Development Theory and Applications Second Edition. SAGE Publications International Educational and Professional Publisher, 2003; chapter 5: 60-96.

6. Ekici E. Öğretmenlere Yönelik Bilişim Teknolojileri Öz-yeterlik Algısı Ölçeğinin Geçerlik ve Güvenirlik Çalışması. Pamukkale Üniversitesi Eğitim Fakültesi Dergisi. 2012; 31:53-65.

7. Naresh, K. M. Questionnaire Design and Scale Development” Chapter 5. http://www.terry.uga.edu/~rgrover/chapter_5.pdf adresinden 24 Aralık 2013 tarihinde ulaşılmıştır.

8. Taşkın, Ç., Akat, Ö. Araştırma Yöntemlerinde Yapısal Eşitlik Modelleme. Bursa: Ekin Yayınevi, 2010, 2.bölüm:16-26.

9. Aydemir Ö. Ölçek Geliştirme, Güvenirlik ve Geçerlik. http://www.saykad.net/wp-content/uploads/2012/11/o_aydemir1.pdf adresinden 24 Aralık 2013 tarihinde ulaşılmıştır.

10. Ercan İ, Kan İ. Ölçeklerde Güvenirlik ve Geçerlik. Uludağ Üniversitesi Tıp Fakültesi Dergisi. 2004;30(3):211-16.

11. Malhan S, Öksüz E. Sağlığa Bağlı Yaşam Kalitesi Kalitemetri. 1-5. Bölüm:29-98.

12. Büyüköztürk Ş. Veri Analizi El Kitabı. Ankara: Pegem yayınları, 2005, 5. Baskı, 1-182.

13. Yılmaz E, Eser E. Miyokart Enfarktüsü

Boyutsal Değerlendirme Ölçeği (MIDAS)
Türkçe Sürümünün Psikometrik Özellikleri.
Anadolu Kardiyoloji Dergisi. 2011;11: 386-401.

14. Şencan H. Sosyal ve Davranışsal Ölçümlerde
Güvenilirlik ve Geçerlik. Ankara: Seçkin
Yayıncılık 2005, 1.baskı:787-88.

15. Akgül A. Tıbbi Araştırmalarda İstatistiksel
Analiz Teknikleri. Ankara: Emek ofset, 2003,
2. Basım, 17. Bölüm:440-54.

16. Büyüköztürk Ş, Faktör analizi: Temel
Kavramlar ve Ölçek Geliştirmede Kullanımı.
Kuram ve Uygulamada Eğitim Yönetimi
Dergisi. 2002;32:470-83.

17. Kurnaz M, Yiğit N. Fizik Tutum Ölçeği:
Geliştirilmesi, Geçerliği. Necatibey Eğitim
Fakültesi Elektronik Fen ve Matematik Eğitimi
Dergisi (EFMED). 2010;4(1):29-49.

18. Arıkan R. Araştırma Yöntem ve Teknikleri.
Nobel yayınevi,2011,1. Basım, bölüm 6:93-7.

19. Bademci V. Tartışmayı Sonlandırmak:
Cronbach'ın Alfa Katsayısı, İki Değerli
Ölçümlenmiş Maddeler ile Kullanılabilir.
Kazım Karabekir Eğitim Fakültesi Dergisi.
2006;13:438-46.

20. Atılğan H. Eğitimde Ölçme ve
Değerlendirme. Ankara: Anıl yayıncılık, 2006,
Bölüm 1- 3:1-131.

21. Bakır O. Bilimsel Araştırmalarda Geçerlik
ve Güven. 80.251.40.59/education.ankara.edu.
tr/aksoy/eay/obakir.doc, adresinden 2 Ocak
2014 tarihinde ulaşılmıştır.

22. Güvenirlik, bölüm 8. www.aoa.edu.tr/
ozder/files/2.2.GÜVENİRLİK.doc adresinden
24 Aralık 2013 tarihinde ulaşılmıştır.