

Psikiyatride Psikometri: Temel Kavramlar

Psychometrics in Psychiatry: Basic Concepts

Hüseyin Güleç¹

¹ Uzm. Dr., Erenköy Ruh ve Sinir Hastalıkları Hastanesi, İstanbul

ÖZET

Bir doğa olayı olan insan zihninin süreçlerini anlamak pozitif bilimin ana uğraş alanıdır. Ölçüm yapma da –ölçme- pozitif bilimin temel özelliği konumundadır. Dar anlamıyla psikometriye zihin ölçümü dersek, ölçeklere de bir tür “zihinmetre” aracı diyebiliriz. Bilim, ölçtüğümüz kadarını iddia edildiği bir oyun olarak kabul edilirse bu oyunun kural kitabını net yazabilmeli ve güncelleyebilmeliyiz. Bugün ölçtüğümüzü iddia ettiğimizizin yarın başka bir şey çıkması bilim tarihinde sık görülen ve olması gereken bir durumdur. Ölçülen olgunun yeterince temsil edilmeden oluşturulması kadar, başka olgularla karışması yani “kirlenmesi”, sınır dâhilinde netleştirilmesi bu kitabın içinde ele alınmalıdır. Bu açıdan bakıldığında ölçeklerin geçerliği ve özellikle yapı geçerliği, temel kavramsal çatıyı oluşturmaktadır. Geçerlik yapısının sağlanması öncesinde ve sonrasında pozitif bilime düşen önemli başka işler de vardır. Geçerlik öncesi güvenilirliğinin gerçekleştirildiği gösterilmektedir. Güvenirlik – yani “nasıl” ölçtüğünüz- zayıf ses verirse geçerlik sinyallerini yeterince alamazsınız. Olgu ölçümünde aynı yönergenin tüm katılımcılara aynı şekilde uygulanması standardizasyon içinde önemli başka bir aşamadır: Tüm dünyada aynı dili kullanmak için aynı ölçeği kullanmamız gerektirdiğini ve dolayısıyla o kültürde uygun dil ve özelliğe getirilmesinin yani uyarlama işleminin önemini ortaya koymaktadır. İlgili bütün değişkenlerin göz önüne alındığı sağlıklı ve hasta gruplardaki normative değerlerin elde edilmesi bir diğer işlemdir. Tüm bu işlemlerin hepsi standardizasyon adı altında toplanır. . Bu yazıda “meraklı” yeni bireylerin, bilimsel bir makalenin okunması ve yorumlanmasında ve/veya kendisi bir araştırma içindeyken, kullanılan ölçeklerin psikometrik özelliklerinden olan geçerlik ve güvenilirliği daha iyi tanımları ve bir ölçeğin standardizasyonu içinde yer alan özelliklerin altının çizilmesi amaçlanmıştır.

Anahtar Sözcükler: psikometri, ölçüm, ölçek, standardizasyon, geçerlik, güvenilirlik

ABSTRACT

Understanding human mental process as a natural event is the mainstay for positive science. Measurement or scaling these processes is the basic component of positive science. If we call the science of psychometrics as “mind measuring”, then the scales should be remarked as the tools of “mind meter”. Besides, if science is considered as a play in which your assertions are only within the limit of your measurements, and then the book of rules ought to be written precisely and updated regularly. In the history of science, there are many incidents in which different tools which had been marked to measure several variables or things were later found to be unable to implement the measurement it had been assigned. Thus the book of rules should set the boundaries of measurements, where to start and to finish, by preventing any confounding factors. Validity and especially construct validity of the scales set the basic conceptual framework. But, positive science should present the proof of reliability before providing the validity. Reliability –how you measure- is considered to put an upper limit on validity because any problem in reliability makes it harder to detect a valid “signal”. Another important step in measurement is to

apply same directives to all participants to assure standardization. Thereby, we could ensure that the entire world keep same standard in the evaluation and application of the scale. Other process is to obtain normative value according to related variables in order to prevent confounding possible cultural factors. All of these steps have been comprised under the title of standardization. With this article, we would like to make sense to the reader concerning what is meant by psychometric properties of scales and focus on the process of standardization and its components.

Key Words: psychometric, measure, scale, standardization, reliability, validity

Psikiyatride Güncel Yaklaşımlar-Current Approaches in Psychiatry 2009; 1: 175 - 186

Çevrimiçi adresi/ Available online at : www.cappsy.org/archives/vol1/

Çevrimiçi yayım tarihi / Online publication date: 01 Eylül 2009 / September 01, 2009

Yazışma Adresi / Correspondence

Dr. Hüseyin Güleç, Erenköy Ruh ve Sinir Hastalıkları Hastanesi, Turkey

E-mail: huseyingulec@yahoo.com

Bir bürokratin, yöneticileri tarafından işyerinde psikolojik baskıya uğradığı iddiasıyla mahkemeye başvurduğunu belki duymamış olabilirsiniz. Gerçi Türkiye’de mahkemeye giden ilk olgu olduğu düşünülürse Türkiye’de çok duymuş sayılmazdı. Ancak son zamanlarda sıkça duyulur ve tartışılır oldu, hatta iş yerinde yıldırma (mobbing) tabirini gündelik kullanıma indirger hale geldik. Acaba daha önce işyerinde yıldırmaya uğrayan olgu ya dab u tür taktikler yok muydu? Muhtemelen vardı ve adı konmamıştı, hatta yapan da (belki) farkında değildi ve onun da kendisine göre olağan, deneyimle edindiği, kendi bulduğu (belki de övündüğü) taktiklerindendi. J. P. Sartre, “siyahlar eziliyor, demedikçe siyahların ezilmesi bir şey demek değildir. O ana kadar kimse, belki siyahların kendileri de farkında değildir. Bir tek sözdür ona bu anlamı kazandıran” sözünü belki de bu yüzden etti.

Bir doğa/evren olayını -özellikle en karmaşık olanlardan biri olan insan zihninin süreçlerini- anlamak, pozitif bilimin emek sarf ettiği alandır. Bilim bunu yaparken “katı” kurallara uymak durumundadır. Bir sürecin pozitif bilim kapsamında sayılabilmesi için gözlemlenebilir, tekrarlanabilir, ölçülebilir olması ve operasyonel olarak tanımlanması ve iletilenin sağlanabilirliğinin gösterilmesi gereklidir.[1] Bu çalışma bu özelliklerin temeli konumunda olan ölçme üzerine bir tartışma niteliğindedir.

Ölçme ve Psikometri

Ölçme, bireylerin ya da nesnelerin belirli özelliklere sahip olup olmadığının, sahipse sahip oluş derecesinin belirlenerek sonuçların sembollerle (özellikle sayı olarak) ifade edilmesidir. Psikometrik kelimesi ise anlam olarak birçok ölçme tarzının performansını tanımlamak için kullanılır. İki temel psikometrik özellik güvenilirlik ve geçerliktir. Ölçmenin sınırlarını daha netleştirmek ve kafa karışıklığını önlemek için değerlendirmeden ayırmak gerekir. Değerlendirme, ölçme sonuçlarını bir ölçütle kıyaslayarak ölçülen nitelik hakkında bir karara varma sürecidir. Sınırları daha netleştirmeye devam edersek, ölçme tanımlama, değerlendirme ise yargılama işlemidir ve ölçme sonucunun bir ölçütle

karşılaştırılmasına dayanır. Bu çalışmada da zaman zaman karışması engellenmemiştir.

Ölçmenin tanı konmasında yeri önemlidir. Yukarıda da belirtildiği gibi pozitif bilim ve dolayısıyla ölçme, öznel bir alan olan klinik görünümü bilimsel kapsamın içine çekmede rol oynar. Karşıt olarak, tanı sistemleri de ölçme yöntemlerinin gelişmelerinde önemli rol oynamaktadırlar. DSM tanı sistemi fenomenolojik olarak sendromları operasyonel olarak tanımlanma dönemine girince –kendi evrimini gerçekleştirince- buna bağlı olarak yapılandırılmış klinik görüşme ve değerlendirme ölçekleri de geliştirilmiştir. Bir şekilde ölçekler tanıya destek sağlarken, tanı sistemleri de ölçeklerin oluşmasına neden olmaktadır. Ölçeklerin kavramsal ile teknik yönü mevcuttur. Ölçeklerin kavramsal yönünü bir başka deyişle oluş nedenini anlamaya çalışalım.

Kavramsal Yapının Ortaya Çıkarılması

Kavramsal olarak ölçeklerde yapı veya yapılar (construct/s) tanımlanabilir. Her bir yapı tanım kümesi, çoklu küme veya alt-kümelere (veya alan=domain) içerebilir. Bu tanı kümeleri birbirlerine bağımlı veya tamamen bağımsız boyutlar (dimensions) veya eksenler (axes) halinde görülebilir. Ölçeklerde her bir küme veya boyuttan elde edilen puanların da geçerlik ve güvenilirliği istenir. Bu yapıardan elde edilmek istenen; tanı, bulgu ve belirtiler, şiddet, bozulma, fonksiyonel kapasite ve yaşam kalitesi gibi değişkenlerin değerlendirilmesidir. Buraya kadarki açıklamalar netleşme yerine kafa karışıklığı yaratabilir. Belki de yapı, küme, boyut ve eksen kavramlarını tahtada yerleri farklı ayrı satranç taşları olarak örneklemek anlamamıza daha fazla yardımcı olabilir.

Başka bir yönüyle, yapı içinde iki temel yaklaşımdan bahsetmek de gerekir. Kategorik sınıflandırma (categorical classification), bir nitel değerlendirme tarzı olup araştırılanın var olup olmadığına bakar. Psikiyatride özellikle tanı için kullanılmaktadır. Sürekli ölçüm (continuous measures) ise nicel değerlendirme tarzıdır. Semptom şiddetinin ve fonksiyonel durum seviyesinin tespiti en sık kullanıldığı yerlerdir. Tamamen ayrı görünen bu iki kavramsal yapı aslında tam olarak ayrılamazlar da. Örneğin, kategorik yapı (katılmıyorum, az katılıyorum, çok katılıyorum, tamamen katılıyorum gibi) içinde olan ordinal sınıflandırma ikisi arasında durmaktadır. Likert tipi bir sınıflandırmanın -10 veya daha fazla sayıda işaretlendirme örneğinde- daha efektif biçimde süreklilik içinde olduğuna yorumlanabilir. Bir örnek de sürekli değişkenlerin kesme puanları sayesinde kategorik yapıya yaklaşmasıdır. Mini Mental'den 23 puan altında alınması demansla uyumlu sayılması örneği gibi. Belki de kategorik tanıların nicel eşiklerden elde edilmesi de (Major depresyon tanısı için 2 hafta süreyle 5 depresyon belirtisinin varlığının gerekliliği gibi) son bir örnek olarak gösterilebilir.[2]

Sartre geçmişse, şikayetçi olan bürokrat günümüzde, herhangi biri (kuvvetle muhtemelen düşünür, sanatçı veya bilim adamı) gelecekte, tanımlanmış veya tanımlanacak olguları karşımıza çıkarmıştır ve çıkaracaktır. Daha önce tanımlanmadıysa kimse ez(il)meyi anlayamayacak ve olguları açıklayamayacaktır. Artık biri çıktı ve siyahlar eziliyor dedi, bana işyerimde psikolojik baskı var dedi. Tanımlanma -doğa veya evren olayı olarak ele alınıp bilimden yar-

dim istenirse ya da bilim kendi tanımlamasını kendi yaparsa- aşamasında artık “olgu”, gözlemlenilebilmeli ve ölçülebilinmelidir. Belki de tekrar yeni tanımlamaları yapılarak, neye yıldırma denileceğinin, neye denilemeyeceğinin sınırları yeniden belirlenmelidir. İşyerinde baskı sonrası, komşusuyla akşam kavga edip bu nedenle ceza alan çalışan, ödemesi gereken tazminatın ne kadarını işvereninden istemelidir. Bu kişi siyah biri olsun ve eve gelirken “beyaz” insanlar tarafından ezilirse; hangisinin yıldırma hangisinin ezilmekten ortaya çıktığını nasıl anlayacağız. Ya da, Sartre çıkıp siyahların ezilmesini işaret ettiğinden fark ettiğimiz veya işyerinde çalışanların üzerinde varlığını keşfettiğimiz baskının daha öncesine sahip çıkması gereken bilimsel vicdanımızdan dolayı, geriye dönük uygulamalarından ötürü iktidarları devirip, mağdur çocuklarının yeterli eğitim alamadığından ötürü tazminat da hak ettiklerini öne sürebilir miyiz? Yoksa bu da yeni başka bir kavram olabilir mi? Algılanan baskı kavramı “olgu”su olabilir mi? Bu “abartılı” sorular gözlemin ve bunu kavramsal hale getirmenin bilimsel önemini altını çizmek için yazılmıştır. Bu olguların tanım olarak yapısının oluşturulmasının önemi gösterilmeye çalışılmaktadır ve nelerin karşımıza çıkabileceğini öngörmemiz gerekebilir. Operasyonel tanım sonrasında da, bunu uygun materyallerle ölçmek gerekecektir. Bilim ayrıca bulgularımızın tekrarlayabilirliğini ve iletilme sonrası sağlanabilirliğini görmeye çalışacaktır.

Kavramsal Yapının Tanı Sistemlerine ve Ölçek Gelişimine Etkisi

“Ne kadar mutsuzsunuz?” sorusu depresyonu ölçer mi, “İşyerinde ne kadar çok psikolojik baskı görüyorsunuz? (1) hiç, (2) çok az, (3) az, (4) orta derecede, (5) fazla, (6) çok fazla, (7) şimdiki kadar gördüğüm baskıların en şiddetlisi”... Bu soru(lar) depresyonu veya moobingi ölçer mi? Bu noktada, daha önce evrim geçirdiği iddia edilen DSM'nin bu sorunun cevaplanmasında yardımcı olabilir. DSM ilk tanımlamalarını yaparken kullandığı paradigma Meyer'in etkisindeydi. Tanısal anlaşmazlığa açıklık ve kavramsal olarak geniş tanısal yaklaşım vardı. Daha geniş ve daha az yapılandırılmış yaklaşım olduğundan tanı sistemleri geniş bir grubu (şizofreni örneğinde; Bleuler'in yaklaşımı tanıda belirsizliği azaltmıyor ve psikotik özellik gösteren depresyon da şizofreni tanısı alabiliyordu) tanı sistemine dahil etmeye meyilliydi. Bu tanı sistemini kullanan Amerikalılar, Avrupalı meslektaşlarından daha az geçerli şizofreni tanısı koyuyorlardı. Tanı sisteminde operasyonel ölçütler DSM-III ile değiştirilmeye ve diğer tanı sistemlerine yaklaşılmaya başlandı ve bilginin paylaşılmasına, aynı dilin daha fazla konuşulmasına başlanılmış oldu.

Yukarıdaki soruya tekrar bakalım. Depresyonun tanı sisteminde operasyonel tanımlanmasını ve ona uygun görüşme yöntemlerinin, tüm dünyada benzerinin hatta aynıısının kullanılmasını ve tanımlanan özelliklerin “ortak” ölçeklerle şiddet değerlendirmesinin sorulması, bu değişimin devamı niteliğindedir. Depresyon sorusunu tekrar göz önüne alalım, muhtemelen bu soru bir şeyi ölçecek ve bu tamamen yanlış olmayacak ama yeterli olup olmadığını tartışmak, geçerlik için daha doğru hale gelmiştir. Kavramsal platform-

da ana amaç geçerlik gibi görülmekte olup, bu değişken ölçtüğümüz “şey”in iddia ettiğimiz “şey” olup olmadığını sorar ve bu sorunun net cevaplarını bulmaya çalışır. Kişi işyerinde yıldırılmaya çalışılmıştır, bu soruya 7 puan verir, ama depresyonu olmayabilir mi? Depresyonun ağırlığından evinden çıkamaz hale gelmiş, işini ve arkadaşlıklarını umursayacak gücü bile kalmamıştır ve son enerjisi ile intihar mektubunu yazmıştır ve 6 puan verir, daha önce daha şiddetli bir atağı vardır ve tedaviden bir dönem iyi cevap görmüştür.

Başka bir örnek, son derece başarılı, sağlıklı ve mutlu biri sadece geçmişinde bitiremediği bir olayın son zamanlarda hesaplaşmasını yapmaktadır ve kendini çok mutsuz hissettiği bir anda bu soruya 7 puan verir. DSM-III’ ten sonra tanı sistemleri bu soruya bugün için en az 9 soruyla yaklaşabileceğimizi ön görmektedir. Depresyonun tanımlanması aşamasında, bir çalışma grubunun bu klinik görünümü dört boyutlu olarak bulduğunu iddia edelim. Bunun doğruluğu ölçüsünde, başka bir araştırma grubu ölçek yardımıyla bu bulguyu destekleyecek veya çürütecektir. Çok iddialı gelebilir; ölçek, depresyon “olgu”sunu, kavramsal olarak dört boyutlu bir yapı olduğunu destekleyecektir. Ya da farklı buldu, daha önceki bulgunun tekrarlanmadığını belki de sadece o kültüre ait bir durum olduğunu söyleyecektir. Daha önce de belirttiğimiz gibi tanı sistemi ölçeğin gelişmesinde sebep olmuştur ancak bu süreç bitmemiştir. Bu bulgunun önemi, diyelim aynı kavramsal yapıyla bir asır önceki depresyon olgusunu araştırıyorduk, ölçme araçlarımızda daha fazla psikomotor aktivasyon değişiklikleri belki de katatonik, somatik yakınma yer alacakken, bugün bu bulguları çok görmediğimizden, motor yavaşlamayı 2 hatta daha fazla soruyla araştırırsak, tanıda kafa karışıklığını arttırabiliriz. Soru havuzu bugünkü duruma göre; hasta profilini ve hastaların çekmekte olduğu bulgu ve belirtilerin daha fazla veya az yer almasını, işaret edecektir. Cinsiyet, yaş, eğitim hatta kültürel yapı aynı olgunun farklı görünümlerine etkide bulun(ma)duğundan, ölçeği geliştirenler bunlara daha dikkat edecektir. Bu da bize ölçeklerin standart uygulamalarında her kültürde, olguyu etkileyen değişkenler ölçüsünde uyarlanması gerektiğini hatırlatacaktır. Buradan hareketle dikkat edilmesi gereken değişkenlerin hesaba katıldığı normative veriler elde edilmeye çalışılacaktır. Psikiyatri dışı tıbbi hastalara vereceğimiz ölçek ile depresyon tedavisindeki ilaç cevabında depresyon şiddetinin değişimini fark eden ölçeklerinin aynı olmamasını kulağımıza fısıldayacaktır.

Bugün psikiyatride tanı sistemleri hâkim paradigmasını geliştiremediği için tanılandaki geçerlik ve güvenilirlik sorunu devam etmektedir. Bugün hala çok iyi gözlemlenmesi ve farklı operasyonel tanımlanması yapılması gereken alanlar mevcuttur. Heterojenitesi olduğu en fazla kabul gören şizofreni örneğini, kavramsal yönüyle tartışırsak belki Andreasen’in şizofreni için geliştirdiği Pozitif Belirtileri Değerlendirme Ölçeği (SAPS) ve Negatif Belirtileri Değerlendirme Ölçeği (SANS)’nin oluşmasını da anlarız, hatta konunun daha iyi anlaşılmasında bize yardımcı dahi olabilir. Bleuler hastalığının disosiyatif temelli olduğu paradigması ile şizofrenin tek bir hastalık olduğunu iddia etmiştir. Bozukluk disosiyasyon değildir, kökünde, natürde bozulan neyse bunun disosiyatif şekilde bozulduğunu iddia etmiştir. Buna göre, olan kapasite

(uyum sağlama gücü) üzerine eklenen çevresel etkileşime (beklenti ve baskı ya da tonus) bağlı olarak bireyde sanrı, varsanı, izolasyon ve azalmış dürtü gibi ikincil fenomenlere (hastalık belirtileri) yol açar. Bu yönüyle hastalık modeli olarak Tip-I diabetes mellitusta (DM) olduğu gibi tek hastalık modelini temsil etmektedir. İnsülin metabolizması bozukluğu temel sorunken hastalık belirtileri olan ikincil fenomenler organ tutulumuna özgüdür. Bleuler şizofreni belirtilerini, temeline asosiyasyon kaybını yerleştirdiği çok bilinen 4A belirtileri ile ifade etmiştir. Bu model sonradan yapılan çalışmalardan gelen kanıtlarla desteklenseydi, çok rahat birçok klinik değerlendirme ölçeğinin bugün var olmayacağını söyleyebilirdik. Doğal olarak da görüşme ve değerlendirme ölçeklerinin bu dört belirtinin sorgulanmasından geçeceği söylenebilirdi.[3]

Bu görüş kanıtlarla desteklenmemiştir ve hastalığa özgü risk faktörlerinin farklılığından gelen veriler, farklı genetik etkilerin varlığını destekleyen veriler, klinik görünüm ve tedaviye yanıtın hatta progresin heterojenitesi ile veriler nedeniyle, hastalık modeli olarak klinik sendrom şeklinde kavramsallaştırılmasını destekler. Bu yönüyle Tip-I DM'den daha çok mental retardasyon örneği hastalık modeli için daha uygundur. Son yıllardaki çalışmalar, mental retardasyonun birçok hastalıktan oluşmuş bir klinik sendrom olduğunu göstermiştir.[3]

Şizofreni hastalık dışı modeller olarak deli bir dünyaya akılcı bir tepki (sosyal tepki teorisi) veya toplumun kendi ürettiği kendine aykırı olanlarla bir başa çıkma yetisi gibi modellerle de kavramlaştırılabilirdi. Şizofreninin biyolojik akrabalarındaki yaygınlığı, erken gelişimsel risk etkenleri, eşlik eden işlevsel ve yapısal beyin bozuklukları, ilaçlara cevap, yaşam boyu yaygınlık gibi verileri bu modeller açıklamaktan uzak ve yetersiz kalmakta ve indirgemeci olduklarından artık dikkate alınmamaktadır. Sadece psikolojik veya sosyal etmenlere indirgemek daha çok geniş model açıklamaların gerektiği modelden uzak kalmaktadır. Sonuçta, bu olguya farklı bir gözlemsel yaklaşım, önce dikkate alınmış fakat yöntem ve ölçüm bu kavramsal çatıya şimdilik onay vermemiş diyebiliriz.[3]

Şizofreni örneği bizlere; tanı sistemleri, yöntem oluşumu, ölçek seçiminin, bilimsel araştırmalarda nasıl birbirleriyle iç içe girebildiğini ve hangisinin neyi ön gördüğüyle uğraşmaktan çok, birbirlerine nasıl destek verdiklerini göstermektedir. Sorun hiç düşünülmediği kadar basit değildir ve kafa karışıklığı belki de artmaktadır. Ölçülen olan değil, ölçenin olguya kendi kurduğu sistemin kavramsal yapısının boyutları, kümeleridir. Olanı ölçmeye çalışır, olmayana ne yapacağı belirsizdir. İşyerinde yıldırımın duyarlılık yoluyla baş eden biri bu yıldırımı yaşamıyorsa işyerinde yıldırma yok denilebilir mi? Duyarsızlığa karşı bizi ne, nasıl uyaracak? Duyarsızlık, istenç kaybı, duygusuzluk nasıl ölçülür. Depresyonunu somatize eden, etmeyene göre "tıkanmış" mıdır, başka bir ağı mı kullanıyordu? Şizofrenide, çalışan bellek ödevinde normal performans yaparken başka bir ağ kullandığını ölçtüğümüzde, benzerlik acaba hem bozukluk hem de sapmışlığın beraberliğiyse bunu nasıl değerlendireceğiz? Sorular, kafa karışıklıkları kavramsal yapının daha ileri işlemlere ihtiyacına ait sinyallerdir.

Ölçmede hatadan kurtulmak imkânsızdır. Elinize aldığınız “olguyu” ölçerken; zaman, yer, kişi faktörlerinden etkilemesinin en aza indirilmesi ve kısıtlılıkların –yaptığınızdan mümkün oldukça en fazla- farkında olunması ve ölçüm için kullandığınız cetvelin daha sonraki uygulamalarınızda ve tüm diğer uygulamalarıyla benzer şekilde uygulanmasının anlam ve önemini altı çizilmeye çalışılmıştır.

Standardizasyon

Kavramsal tartışmanın kendi kısıtlılığını akılda tutarak bir sonraki aşamada ölçeklerin standart uygulamaları yapma garantisini verme derecesini ne derece sağladığından söz edebiliriz. Bu yolla bilim iletilen bulguların tekrarlanmasına ve aynı zamanda sağlanmasına izin verebilecektir. 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 yorum hatalarını en aza indirmeye çalışır.[4] Ölçeğin standardize olabilmesi ve sonrasında uygun bilgiler üretme yeteneğine sahip olması için “güvenirlilik” ve “geçerlik” olarak nitelendirilen iki özelliğe sahip olması istenir.

Güvenirlilik

Ölçeğin taşınması gereken özelliklerden birisi olan güvenirlilik, bir ölçme aracıyla aynı koşullarda tekrarlanan ölçümlerde elde edilen ölçüm değerlerinin kararlılığının bir göstergesidir.[5] Ölçümlerin kararlılığının yorumlanmasında, iki farklı yaklaşım vardır. Birinci yaklaşım, birbirini izleyen ölçümlerde, 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 etmesinin gösterilmesi ile ilgilidir.[6] Güvenirlilik sadece ölçme aracına ait bir özellik değildir, ölçme aracı ve aracın sonuçlarına ilişkin bir özelliktir. Bir ölçme aracının ölçtüğü özelliği doğru olarak ölçebilme derecesidir. Ölçme aracının ölçtüğü özelliği ya da özellikleri ne derecede bir kararlılıkla ölçmekte olduğunun göstergesidir.[7] Bir ölçme aracının ölçümlerine birçok kaynaktan hata karışabilir. Güvenirlilik ölçme aracının hatalardan arınlık düzeyidir. Ölçme sonucunda hata miktarı düştükçe güvenirlilik artar. Hata miktarı arttıkça güvenirlilik de düşer.[8,9]

Ölçme araçlarının güvenirlilik düzeyleri güvenirlilik analizi ile belirlenir. Güvenirlilik katsayısı 0.00 ile +1.00 arasında değer alır. Ölçme aracının güvenirlilik katsayısı 1'e ne kadar yakınsa güvenirlilik o kadar yüksek, 0'a ne kadar yakınsa güvenirlilik o kadar düşük olur. Güvenirlilik katsayısı hiçbir zaman tam 0.00 ya da tam +1.00 olmaz. Çünkü ölçme sonuçlarına az da olsa bir miktar hata mutlaka karışır o yüzden tam +1.00 değer almaz. Aynı şekilde ölçmelerin tamamına hata karışamayacağı için tam 0.00 değeri almaz. Ölçme araçlarının güvenirlilik düzeylerini belirlemek için kullanılan bazı yöntemler şunlardır;

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 uygulanmasıdır. İki uygulamadan elde edilen ölçüm değerleri korelasyon katsayısı ölçeğin güvenilirlik katsayısıdır.[10]

Paralel (Eşdeğer) Formlar Yöntemi (Parallel-Forms Method)

Paralel formlar yönteminde aynı davranış kalıplarını temsil edebilecek farklı maddeler örneklenerek iki eşdeğer form oluşturulmaktadır. İki formun eşdeğer olabilmesi için, formların kapsamlarının, yapısının, zorluk derecesinin, talimatların, puanlamanın, madde sayısının ve yorumlamanın aynı olması gerekir.[5] Eşdeğer iki form aralıksız olarak aynı anda ya da aralıklı olarak farklı iki zamanda uygulanır. Formlar arasındaki korelasyon hesaplanır ve güvenilirlik katsayısı olarak yorumlanır.[5]

İç Tutarlılık (Internal Consistency)

Bir ölçeğin bir kez uygulanmasıyla güvenilirlik tahmini yapılıyorsa, diğer güvenilirlik tahmini yöntemlerine göre, güvenilirlik tahmininde meydana gelebilecek hata daha az olacaktır. Yarıya Bölme Yöntemi (Split-Half Method), formu iki eş parçaya bölerek, iki yarının deneklere aynı anda uygulanması sonrası, deneklerin yarılarından aldıkları puanlar arasındaki korelasyon ile güvenilirlik tahmini yapılmasını sağlar. İçsel tutarlılığın güvenilirliğini belirlemek için en sık kullanılan yöntemlerden birisi de Kuder-Richardson yaklaşımıdır. Yöntem tüm maddelerin birbirleriyle ve ölçeğin tamamıyla iç tutarlılığını tahmin etme amacı üzerine kuruludur. Bu yöntemle, iç tutarlılığa yönelik güvenilirlik kestiriminde bulunmada belirli ölçütler dikkate alınarak Kuder-Richardson 20 veya 21 formlerinden uygun olanı kullanılır.

Cronbach Alfa Güvenirlik Katsayısı, maddelerin 1-3, 1-4, 1-5 gibi puanlandığında, kullanılması uygun olan bir iç tutarlılık tahmin yöntemidir. Cronbach alfa katsayısı, ölçekte yer alan maddelerin varyansları toplamının genel varyansa oranlanması ile bulunan bir ağırlıklı standart değişim ortalamasıdır.[10]

Geçerlik

Uygun bir yöntemle, ölçme aracının güvenilirliği saptansa bile, güvenilirliğin 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. Bu nedenle, davranışsal özellikleri ve bunlardan da özellikle bilişsel ve duyuşsal yönü baskın olanları ölçerken kullanılan ölçme aracının, ölçmek istediğimiz özelliğe yönelik ölçme dereceleri araştırılmalıdır. Ölçme araçlarının geçerliğinin dayandığı gerekçe: Doğrudan ölçme yapılabilen somut özelliklerde, amaca hizmet gücü yüksek olan ölçme araçlarıyla çalışılmaktadır. Ölçülmek istenen özellik soyutlaştıkça amaca hizmet gücü olmayan ya da bu gücü sınırlı olan ölçeklerle çalışma tehlikesi artar. Ölçümlerin amaca hizmet edebilmesi,

ölçme aracının ölçülmek istenen değişkeninin ölçüsü olabilecek ölçümler vermesine bağlıdır.[11,12]

Geçerlik, bir ölçme aracının ölçmeyi amaçladığı özelliği, başka herhangi bir özellikle karıştırmadan, doğru ölçebilme derecesidir. Geçerliğin davranışsal özelliklerin ölçülmesinde önemli bir yeri vardır. Davranışsal özelliklerin ölçülmesinde kullanılan ölçüklerin hiç birinin geçerliği tam değildir. Fakat ölçüklerin geçerliği yeterli sayılabilecek doğrulukta ölçümler verecek düzeye getirilebilir. Bir ölçme aracının geçerliği, 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.

Ölçümlerdeki değişkenliğin tümünün bireylerin ilgili özelliğe sahip oluş dereceleri arasındaki gerçek farklardan gelmekte olması demek, ölçüğün başka hiçbir özelliğin etkisinden etkilenmemiş olması ve ölçme hatasının tam olarak sifira indirgenmiş bulunması demektir. Ölçme hatasının tam olarak sifira indirgenmesi imkânsızdır. 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. Bir ölçüğün geçerliği, daha önce de belirtildiği gibi bir derece problemdir, yani ölçüğün geçerliği vardır veya yoktur şeklinde düşünülemez. Bu nedenle ölçüğün geçerliğini özel bir amaç dışında yüksek, orta ve düşük olarak da nitelememek gerekir. 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 ölçütlerinin yeterince duyarlı olmaması nedeniyle, geçerlik daha düşük olabilmektedir.

Ölçüğün geçerlik düzeyi onun geçerlik katsayısının hesaplanmasıyla anlaşılır. Geçerlik katsayısı, ölçekten elde edilen değerlerle ölçüğün 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. 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. Bu nedenle geçerlik katsayıları güvenilirlik katsayılarıyla birlikte yorumlanır. Bir ölçmenin geçerli sayılabilmesinin ilk koşulu onun güvenilir olmasıdır. Güvenirlik, geçerlik için gerekli koşul olmasına rağmen, yeterli koşul değildir. Ölçüm sonuçlarının ne kadarıyla ölçüm hatasını yansıttığı sorusuna yanıt verebilmek için ölçme aracının geçerliğinin saptanmasına gerek vardır.[11,12]

Kapsam Geçerliği (Content Validity)

Kapsam geçerliği, bir bütün olarak ölçüğün ve ölçekteki her bir maddenin amaca ne derece hizmet ettiği. Bir ölçüğün kapsam geçerliği mantıki yoldan ve istatistikî yoldan olmak üzere iki türlü incelenebilir.[13] Mantıki yoldan ölçüğün kapsam geçerliğini araştırmak; bu yaklaşımla ölçüğü uygulamadan, ölçüğün 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. Diğeri ise istatistikî yoldan ölçeğin kapsam geçerliğini araştırmaktır. 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.

Kriter (Ölçütsel) Geçerliği (Criterion Validity)

Bu geçerlik türü, ölçeğin etkinliğini belirlemek amacıyla, ölçekten elde edilen puanlarla belirlenen kriter arasında, gelecekteki veya o andaki, ilişkiyi inceler. Kriter geçerliğin, yordama geçerliği (predictive validity) ve uyum geçerliliği (concurrent validity) olmak üzere iki alt grubu vardır.[14]

Yordama (Kestirimsel) Geçerliği (Predictive Validity)

Bir ölçeğin kestirimsel geçerliği, o ölçekten elde edilen kestirimsel puan ile ölçülmek istenen özellikleri ölçtüğü bilinen kriter arasındaki korelasyonun hesaplanmasıyla elde edilir.[11]

Uyum Geçerliği (Concurrent Validity)

Eşzamanlı olarak geliştirilen ölçekten elde edilen puanlarla, belirlenen kriter arasındaki korelasyon uyum geçerliği olarak değerlendirilir.[11]

Yapı Geçerliği (Construct Validity)

Yapı birbirleriyle ilgili olduğu düşünülen belli kümelerin ya da kümeler arasındaki ilişkilerin oluşturduğu bir örüntüdür. Bir ölçeğin yapı geçerliğini belirleme süreci bir ölçüde, bilimsel kuram geliştirme süreciyle aynıdır. Yapı geçerliğini belirleme süreci aşağıdaki adımları kapsar.[12] Kavramsal yapıyı bir daha burada tekrarlırsak; ölçmede kullanılacak yapının tanımlanması, yapının altında yatan teoriden ölçek performansı ile ilgili sınanabilir hipotezler çıkarılması ve çıkarılan hipotezleri sınamak için deneysel ve istatistikî çalışmalar, yorumlamalar yapılması. Yapı geçerliği belirlenirken birçok yönetime başvurulabilir. Faktör analizi ve hipotez testi yapı geçerliğini belirlemek için sıklıkla kullanılan iki yöntemdir.

Faktör Analizi: Aynı yapıyı ya da niteliği ölçen değişkenleri bir araya toplayarak ölçmeyi az sayıda faktör ile açıklamayı amaçlayan bir istatistiksel tekniktir. [12] Faktör analizi, başlıca amacı aralarında ilişki bulunduğu düşünülen çok sayıda değişken arasındaki ilişkilerin anlaşılmasını ve yorumlanmasını kolaylaştırmak için daha az sayıda temel boyuta indirgemek veya özetlemek olan bir grup çok değişkenli analiz tekniğine verilen genel bir isimdir. Diğeri bir ifade ile faktör analizi, aralarında ilişki bulunan çok sayıda değişkenden oluşan bir veri setine ait temel faktörlerin (ilişkinin yapısının) ortaya çıkarılarak araştırmacı tarafından veri setinde yer alan kavramlar arasındaki ilişkilerin daha kolay anlaşılmasına yardımcı olmaktır.

Faktör analizinde araştırmacı öncelikle araştırma bağlamında kullanılan değişkenler setinin temelini oluşturan ana (temel) faktörlerin neler olduğu ve

bu faktörlerden her birinin değişkenlerden her birini açıklama derecesini de görme imkânına sahip olmaktadır. Bu yolla, araştırmacı elinde bulunan çok sayıda değişkenden oluşan değişkenler setini daha az sayıda yeniden oluşturulmuş değişkenler (faktörler) cinsinden ifade etme ve anlama imkânına sahip olacaktır. Etkin bir faktör analizinin toplam veri setini en iyi temsil edebilen ancak mümkün olduğunca az sayıda faktörden oluşan bir çözüm olması arzu edilen özelliktir. İyi bir faktör analizi sonucunun hem basit hem de yorumlama becerisinin iyi olması arzu edilir. Faktör analiziyle ilgili teknikleri Açıklayıcı Faktör Analizi ve Onaylayıcı Faktör Analizi olmak üzere iki başlık altında incelemek mümkündür. Açıklayıcı faktör analizinde araştırmacı araştırma yaptığı konuyla ilgili olarak değişkenler arasındaki ilişkiye yönelik olarak herhangi bir fikrinin veya öngörüsünün olmaması sebebiyle değişkenler arasındaki muhtemel ilişkiyi ortaya çıkarmaya çalışır. Onaylayıcı faktör analizinde ise araştırmacı tarafından daha önceden belirlenen bir ilişkinin doğruluğunu test etmek amaçlanmaktadır.[12]

Hipotez Testi: Benzer ölçeklerden elde edilen puanlar arasındaki farkın anlamlılığın belirlenmesidir.[11]

Görünüş Geçerliliği (Face Validity)

Ölçme aracının neyi ölçtüğünü değil de neyi ölçer göründüğünü belirtmektedir. Bir ölçeğin görünüş geçerliliği, o ölçeğin ölçmek istediği özelliği ölçüyor gözükmesidir. Ölçeğin görünüş geçerliliğinin bazı durumlarda yükseltilmesi bazı durumlarda ise gizlenmesi gerekir.

Bir ölçeğin/testin kullanışlığı -onun geliştirilmesinin, çoğaltılmasının, uygulanmasının kolay ve ekonomik olması-, dikkate alınması gereken bir durumdur. Uygulama ve puanlamanın standart olması, başka toplum ve dilde geliştirilmiş ölçme araçlarının, kullanılan kültüre uyarlanması gerekliliğinin altı ayrıca çizilmelidir. Aynı ölçekle farklı şartlar altında, farklı yönergelerle farklı sonuçların alınması bu şekilde engellenmiş olacaktır. Tüm dünya "aynı dili" kullanma amacıyla geliştirilen ölçeklerin kendi kültüründe uygulanması amacıyla yürütülen uyarlama çalışmaları özel öneme sahiptir. Bu konu sadece bir çeviri yapmak ve örneklem(ler)e uygulamak değildir. Sözel veri dışında sözel olmayan verinin de kültürden etkilenmesi, sadece çeviri işlemi (çeviri-geri çeviri/translation-back translation) yapılmasının sakıncalarının gösterilmesi bu konuların ayrı ayrı ciddi bir basamak olduğunu göstermektedir.

Sonuç ve sınırlar

Testin mükemmel psikometrik özellikler göstermesi, yapılacak bir araştırmada uygun bir materyal bile olsa, o ölçeğin o örneklemde iyi işlemeyebilir. Bazen kullandığımız yöntemin -Wisconsin Kart Eşleme Testinde güvenilirliğe uygun yöntemi oluşturamamak gibi- bazı özelliklerini gösteremeyebiliriz. Psikometrik özelliklerini incelerken, hangi örneklemle, nasıl bir yöntem izleneceği, örneğin, faktör yapısının nasıl olduğu -hangi yöntemle inceleneceği-, altın standart yokluğunda hangi değişkenlerle ilişkilendirileceği gibi özellikler bu çalışmanın içine direkt olarak dâhil edilmemiştir. Kullanışlık sorununu çözen ve özelliklerini iyi bildiğimiz bir ölçeğin örneğin, bedensel bir hastalıkta

depresyon şiddetini, bu maddeleri (bedensel) fazla içeren bir ölçeği kullanması uygun olmayacaktır. Ya da cinselliği hesaba katmayan bir yaşam kalite ölçeğiyle bu yönden sorun yaşayan grubun değerlendirilmesi gibi. Hangi ölçeğin seçileceği, kişinin kendini değerlendirme veya uygulayıcı-yakın değerlendirme ölçekleri, tarama ölçekleri benzer önem gerektiren durumlardır ve bu çalışma içinde direkt olarak tartışılmamıştır.

Klinik psikiyatride araştırmalarda ölçeklerin tartışılmaz bir yeri vardır. Her ne kadar bilimsel duruş tartışılmazlığa şüphe ile baksa da, ölçeklerin öneminden çok kullanılacak türü, uygulama örnekleme, yeri ve zamanı konusunu daha önceden tartışmaktadır. Bu yaklaşım çalışmaların bilimsel düzeyine tekrarlanırlığına ve bilimsel etkisine önemli katkılar sağlayacaktır.

Kaynaklar

1. Karakaş S. BİLNOT bataryası el kitabı. Nöropsikolojik testler için araştırma ve geliştirme çalışmaları. Eryılmaz Matbaası, Ankara. 2004.
2. Blacker D, Endicott J. Psychometric properties: concept of reliability and validity. Hollander E, Cohen L, Simon L (Eds): Handbook of psychiatric measures. Rush AJ, Pincus HA, First MB (Eds), APA, Washington, DC: p. 7-14, 2005.
3. Buchanan RW, Carpenter WT. Concept of schizophrenia. Comprehensive Textbook of Psychiatry, 7. baskı, 1. cilt, BJ Sadock, VA Sadock (Eds), Philadelphia, Lippincott Williams&Wilkins, 2004.
4. Aydemir Ö. Psikiyatride değerlendirme araçları: özellikleri, türleri kullanımı. Psikiyatride kullanılan klinik ölçekler. Aydemir Ö ve Köroğlu E. (Ed.) Hekimler Yayın Birliği. Ankara. 2007.
5. Öncü H. Eğitimde ölçme ve değerlendirme. Matser Basımevi. Ankara 1994.
6. Thorndike RM, Cunningham GK, Thorndike RL, Hagen EP. Measurement and evaluation in psychology and education. 5th ed. Macmillian Publishing, 1991.
7. Tekin H. Eğitimde ölçme ve değerlendirme. Yargı Yayınları, Ankara. 2000.
8. Tavsancıl E. Tutumların ölçülmesi ve SPSS ile veri analizi. Nobel Yayınevi, Ankara. 2006.
9. Mehrens WA, Lehmann I.J. Using standardized tests in education, Longman, New York, USA. 1987.
10. Karasar N. Bilimsel Araştırma Yöntemleri. 10. baskı. Nobel Yayınevi, Ankara. 2000.
11. Demircioğlu G. Ölçme ve değerlendirme. Pegem A Yayıncılık, Ankara. 2007.
12. Büyüköztürk S. Sosyal bilimler için veri analizi el kitabı: istatistik araştırma deseni, SPSS uygulamaları ve yorum. Pegem A Yayıncılık, Ankara. 2008.
13. Kan A. Eğitimde ölçme ve değerlendirme. Ani Yayıncılık, Ankara. 2006.
14. Tan S. Ölçme ve değerlendirme, öğretimde KPSS el kitabı. Pegem A Yayıncılık, Ankara. 2008.