

SOSYAL BİLİMLER ARAŞTIRMALARINDA KULLANILAN ARAÇLARIN GEÇERLİK VE GÜVENİRLİKLERİ

Doç. Dr. Şefik UYSAL

Bir arařtırmanın, arařtırılan sorunu yansıtabilmesi, dikkatli bir arařtırma plânu hazırlanmasına baęlı olduęu kadar, arařtırmada kullanılacak veri toplama ve ölçü araçlarının etkililięini de gerektirmektedir. Bir arařtırma sorununun incelenmesinde gerekli verileri toplayabilmek için bir takım araçlara ihtiya olacaęı doęaldır. Ölme araçları ise, bireylerin ya da objelerin özelliklerini nicelik olarak saptamaya yardım ederler. Ölme, kısaca, belirlenen niceliklere belli kurallara göre sembolik deęerler verme sürecidir.

Bir arařtırmada söz konusu olan sorunlar için uygulanan anlamlı ve uygun bir ölçme sürecinde, ölçülmesi istenen teorik kavram ya da kavramlara bir açıklık kazandırmak gerekir. Dięer bir deyişle ne ölçüleceęini bilmek önemlidir. Bir ok kavramların eşitli teorik yaklaşımlarla açıklandığı Sosyal Bilimlerin bazı alanları için, kavramlara açıklık kazandırılması sorunu önemli bir uğraşdır. Ancak kavram açıkca belirlendikten sonra, bunun nasıl ölçüleceęini düşünmek ve saptamak mümkün olur.

Bilindięi gibi Sosyal Bilim arařtırmalarında veriler, davranışların gözlenmesi, anketler, mülâkatlar, testler veya mevcut dokümanlardan yararlanmak gibi eşitli yöntemlerle elde edilmektedir. Bir arařtırmada hangi veri toplama yönteminin yararlı olacaęı sorusuna en olumlu cevabı, ölçülmek istenen kavram ya da sorunda aramak en doęru yol olarak kabul edilmektedir.

Ölme, oęu kere, oldukça karmaşık durumlarda yapılır ve ok eşitli faktörler ölçülen özellięe etki eder. Bu yüzden, bir ölçme aracı kullanıldıktan sonra elde edilen gözlem deęerleri, bireyler ya da objeler arasındaki gerek farklılıkları yansıttığı gibi dięer başka faktörlerin de bir sonucu olabilir. Başka bir deyişle, ölçme sonucu elde edi-

len farklılıkların dağılımını doğuran nedenler oldukça çeşitlidir. Aslında bir dağılım ölçüsü olarak varyans'ı etkileyen bu çok çeşitli nedenler, maksada göre değişik biçimlerde sınıflandırmaktadır. Konumuzla ilgili olduğu için biz burada, Sosyal Bilimlerde genellikle yapılan ölçmeleri dikkate alarak, varyansı doğuran nedenleri sekiz kategoride inceleyeceğiz.

1) Gerçek Farklılıklar

Gerçek farklılıklar ölçülen özelliğe göre bireyler ya da objeler arası gerçek farklılıkların varyans'ta yansımalarıdır. Bir örnek verelim; 1962 yılında Türkiye'de yapılan ve "Köy Araştırması" adı ile bilinen araştırma için hazırlanan ankette, köylülerin dinsel uygulamalara karşı tavırlarını saptamaya çalışan sorunlar yer almıştır. Burada, dinsel uygulamalara karşı tavırlar yönünde verilen tepkilerin, bir ölçüde bireylerin gerçek tavırlarını yansıtır nitelikte olması beklenmektedir. İşte bu beklenti doğrulandığı oranda varyansın bir kaynağı olmaktadır. Aslında bütün Sosyal Bilim araştırmalarında araştırmacıların gayreti, varyansın çok büyük ölçüde bu gerçek farklılıkların bir sonucu olmasını sağlamak yönünde yoğunlaşmaktadır.

Diğer yandan, aşağıda görülecek öbür varyans kaynaklarından da anlaşılacağı gibi, özellikle Sosyal Bilimlerde, varyansın tamamen gerçek farklılıkları yansıttığını iddia etmek mümkün olmamaktadır.

2) Bireye Ait Oldukça Sabit Özellikler

Bireyin bir uyarıya gösterdiği tepki genellikle bireyin sabit özelliklerinin etkisi altında olmaktadır. Tepkiyi bu sabit özelliklerin etkisinden arıtmak çoğu kere mümkün değildir. Yukarıda sözü edilen "Köy Araştırması"nda bu sorunu inceleyelim; Dinsel uygulamalara gösterilen tepki, bireyin gerçek tavırlarını yansıtabileceği gibi, içinde bulunduğu toplumun bu yöndeki değer yargılarını nasıl algıladığını da içerecektir. Diğer bir deyişle, toplumca benimsenen bir değer yargısının dışında tavır belirtmek birey için oldukça güç olacaktır.

Psikometride hazırlanan bir çok "tipik tepki" ölçeklerinde bu duruma daha sık raslanmaktadır. Örneğin, bir şahsiyet ölçeğinde, toplumca olumsuz olarak nitelendirilen bazı durumlarda birey gerçeği yansıtmaktan ziyade toplumca beklenen tepkileri göstermektedir. Bireyin "sık sık yalan söyler misiniz?" uyarımına "evet" tepkisini vermesi, toplum beklentileri dışında bir tavır gösterebilme davranışını kazanmış olmasına çok bağlıdır.

Varyansa etki eden, bireye ait oldukça sabit özellikleri sıralarken şahsiyet dışında, bireyin zekâ seviyesi, cinsiyeti, yaşı, içinde bulunduğu sosyo-ekonomik kesim ve ilgileri söz konusu edilebilir.

3) Kişisel Özellikler

Araştırmada kendilerinden veri toplanacak bireylerin tepki gösterdikleri zaman süresi içindeki sağlık durumları, zihinsel hazır oluşları ve yorgunluk ortaya çıkan farklılıkların-varyansın- bir diğer kaynağı olarak belirmektedir. Bir araştırma için köye giden araştırma ekibinin, örnekleme giren köylülerle, onları köyde belli bir yere çağırarak sureti ile mülâkat yapması bu konuda bir örnek olarak gösterilebilir. Hasat zamanı çalıştığı tarlasından köye çağrılan bir köylünün mülâkatın bir an önce bitmesini beklemesi doğaldır. Ayrıca yorgunluk, ve mülâkata hazır olmama gibi durumlar da söz konusudur. Mamafih, araştırmacıların alacakları tedbirler, bu tür nedenlerin varyansa olumsuz etkisini asgariye indirmektedir.

4) Durumsal Özellikler

Ölçmenin yapıldığı durumlar veya durumsal koşullar bireylerin tepkilerini büyük ölçüde etkilemekte ve bu varyansa bir hata kaynağı olarak girmektedir. Örneğin, muhtar ve benzeri resmî kişiler önünde yapılan mülâkatta, araştırmaya konu olan bir köylünün gerçek tepkilerini göstermesi ne dereceye kadar mümkündür? Bu durumda bireyler, bir büyük ihtimalle, bu resmî kişilerce de tesvip görecektir tepkileri gösterme eğilimi içine gireceklerdir.

Diğer taraftan, bir araştırma için yapılan mülâkatların bir kısmının bireyle yalnız olduğu, diğer bir kısmının yanında başkaları olduğu zaman yapılmasının da, mülâkatların farklı durumlarda yapılmasından ötürü sonuçlarda bir farklılık meydana getireceği düşünülmalıdır. O halde, ölçmenin yapıldığı durumlarda mümkün olduğu kadar, bireylerin tepkilerini etkileyecek koşulların yaratılması önlenmelidir.

5) Uygulama Farklılıkları

Ölçü aracının uygulanmasında hemen bütün ayrıntılarına kadar düzenlenmiş bir uygulama yöntemi bulunmalıdır. Aksi halde uygulamanın ortaya koyduğu değişik uyurım düzeni, sonuçta istenmeyen bir farklılığın doğmasına sebep olacaktır. Örneğin, bir mülâkatçının aynı soruyu bir bireyden diğerine değişen ifadelerle yöneltmesi sonucu ortaya çıkan tepki farklılıkları, değişik ifadelerle verilen uya-

rımların bir sonucu olabilir. Ayrıca mülâkatçıların yaptıkları mülâkatlarda değişik psikolojik durumlarda bulunmaları da farklılıklara olumsuz etkiler yapmaktadır. Örneğin, bir mülâkatta neşeli, birinde yorgun, bir diğesinde ise çok ciddî olan bir mülâkatçının bu değişik durumları, sonuçta elde edilen varyansta önemli bir kaynak olarak görülebilir.

6) Soruların Örnekleme

Ölçülmek istenen özelliğe ilişkin bir çok tepki söz konusudur. Bu tepkilerin saptanabilen tümüne “davranış popülasyonu” denmektedir. Sosyal Bilim alanlarında yapılan araştırmalarda herhangi bir özellikle ilgili bütün tepkilere ait veri toplaması düşünülemez. Böyle bir yaklaşım birçok bakımdan sakıncalıdır. Bu yüzden hazırlanan ölçek veya diğer veri toplama araçlarında, ölçülmek istenen özelliğe ait en belirgin ve en temsili olan sorular ya da ifadeler yer alır. İşte bunlara “davranış örnekleme” veya “sorular örnekleme” denmektedir. Bu kısa açıklamadan belirgin olarak görüleceği gibi, ölçülen bir özelliğe ait tek bir soru veya ifade koymak yerine örneğin bir çok sorular veya ifadeler koymak, örnekleme hatasını azaltacak, diğer bir deyişle, davranış popülasyonu yeterince temsil edilmiş olacaktır. Böylece bireyler arası farklılıklar, örneklemin darlığının ortaya koyacağı yetersizlikten büyük ölçüde etkilenmeyecektir. Örneğin, daha önce sözü edilen “Köy Araştırması”nda dinsel uygulamalara karşı tavırlar ölçülürken, bir yandan bu konuda önemli olan ibadet biçimleri-namaz, oruç v.b. -dikkate alınırken diğer yandan da batıl inançlarla ilgili uygulamalar -adak, türbe ziyareti v.b. ile ilgili davranışlarla tavırların ölçülmeye çalışılması, davranış örneklemini genişletmek sureti ile varyansta gerçek farklılıkların yansımaya imkân hazırlanmış olmaktadır.

7) Ölçme Aracının Anlaşılabilirliği

Sosyal Bilim araştırmalarında kullanılan ölçme araçlarında yer alan ifade veya soruların cevaplandırana göre değişik tanım kazanmasından ötürü ortaya çıkan durumun varyansa etki etmesi, çoğu kere, ölçü aracının gereğince açık ve anlaşılır biçimde ifadelendirilmemiş olmasındandır. Bu durum bazen çok kolay sanılan sorularda bile görülebilir. Örneğin, bir bağımsız değişken olarak, öğrenim görülen yeri coğrafî bölgelere göre dikkate almayı düşündüğünüz bir araçta, “Son mezun olduğunuz okul?” gibi bir soru, çoğu kere istenen bilgiyi getirecek açıklıkta olmayabilir. Çünkü, cevaplar, daha

ziyade lise, öğretmen okulu veya üniversite gibi öğrenim seviyesine ilişkin bilgileri kapsayacaktır. Bu bilgiler ise, sözü edilen değişken ile ilgili duruma bir açıklık getirmeyecektir.

Ayrıca, “sık sık”, “çoğu kere,” “nadiren” gibi davranışların frekansını ifade eden terimler de bireylere göre değişen tanımları içermektedir. Yapılan bir araştırmaya göre, örneğin “sık sık” deyimini bireylere göre yüzde 70’ten yüzde 95’e kadar değişen bir sıklığı belirtmektedir. “Çoğu kere” deyiminde ise frekans, yüzde olarak (60-85) arasında değişmektedir. Görülüyor ki bir birey tarafından “sık sık” belirtilen bir davranış bir diğer bireyde benzer frekansta olmasına rağmen “çoğu kere” ve hatta “bazen” olarak nitelendirilebilmektedir. O halde ölçek veya anketlerdeki soru ya da ifadelerin açık, kesin ve mümkün olan ayrıntıları ile belirtilmesi, varyansa bu açıdan olumsuz etkiyi azaltmış olacaktır.

8) Diğer Faktörler

Ölçü araçlarının uygulanması ve sonuçların analizinde ortaya çıkması muhtemel küçük mekanik hataların varyansta yansıtacağı düşünülmelidir.

a. Çoğu kere, ölçü araçlarının cevaplandırılmasında, dikkatsiz işaret konulması, yanlış kutuların işaretlenmesi veya açıklamaya uygun işaretlenmemesi gibi mekanik hatalar sonuçları değiştirebilecek nitelikte olmaktadır. Nadiren görülen bir örnek verelim, Bir araştırmada kaleminin mürekkebi biten bir mülakatçının ruj kullanmak sureti ile koyduğu (X) işaretlerinin hangi cevaba ait olduğunu saptamak bir hayli güç olmuş ve bazı ahvalde anketlere verilen cevapları iptal etmek gereği duyulmuştur.

b. Analizde görülebilen ve sonuca etkisi muhtemel bazı hata kaynakları şunlardır.

- i. mülakatçının ankete sorular ekleyerek aldığı cevapları kaydetmesi,
- ii. Soru ve ifadeleri aslındaki sıraya göre vermemesi,
- iii. Bazı soru veya ifadelerin atlanması,
- iv. Soru ve ifadelerin aslından farklı verilmesi,
- v. Cevapların verildiği şekilden farklı olarak yorumlanması veya değiştirilerek kaydedilmesi.

Bu ve benzeri durumlarda ölçme sonucu elde edilen verilerin ortaya koyduğu durumun gerçeği yansıtması ihtimalinin az olması beklenmelidir.

Buraya kadar sözü edilen hususlar hemen her ölçme sürecinde söz konusu olmaktadır. Yalnız farklı disiplinler ve bilim alanlarındaki farklı ölçümler için değişik varyans kaynakları sınıflaması düşünülebilir. Nitekim Cronbach, özellikle Psikometri için varyans kaynaklarını sürekli ve geçici olmak üzere iki genel kategoride toplamaktadır. Her bir kategoride de genel ve spesifik olmak üzere iki değişik varyans kaynağı incelenmektedir. (Cronbach, 1960).

Yukarıdaki tartışmalardan anlaşılacağı gibi, ilk maddede açıklanan; bireyler arası gerçek farklılıklara göre varyansın bulunması ölçmenin temel amacı olmaktadır. Bunun dışındaki varyans kaynakları değişen türde ve ölçülerde hata varyansını oluşturmaktadır. Diğer bir deyişle, ölçmeye hatayı getirmektedirler. Geçerlik ve güvenilirlik kavramlarının tartışmasına bir hazırlık olmak üzere ölçmede, sözü edilen, bu hataları iki genel kategoride toplamak uygun olacaktır.

1) Sistematik Hata (Sistematic Error)

Her tür ölçmeye sistematik bir etkide bulunan türde bir hatadır. Yukarda sözü edilen varyans kaynaklarından ikinci maddedeki bireyin sabit özellikleri, sistematik hatanın en tipik bir kaynağıdır.

2) Şans Hatası (Random Error)

Ölçmede bireyin, ölçme çerçevesinde, durumdan duruma ve zamandan zamana değişen özelliklerinin ve tepkilerinin ortaya çıkardığı türdeki hatalardır. Bu yüzden devamlılık gösteren bir yönü yoktur. Bir ölçmeden diğerine görülen farklılıklar şans hatasını yansıtmaktadır. Varyans kaynaklarından yukarda sıralanan son altı tanesini içeren bu tür hataların yarattığı sorun daha ziyade güvenilirlik kavramı ile yakından ilişkilidir. Geçerlik ise hem sistematik hem de şans hatalarından etkilenmektedir.

Şimdi geçerlik ve güvenilirlik kavramlarını bazı önemli yönleri ile tartışalım.

A. GEÇERLİK (VALIDİTY)

Daha önce de kısaca değinildiği gibi, Sosyal Bilimlerde ölçülen kavramlar çoğu kere soyut ve karmaşık olmaktadır. Bu temel nedenden ötürü, ölçmeler daha ziyade belirtilere dayanılarak dolaylı yapılmaktadır. O halde, Sosyal Bilimlerde kullanılan ölçü araçları için cevaplandırılması gereken bazı temel sorular bulunmaktadır. Örneğin,

ölçü aracı ne ölçmüştür? Ölçülen özellik yönünde elde edilen sonuçlar ve sayısal değerler, ne dereceye kadar bireyler arası gerçek farklılıkları yansıtmaktadır? Veya, elde edilen veriler ölçülmek istenen özellikleri yansıtacak nitelikte midir? İşte bu ve benzeri soruların cevapları büyük ölçüde, ölçü araçları ile ilgili olarak yapılan geçerlik çalışmaları ile elde edilmektedir.

Geçerliğin, bir ölçü aracının ölçmek istenen özelliği ölçebilmesi niteliği, şeklinde benimsenen bir genel tanımı vardır. Buraya kadar yapılan açıklamaların ışığında bir ölçme aracının geçerliği, ölçülmek istenen özelliğe göre, bireyler, gruplar ya da durumlar arasındaki gerçek farklılığı yansıtan bir niteliği olarak da tanımlanabilir. Geçerlik saptanmasındaki yöntemlere göre, 1). Mantıkî geçerlik, ve 2). Deneysel geçerlik olmak üzere iki genel kategoride düşünülmektedir. Bir de bunların dışında, son yıllarda önem kazanan Yapı Geçerliği söz konusu edilmektedir.

1. Mantıkî Geçerlikler (Logical Validity)

Bir ölçü aracının ne ölçtüğünü görmek için tutulan yollardan birisi, ölçülmek istenen özelliğe göre ölçü aracının, sistematik ve anlamlı bir şekilde dikkatle incelenmesidir. Daha ziyade uzmanlarca yapılan böyle bir inceleme sonucu, ölçü aracının neyi ölçtüğü yönündeki görüş, mantıkî geçerlik açısından bir bilgi vermektedir. Görüldüğü gibi mantıkî geçerlik'te, geçerlik seviyesini bir sayısal değerle belirtmek imkânı yoktur. Mantıkî geçerlikler, aslında, diğer deneysel geçerliklere güvenle başvurmak imkânını hazırlamış olmaktadır. Burada mantıkî geçerliklerden en çok kullanılan iki tanesi kısaca tanıtılacaktır.

a. Yüzeysel Geçerlik (Face Validity)

Yüzeysel geçerlik geliştirilen bütün ölçü araçlarında ilk başvuru olan bir geçerlik türüdür. Bir veya bir kaç uzmanın görüşüne başvuru olarak, ölçme aracının kullanılacağı araştırma sorunları için gerekli nitelikte veriyi toplayacak durumda olup olmadığı yönünde alınan bilgi yüzeysel geçerlik için çoğu kere yeter kabul edilmektedir. Oduka basit bir yöntemle toplanan bu bilgi, anlaşılacağı gibi, ölçü aracının neyi ölçtüğünü değil de, neyi ölçer göründüğünü belirtmektedir. Yüzeysel geçerliği bilimsel açıdan savunmak güçtür. Ne kadar dikkatli incelenirse incelenirse, deneysel yolla toplanan bilgilere sahip olunmadıkça ortaya konan geçerlik inandırıcı bir düzeyde olmamaktadır. Ama, geliştirilmekte olan bir ölçü aracının ne dereceye kadar

geçerli görüldüğünün araştırma amaçları ile tutarlılığının böyle bir yüzeysel geçerlik incelenmesi ile önceden bilinmesi araştırmacılar için oldukça önemlidir. Ancak, bu yolla araç daha ilk adımda geliştirilmeye açık bir nitelik kazanmış olmaktadır.

Anlaşılabacağı gibî yüzeysel geçerlik, ilerde elde edilecek deneysel geçerliğin yerini alabilecek bir tür değıldir.

Yüzeysel geçerlik ölçü aracının geliştirilmesinde, örneğın, bir anket için hazırlanan soruların amaçlarla tutarlılığını önceden görmek yönünde bir yarar sağladığı gibi, ölçü aracının uygulanmasında istenilen ortamın ve bireysel ilişkiler düzeninin yaratılmasında da yardımcı olmaktadır. ölçü aracının araştırma amaçları ile tutarlılığı, elbette ki tabii olarak, aracı alanlar üzerinde olumlu bir etki yaratacaktır. Bu etkinin uygulama süresince arzulanan atmosferin sağlanmasında ve toplanan bilgilerin geçerliğinin artırılmasında önemli bir rolü vardır.

b. Muhteva Geçerliğı (Content Validity)

Bu tür geçerlikte aranan, ölçü aracının, tanımlanan davranış popülasyonunu yeterince temsil edebilmesidir. Bu nedenle, muhteva geçerliğı, örneklem geçerliğı (sampling validity) olarak ta adlandırılmaktadır.

Muhteva geçerliğinin saptanması görüldüğü kadar kolay olmamaktadır. Önce, ölçülmesi istenen davranışlar popülasyonunun dikkatli ve sistematik bir analizinin yapılması ve içerdiği faktörlerin oldukça açık olarak saptanması gerekir. Sonra da, her faktörün davranışlar popülasyonundaki oranlarına göre ölçü aracının bir örneklem alıp almadığına bakılmaktadır. Örneğın, dinsel uygulamalara karşı tavırları ölçen bir araçta, bu uygulamaların herbirinin dinsel tavırların tanımına olan katkı derecesine göre, azalan veya çoğalan sayıda davranışları ölçecek nitelikte soru ya da ifadelerin bulunması gerekmektedir. Diğer yandan, verilerin muhteva analizinde, ölçü aracının örnekleme alınan davranışların ölçülmesinde aşırı genellemelere zemin hazırlayacak nitelikte olmamasına da dikkat edilmelidir.

Muhteva geçerliğinin saptanmasında çeşitli yöntemler vardır. Bunlardan, Sosyal Bilim araştırmalarında en yaygın görüneni, daha önce yapılan benzeri araştırma bulgularından, özellikle değışkenler arası ilişkileri gösteren istatistik verilerden yararlanarak ölçülmek istenen popülasyona ait davranış örnekleminin muhteveya uygunluğunu saptamaktır. Bu daha ziyade, ölçü aracının, alan uzmanlarınca yapı-

lan dikkatli ve sistematik bir değerlendirilmesi sonucu olmaktadır. Bu yüzden araştırmacılar genellikle, geliştirdikleri ölçü araçlarını ilgili alan uzmanlarının incelemelerine açık tutma ve muhteva geçerliği açısından yapılacak eleştirileri dikkate alma eğilimindedirler.

Sosyal Bilim araştırmalarında kullanılan kişilik envanterleri ve yetenek ölçekleri gibi psikometrik araçlarda, muhteva geçerliği aynı derecede önemli olmayabilir. Çünkü, bir çok psikometrik araçlarda ölçek, davranış alanına çok az benzerlik gösterebilir. Bu yüzden psikolojide kullanılan ölçme araçlarında deneysel geçerlikler daha büyük önem kazanır.

2). Deneysel Geçerlikler (Empirical Validities)

Geçerlik saptanmasında deneysel yaklaşım, ölçme aracının ölçmek istediği davranışları yansıtıcı veya ileriye ait tahmin (prediction) imkânını verecek nitelikte olup olmadığı yönünde bilgi verir. Örneğin, bir zekâ ölçeği ya geri zekâlılığı teşhis veya öğrencilerin bir üst öğrenim seviyesindeki akademik başarısını önceden tahmin edebilme amacı ile kullanılabilir. Böyle bir durumda bir zekâ ölçeğinin ya öğrencilerin davranışlarını yansıtması, ya da ikinci maksat için, ilerdeki akademik başarısını önceden tahmin edebilme gücüne sahip olması gereklidir. Bu yüzden ölçü aracının bir dış kriter ile ilişkisine bakılmaktadır. Deneysel geçerliklerin saptanmasında, görüldüğü gibi, iki devamlı değişken arasında ilişki arandığı için, en çok korelasyon teknikleri kullanılır. Bunların en yaygın olanı Pearson Momentler Çarpımı Korelasyonu tekniğidir.

Deneysel geçerliğin saptanmasında kriter çok önemlidir. Kriter ölçülmek istenen kavramı ne dereceye kadar yansıtırsa, bulunan ilişki, o derecede ölçü aracının geçerliği yönünde sağlam bilgi verir.

Kriterlerin saptanması ölçü aracının mahiyetine ve araştırmanın amaçlarına göre farklılık göstermektedir. Örneğin, üniversite giriş sınavlarında kullanılan bir zekâ ölçeği için, öğrencilerin aldıkları ders notları ortalaması ya da bu maksatla verilen başarı testleri puanları birer kriter olarak kullanılabilir. Bir ilgi envanteri için, işbaşında gösterilen başarı değerlendirilmesi sonuçları bir diğer kriter olarak düşünülebilir.

Bazen bir kriter ölçümü elde etmek güç olabilir veya sağlam sonuçlar elde edilmeyebilir. Bu takdirde önceden tanımlanmış gruplardan, aynı ölçü aracı ile, elde edilmiş puan ortalamaları deneysel geçerliğin saptanmasında kullanılabilir. İlgili ölçülmesinde kullanılan

bir yöntem buna en tipik bir örnek olabilir. Ölçme aracı bir genel gruba, bir de ilgi alanına giren bir meslekteki başarılı kimselere uygulanır. Eğer ölçme aracı, deneysel olarak geçerli ise, o meslekteki başarılı kimselerin puan ortalamasının, genel grubun elde ettiği puan ortalamasından daha yüksek olması gerekir.

Kriter saptanmasında kullanılan diğer bir yöntem de gruplamadır ve en çok ilgi ölçülmesinde kullanılır. Örneğin, eğer bir ilgi envanteri, çeşitli mesleklerdeki kimseleri oldukça yüksek oranlarda birbirlerinden ayırt edici nitelikte sonuçlar vermişse, o takdirde bu envanterle ilginin ölçülmesinde geçerli sonuçlar elde edildiği kabul edilmektedir.

Deneysel geçerliklerin tartışmasında kriter saptanması önemli bir yer tutar. Bu yüzden kriterleri daha sistematik bir şekilde vermek yararlı olacaktır. Çeşitli kriter sınıflamaları yapılmaktadır. Biz bunlardan en yaygın olanını kısaca özetlemeye çalışalım.

i. Yaş:

Özellikle zekâ ölçeklerinde geçerliği saptamak için kullanılan bir kriterdir. Diğer bir deyişle, yaş ilerledikçe gösterilen doğru tepkilerde azalma görülürse ölçü aracının geçerliğinin düşmekte olduğu kabul edilmektedir.

ii. Akademik Başarı:

Yine özellikle zekâ ölçeklerinde kullanılan bir kriter ölçüsüdür. Ölçekten yüksek puan alanların üst başarı seviyesinde görülmesi geçerliğin bir belirtisi olarak kabul edilebilir.

iii. Mesleğe Yönelik Eğitimde Başarı:

Özel yetenek ölçekleri ve ilgi envanterleri için genellikle uygun görülen bir kriterdir. Örneğin, mekanik yetenek ölçeği verildikten sonra, ölçeği alanların teknik üniversitedeki öğrenimlerinin başarı ölçüsü bir kriter olarak alınabilir.

iv. İşbaşında Başarı:

Yine, özellikle ilgi ve özel yetenek ölçekleri için düşünülebilecek kriterlerdir. İş başında başarı kriteri, sınırlı bir davranış örnekleri için alınabileceği gibi, bireylerin işbaşındaki tüm başarılarını belirten davranışları da içerebilir. Örneğin, bireyin sadece çalışma arkadaşları ile olan ilişkisini gösteren davranışlar veya görevin gerektirdiği bütün davranışlar açısından kriter ölçüleri alınabilir.

v. Bireylerin Belirli Davranışlara Göre Sıralanması:

Bu tür, Psikometride kullanılan hemen hemen bütün ölçme araçları için düşünülebilecek bir kriter ölçüsüdür. Bu tür bir kriter ölçüsünde, bireylerin, önceden tanımlanmış davranışlarına göre, öğretmenleri, üstleri, komutanları veya kendilerini gözleyen kişiler tarafından belli esaslara göre bir sıralaması yapılır.

vi. Karşıt Gruplar:

Deneyssel geçerliğin saptanmasında karşıt grupların kriter olarak alınması ile, günlük hayatın kontrol edilen ve seçilmiş etkilerinin yığılması yansıtılmaktadır. Politik davranışları ölçülen politikacıların yanında, politikaya karşı ilgisiz kimselerin aynı konudaki davranışlarının ölçülerek bir karşılaştırma yapılması bu tür kritere tipik bir örnek olabilir. Kişilik envanterlerinin geçerlik saptanmasında oldukça yaygın olarak kullanılan bir kriter ölçüsüdür.

vii. Diğer Benzer Ölçü Araçları:

Oldukça benimsenmiş bir yöntemdir. Önceden geliştirilmiş ve çeşitli denemelerde geçerlik ve güvenirliliği oldukça yüksek bulunmuş bir ölçü aracı, sonradan benzer maksatlar için hazırlanmış ve aynı değişkeni ölçmeye yönelik bir ölçü aracına kriter olarak kullanılabilir. Yahut da gruba uygulanması plânlanan bir anketin, daha önce bireylere tek tek uygulanmak üzere geliştirilmiş bir mülakat anketine göre geçerlik çalışmasını yapmak da mümkündür.

viii. Ölçü Aracının İçtutarlılığı:

Kriter olarak bir de ölçü aracının içtutarlılığı söz konusudur. Bu yöntemin temel özelliği ise ölçü aracına gösterilen tepkilerin toplamının kriter olarak alınmış olmasıdır. Anlaşılacağı gibi, toplam puan, geçerlikten ziyade ölçümün homojen olup olmadığı yönünde bir kriter olarak düşünülmektedir. Bu yüzden daha ziyade bir homojenlik ölçüsü olan iç tutarlılık, tam anlamı ile geçerlik olarak benimsenemez. Kaldı ki, bir ölçü aracının fazla mütecanis olduğu durumlarda, bir dış kriter söz konusu ise, geçerliğin düşmesi dahi beklenebilir.

Ölçü aracı ve kriter arasındaki ilişki sonucu olarak, belirtilen korelasyon katsayılarının ne kadar yüksek olması gerektiği yönünde bir karar vermek, ölçü aracının mahiyetine ve araştırma amaçlarına bağlı olmakla birlikte, elde edilen korelasyon katsayısının içerdiği

hata miktarı ile de ilgilidir. Nitekim korelasyonun içerdiği hata miktarlarını tahmin etme yönünde çeşitli istatistik teknikler de geliştirilmiştir.

Şimdi kısaca, Sosyal Bilim araştırmalarında kullanılan ölçü araçlarında daha çok aranan şu iki tür deneysel geçerliği tanıyalım. a. ileriye ait tahmin geçerliği (Predictive validity), ve b. Halihazır geçerlik (Concurrent validity).

a. İleriye Ait Tahmin Geçerliği:

Sosyal Bilimlerin bir çok araştırmalarında kullanılan ölçü araçlarının temel fonksiyonlarından birisi, bireylerin ya da grupların ileride gösterecekleri davranışlarını önceden tahmin edebilmektir. Diğer bir deyişle, ölçü araçları çoğu kere bu amaçla geliştirilmektedir. Örneğin, kullanılan bir zekâ ölçeği öğrencilerin ilerdeki akademik başarılarını, bir ayrıcalı yetenek ölçeği ilerdeki meslekî başarıyı veya bir politik tavır ölçeği belli niteliklere sahip kişilerin hangi partilere oy verebileceklerini önceden tahmine olanak sağlayabilmelidir. İşte bu ve benzeri sorunlarda, araştırmada kullanılan ölçü araçlarının böyle bir tahmin gücünü vermesi beklenmektedir. Bu yüzden ileriye ait tahmin geçerliği bir çok ölçü araçlarında aranmaktadır. Özellikle, psikolojide kullanılan ölçü araçlarından, bu tür bir geçerlik, kaçınılmaz bir teknik bilgi olarak istenmektedir.

İleriye ait tahmin geçerliğinin saptanmasında şöyle bir yöntem uygulanır. Ölçü aracı belli bir örneklem grubuna uygulandıktan sonra, ölçülen davranışın belirgin olarak görülebileceği belli bir süre beklenir. Bu süre sonunda kriter ölçüsü elde edilir ve ölçekten elde edilen puanlar ile kriter puanlar arasındaki ilişki, korelasyon teknikleri ile aranır. Eğer, ölçü aracında bireylerin elde ettikleri sonuçlara göre dağılımdaki yerleri, kriter ölçüsüne göre açık farklılıklar göstermemişse, ölçü aracının ileriye ait tahmin gücü (predictive power) olduğu kabul edilir. Diğer bir deyişle, bu ölçü aracının ileriye ait tahmin geçerliği yüksektir. Bu konuda, yıllardır uygulanmakta olan Üniversite Giriş sınavı en tipik bir örnek olabilir. Üniversite giriş sınavında kullanılan başarı ölçeği, üniversite öğretiminde başarılı olması muhtemel öğrencileri seçmek için kullanılmaktadır. Diğer bir deyişle, amaç ölçeği alan lise mezunlarından üniversite öğretiminde başarılı olma ihtimali yüksek olanları seçmektir. Böylece, bu ölçeğin, lise mezunlarının üniversite öğretimindeki başarılarını önceden tahmin etme imkânı vermesi beklenmelidir, yani ileriye ait tahmin

geçerliği yüksek olmalıdır. Böyle bir ölçeğin, ileriye ait tahmin geçerliğini saptamak için, hazırlanan başarı ölçeği bir öneklem grubuna uygulandıktan sonra en az bir öğretim yılı beklenir. Öğretim yılı sonunda, öğrencilerin üniversitede aldıkları not ortalamaları elde edilir. Bu durumda not ortalamaları kriter ölçüsü olmaktadır. Korelasyon yolu ile aranan ilişki yüksek çıkmışsa, bu başarı testinin belli bir hatanın sınırları içinde geleceğe ait tahmin geçerliğine sahip olduğu kabul edilir.

b. Halihazır Geçerlik (Concurrent Validity)

Her zaman, önce ölçü aracının ve bir süre sonra kriter ölçüsünün uygulanması mümkün olmayabilir. Ayrıca bazı durumlarda ileriye ait tahmin geçerliğinin saptanması, zaman alıcı, pahalı ve hatta belirli bir maksat için uygun görünmeyebilir. O zaman, halihazır geçerlikle yetinilebilir ve belki de tercih edilebilir. Halihazır geçerlikte, ölçü aracı kullanıldığı zaman, kriter ölçüsü de aynı zamanda elde edilir. Bu tür geçerliğin saptanmasında, örneğin üniversite giriş sınavı için hazırlanan başarı testi, bir grup lise son sınıf öğrencisine uygulanır ve testi alan öğrencilerin lise dersleri not ortalamaları da bir kriter ölçü olarak saptanır. Böylece test ve kriter ölçüsü (not ortalaması) arasındaki korelasyon hesapları ve bulunan sonuç, test için bir geçerlik oranı olarak kabul edilir.

Sosyal Bilimlerin bir çok alanlarında kullanılan anketler için halihazır geçerlik çoğu kere uygulanabilir bir nitelik göstermektedir. Belli bir konuda hazırlanan anket sonuçları ile, aynı konuda başka bir araştırmacı tarafından, ya da bir kamu kuruluşunca toplanmış bilgiler arasında ilişki kurmak sureti ile anketin halihazır geçerliği hakkında bir bilgi sahibi olmak mümkündür.

3) Yapı Geçerliği (Construct Validity)

Sosyal Bilim araştırmalarında kullanılan ölçü araçlarından elde edilen sonuçlara teorik ve bilimsel açıdan, daha güvenle yorum kazandırabilmek için, ölçülen kavrama açıklık getirmek amacı ile yapılan çalışmaların son yıllardaki en önemli ürünü yapı geçerliği olmuştur. Bu konudaki tartışma, özellikle psikometride, giderek yoğunlaşmaktadır.

Bilindiği gibi, Sosyal Bilimlerdeki bir çok kavramlar, zekâ, şahsiyetteki esneklik, şahsiyet, kaygı (anxiety) ve önyargı (prejudice) gibi soyut kavramlardır. Bunların her biri yapısal olarak, çeşitli du-

rumlarda bireylerin gösterebilecekleri davranışlardan oluşan bir tepki kalıbı ya da bir gruplaşmış davranışsal özellik olarak nitelendirilmektedir. Fakat bütün bu ve benzeri kavramları gerekli açıklıkla ölçebildiğimizi savunmak için, bilimsel çevrenin kolaylıkla benimseyebileceği ve kriter olarak kullanılacak nitelikte bir ölçü getirmek çoğu kere mümkün olamamaktadır. Sosyal Bilimlerde söz konusu kavramların karmaşık oluşu ve teorilere göre farklı ele alınışı, ölçmede böyle bir güçlüğü ortaya çıkmasında temel bir neden olarak görülebilir. İşte bu yüzden ölçme sorunlarına yaklaşımda, ilk aşama olarak, ölçülen kavram hakkında yapılan çeşitli bilimsel çalışmalar sonucu geliştirilen önerilerden biri veya bir kaç, aracı geliştiren tarafından benimsenmektedir. Diğer bir deyişle, ölçülen kavramı tanımlayan teorilerden biri tercih edilmektedir. Böylece, ölçü aracının ölçmek istediği kavramın yapısal görünümü ortaya çıkmaktadır. Bu görünümü teorik açıdan savunulması mevcut araştırma bulgularına dayandırılacağı gibi, yeni bazı araştırmaların yapılmasını da gerektirebilir.

İkinci aşamada ise belirlenen yapıya göre bireylerin beklenen davranışları belli bir sistematik düzen içerisinde ortaya konmaktadır. Böylece, sistematik davranışlara göre, hazırlanan ölçü aracında yüksek puan alanların, düşük puan alanlardan farklılığını göstermek sureti ile geçerlik çalışması yapılmış olur.

Bir örnekte bunu görelim: Bir araştırmacı, zekâyı bir çok faktörler yanında şu beklentilere göre tanımlayabilir, yani yapıyı belirler;

- i. Zeki insanların kelime hazineleri geniştir.
- ii. Zeki insanlar, daha uzun ifadeleri, düzgünce ve güçlük çekmeden konuşur ve yazarlar.
- iii. Zeki insanlar propagandadan kolayca etkilenmezler.
- iv. Zeki insanlar çevrelerine uyumda güçlük çekmezler.

Böylece, bu yapıya göre hazırlanan bir ölçü aracında yüksek puan alanlar bu davranışları gösterirlerse bu "zekâyı ölçen aracın geçerliği yüksek olabilir" şeklinde düşünülmektedir.

Eğer bu teorik öneri, diğer bir deyişle, bu teoriye uygun yapıya göre hazırlanan bu ölçü aracı, beklenen sonucu vermezse, o zaman şu ihtimaller akla gelebilir:

- i. Hazırlanan ölçü aracı önerilen yapıyı ölçmekten uzaktır.

- ii. Ölçü aracının dayandırıldığı teorik yapı gerçeği yansıtmaktadır, diğer bir deyişle, gerçek değildir.
- iii. Benimsenen teorik yapıdan oluşan hipotezlerin doğru olup olmadığını saptamak için düşünülen ve uygulanan araştırma planı yeterli değildir.

Görüldüğü gibi, yapıya uygun olan sonuçlar alınmadığı zaman, nedenler, ölçü aracında, teorik yapıda veya araştırma planında aranmaktadır. Böylece yeni ve daha yaratıcı çalışmalar yapılması motive edilmektedir. Yapı geçerliği, bu yüzden, ölçme alanında yeni ve yaratıcı araştırmaları teşvik eden bir fonksiyona da sahip görünmektedir.

Yapı geçerliğinin bu genel yararı yanında böyle bir geçerlik çalışmasının çeşitli açılardan bilimsel yaklaşımı ve düşünmeyi de pekiştirdiği görülmektedir.

Görüldüğü üzere, yapı geçerliğini saptamak için tek bir yöntem olmadığı gibi, sonucu da tek bir sayısal değer ile ifade etmek de mümkün değildir. Çünkü, sorunun gereği olarak, çok yönlü bilgiler, çeşitli yöntemlerle sağlanmakta ve sonuç açıklayıcı bir ifade olarak verilmektedir.

B. GÜVENİRLİK (RELIABILITY)

Daha önceleri de söz konusu edildiği üzere, bir ölçme aracından elde edilen puan, aracın ölçmek istediği özelliği yansıttığı gibi, diğer bazı devamlı ve şans hatalarını da içermektedir. Bir araştırmacı, ölçme sonucu elde ettiği puvana, ne dereceye kadar güvenebileceğini, diğer bir deyişle, bu sonuçların ne kadar şans hatası dışında olduğunu bilmek durumundadır. İşte bir ölçmede, bireyden ve aynı ölçme aracının bir uygulamasından diğerine ne kadar şans hatası içerdiği hususu güvenilirlik sorunudur. Daha kısa bir tanımı ile güvenilirlik, bireylerin elde ettiği sonuçların ne kadarının ölçmedeki şans hatasına bağlanabileceğini göstermektedir. Yani güvenilirlik, bir ölçmenin devamlılık ölçüsü olmaktadır.

Teorik olarak güvenilirliği şöylece açıklamak mümkündür: Örneğin, bir zekâ testini aynı kişiye çeşitli defalar vermek mümkün olsa idi birey her seferinde birbirinden çok farklı olmasa da değişik puanlar alacaktır. Çünkü ölçmenin her uygulamasında bireyin kendisinden, ölçme arasından ve uygulamadan dolayı değişik hata kaynakları rol oynayacaktır. Böyle bir husus varit olmasa idi, o zaman bire-

yin aldığı puanın her uygulamada aynı olması gerekirdi. Bu ise imkansızdır. İşte bireyin aldığı bu değişik puanların dağılımının standart kayması ölçmenin standart hatasını verecektir. Bir ölçmede standart hata küçüldükçe, ölçmenin güvenilirliği artmış olacaktır. Bu yönde geliştirilmiş formülden yararlanarak bu ilişkiyi görelim:

Ölçmenin standart hatası

$$SE_m = \sigma \sqrt{1-r_{11}}$$

olarak tanımlanmaktadır.

Burada:

SE_m = ölçmenin standart hatası, ve

r_{11} = elde edilen güvenilirlik katsayısıdır.

Bir ölçü aracı uyguladığımızı ve elde ettiğimiz puanların dağılımının standart kaymasının (10) ve güvenilirlik katsayısının 0.70 olduğunu kabul edelim. Bu halde:

$$\begin{aligned} SE_m &= 10 \sqrt{1-0.70} \\ &= 10 \times 0.55 \\ &= 5.5 \text{ tir.} \end{aligned}$$

Yeni bireyin aldığı bir puan geniş bir sınır içinde değişebilecektir. Diyelim ki, bu ölçü arasından birey 100 puan almıştır. Ölçmenin standart hatası 5.5 olduğuna göre bu bireyin puanını büyük bir ihtimalle 94.5 (100-5.5) ile 105.5 (100+5.5.) arasında değişecektir.

Şimdi aynı ölçü arasının güvenilirlik katsayısının 0.90 olduğunu ve standart kaymayı aynı kabul edelim. Bu defa:

$$\begin{aligned} SE_m &= 10 \sqrt{1-.90} \\ &= 10 \times 0.31 \\ &= 3.1 \quad \text{olmaktadır.} \end{aligned}$$

Görüldüğü gibi ölçmenin standart hatası düşmüştür. Böylece daha güvenilir bir ölçü elde edilmiş olunmaktadır. Diğer bir deyişle, bireyin bir ölçmede aldığı puan daha dar bir sınır içinde, yalnız şans hatasına bağlı değişiklik gösterecektir.

O halde konuya yeni bir sorun daha girmektedir. Aynı aracın bir çok uygulamalarında değişen sonuçlar alındığına göre bireyin o değişiklerde gerçek puanı ne olmaktadır? Gerçek puan (true score) bireye bir defadan fazla verilen ölçü aracından aldığı puanların dağılımının

ortalaması olarak tanımlanmaktadır. En çok tekrar eden-normal dağılımda-puvar olarak ortalamanın gerçek puvarı yansıtması ihtimali daha yüksektir.

Alınan puvarların farklılığı bir hatanın söz konusu olduğunu açıkça göstermektedir. O halde bir çok defalar verilen aynı ölçü aracından bireyin elde ettiği sonuçların değişimi de bir hatayı yansıtmaktadır. Bu hata, daha önce de belirtildiği gibi ölçmenin standart hatası olarak görülmektedir. Bu yüzden elde edilen puvar dağılımının varyansı (s^2), gerçek puvarlar dağılımı varyansı (s_t^2) ile hata varyansı (s_e^2) toplamına eşit olmaktadır.

Böylece:

$$s^2 - s_e^2 = \text{gerçek varyans olmaktadır } (s_t^2).$$

Güvenirlik katsayısı ise:

$$r_n = \frac{s^2 - s_e^2}{s^2}$$

olarak tanımlandığına göre, bir önceki örneğimizi burada da uygulayalım:

$$\begin{aligned} s &= 10 & s^2 &= 100 \\ s_e &= 3.1 & s_e^2 &= 9.61 \end{aligned}$$

Buna göre:

$$\begin{aligned} r_n &= \frac{100 - 9.61}{100} \\ &= \frac{90.39}{100} \\ &= 0.90 \quad \text{çıkılmaktadır.} \end{aligned}$$

O halde, hipotetik dağılımda $s = 10$ ve $s_e = 3.1$ olduğu takdirde güvenilirlik katsayısı 0.90 olmaktadır. Böylece, bir ölçü aracının güvenilirlik katsayısı, elde edilen varyansın hangi oranda bireyler arası gerçek farklılığı yansıttığını göstermektedir.

Güvenirlik ile hata arasında bu ilişki dikkate alındığı zaman, hatayı azaltma yönünde alınan her tedbirin güvenilirlik katsayısını yükseltmesi beklenmelidir. Bu hata kaynaklarının başında örneklem hatasını ortaya çıkaran davranış örnekleme gelmektedir. Davranış örnekleme evreni daha iyi temsil edebildiği oranda, bunun güvenilirlik katsayısı

sına olumlu bir etkisi görülecektir. Kısaca, geliştirilen ölçü aracı, ölçülen özellikle ilgili daha çok madde (soru, ifade v.s.) içerdiği zaman güvenilirlik katsayısı artmaktadır. Fakat doğal olarak, ölçü aracındaki madde sayısını arttırmak sureti ile güvenilirlik katsayısının 1.00'e ulaşacağını iddia etmek mümkün değildir. Çünkü bir noktadan sonra, ölçü aracındaki madde sayısını arttırmak başka şans hatası kaynaklarını harekete geçirmiş olacaktır. Örneğin, bireylerin yorgunluğu, dikkatin dağılması, v.b. gibi. Spearman Brown formülü, ölçü aracındaki madde sayısı çoğaltıldığı zaman elde edilmesi muhtemel güvenilirlik katsayısı hakkında bir bilgi edinmeye yardım etmektedir.

Formül:

$$r_n = \frac{nr_n}{1 + (n-1)r_n} \quad \text{dır.}$$

Burada:

r_n = elde edilecek muhtemel güvenilirlik katsayısını

n = madde sayısının arttırılma katı, ve

r_n = halihazırda elde edilen güvenilirlik katsayısını göstermektedir.

Bir örnek verelim. Diyelim ki, 100 maddelik bir tavır ölçeği geliştirildi ve $r_n = 0.50$ elde edildiği. Araştırmacı bu ölçeğin madde sayısını 200'e çıkarırsa elde edilmesi muhtemel güvenilirlik katsayısı (r_n) ne olabilir. Yalnız burada r_n 'nin bir tahmin olacağını unutmamak gerekir.

Formüle uygulayalım:

$$n = 2 \text{ (2 katı arttırılmaktadır)}$$

$$r_n = 0.50 \text{ olduğuna göre;}$$

$$\begin{aligned} &= r_n \frac{2(0.50)}{1 + (2-1)0.50} \\ &= \frac{1 + 0.50}{1 + 0.50} \\ &= 0.67 \text{ olmaktadır.} \end{aligned}$$

Aynı işlemi, ölçeği 300 maddeye çıkarılacağına göre, uygularsak, $r_n = 0.75$ olmaktadır.

Güvenirlik katsayısı, ölçülen örneklemin mahiyetine göre de bir farklılık gösterecektir. Örneğin, politik tavırları ölçmek için hazırlanan bir ölçek, oldukça kaba bir ayırım yapmak maksadı ile politi-

kacılar dışında genel nüfustan alınan bir örnekleme uygulanmış ve güvenilirlik katsayısı 0.90 bulunmuş olsun. Burada dikkat edileceği gibi oldukça heterojen bir grupta farklılıklar aranmaktadır. Aynı ölçek, bir defa da bir politik partinin mensuplarına uygulandığı zaman güvenilirlik katsayısının farklı ve muhtemelen daha düşük olması deklennmelidir. Çünkü farklılıkların büyük olduğu bir durumda, aslında çok dikkatli hazırlanmış bir ölçeğe ihtiyaç duyulmaktadır.

Ama, aynı politik partide birbirine oldukça yakın tavırlara sahip bireylerin birbirinden olan farklılıkları ayırmak için çok dikkatli ve ayrıntılarına inilen bir ölçeğe ihtiyaç duyulacağı doğaldır. Tıpkı bakkal terazisinin kuyumcularda kullanılması imkanı olmadığı gibi. Kullanılırsa eğer, kuyumcuların hata yapmaları ihtimali artar ve ölçeğe güven azalır.

Güvenirliğin saptanmasında, maksada göre değişen çeşitli yöntemler uygulanmaktadır.

1) Devamlılık Katsayısı (Coefficient of Stability)

Bir çok durumlarda, bireylerin devamlı özellikleri ile ilgili ölçmeler yapılmaktadır. Zekâ, özel yetenekleri, şahsiyet gibi. Bu ve benzer durumlarda aranan güvenilirlik türü ölçünün devamlılığını yansıtmaktadır. İşte devamlılık katsayısı, ölçmenin böyle bir niteliğinin göstergesidir.

Bu tür güvenilirlik katsayısının saptanmasında, ölçü aracının belli bir zaman aralığı sonunda aynı örneklem grubuna uygulanması ve elde edilen iki seri puanın korelasyonunun bulunması ile olur. Yalnız araç iki uygulama arasındaki süre, yaş, eğitim, büyüme ve öğrenme gibi faktörlerin etkisini önemli miktarda azaltacak uzunlukta olmalıdır. Böylece bu faktörlerin etkisi ile puan farklılığının ortaya çıkması ihtimali azalmaktadır. İki uygulamadan elde edilen puanların aralarındaki tutarlılık, ölçü aracının ne dereceye kadar bireyin devamlı (kalıcı) özelliklerini ölçtüğünü göstermektedir.

2) Eşdeğerlilik Katsayısı (Coefficient of Equivalence)

Eşdeğerlilik katsayısı, farklı araştırmacıların aynı zamanda kullandıkları aynı ölçü aracının, yahutta farklı fakat benzer ölçü araçlarının aynı gruba aynı zamanda uygulanmasından elde edilen katsayıyı belirtmektedir. Özellikle iki ayrı gözlemin Örneğin, aynı ölçü aracının iki formu gibi ne dereceye kadar birbirleri ile tutarlı olduğunu görmek için bu tür geçerlik katsayısı tercih edilmektedir.

3) . Devamlılık ve Eşdeğerlilik Katsayısı

Birinci ve ikinci türde güvenilirliği tek bir süreçte birleştirmek sureti ile elde edilmektedir. Bu defa, birbirine benzer iki ölçü aracı aynı gruba bir süre sonra verilmektedir. Devamlılık ve eşdeğerlilik katsayısı anket, mülakat ve benzeri ölçü araçlarında genellikle tercih edilmektedir, Çünkü, ölçülen özellikle ilgili belirli yaşantılar sonuca daha az etki eder görünmektedir.

4) . İç-tutarlılık Katsayısı (Coefficient of internal consistency)

Bazen de aynı ölçü aracı bir gruba uygulandıktan sonra, sonuçların puvanlanması ölçü aracının belli bir yöntemle göre ikiye ayrılması sureti ile bir güvenilirlik elde edilmektedir. Bu tür bir uygulama sonucu elde edilen katsayıya iç-tutarlılık katsayısı adı verilmektedir. Örneğin, bir ölçekte sonuçlar puvanlanırken, tek ve çift numaralı sorular için iki puvanlama anahtarı hazırlanır. Buna göre her birey için iki puvan elde edilir. Böylece, ortaya çıkan iki seri puvan arasındaki ilişki ölçüğün kendi içindeki tutarlılığını vermiş olmaktadır. Bilindiği gibi burada, davranış örnekleme küçültülmüştür. Bu yüzden elde edilen katsayının, daha önce söz konusu edilen Spearman-Brown formülünden yararlanarak bütün ölçek için düzeltilmesi gerekmektedir. Anlaşılacağı gibi, bu yöntem, ölçüğün maddelerinin homojen bir görünüm gösterip göstermediği yönünde bir bilgi vermektedir. Bu bilgi de bazı maksatlar için güvenilirlik katsayısı olarak kullanılmaktadır.

Buraya kadar yapılan tartışmalardan edinilen bilgiye göre, sosyal Bilimlerde kullanılan ölçü araçlarında geçerlik ve güvenilirlik, her zaman araştırılması gereken teknik niteliklerdir. Bu bilgilere sahip olunmadıkça, neyin ne kadar doğrulukla ölçüldüğü iddiası dayanaksız kalmaktadır. Bir ölçü aracı, şans hatası ile yüklü ise, bireylerin ölçülen özellik içindeki yerlerini doğru olarak saptamak mümkün olmayacaktır. Bu durumun, doğal olarak geçerliğe de olumsuz etkide bulunması beklenebilir. Mamafih bunu kesin bir genel kural olarak düşünmek mümkün değildir. Çünkü, örneğin, iki yarım-test yöntemine göre elde edilen güvenilirlik katsayısı düşük olmasına rağmen, geçerliğin yüksek çıkması olağandır. Ölçü aracının kullanılış maksadı burada büyük rol oynar.

Sosyal Bilimlerde kullanılan anket, mülakat ve benzeri tekniklerde gerçeklik ve güvenilirlik bazen sayısal olarak saptanamayabilir. Bu gibi durumlarda, ölçü araçları geliştirilirken sistematik ve özellikle şans hatalarını azaltacak bir çok tedbirlerin önceden alınması ve ön deneme sonuçlarından yararlanılması gerekmektedir.

YARARLANILAN KAYNAKLAR

- ANASTASI, Anne**, *Psychological Testing*, The Macmillan Company, New York, 1956.
- BEST, John, W.** *Researrh in Education*, Prentice-Hall, Inc., New Jersey, 1970.
- CRONBACH, Lee J.** *Essentials of Psychological Testing*, Harper and Brothers, Publishers, New York, 1960.
- GALTUNG, Johan**, *Theory and Methods of Social Research*, Columbia University Press, New York, 1969.
- GLOCK, Charles Y.**, *Survey Research in the Social Sciences*, Russell Sage Foundation, New York, 1967.
- GOOD, Carter V., and SCATES, Douglas E.**, *Methods of Research*, Appleton Century Crofts, New York, 1960.
- GOODE, J. William and HATT, Paul.**, *Methodes in Social Research*, Mc Graw-Hill Book Company, Inc. New York, 1952.
- GOPAL, M.H.**, *An Introduction to Research Procedure in Social Sciences*, Asia Publishing House, New York, 1970.
- HELMSTADTER, G.C.**, *Research Concepts in Human Behavior*, Appletion-Century-Crofts, New York, 1970.
- HYMEN, Herbert**, *Survey Design and Analysis*, The Free Press, Publishers, Glencoo, Illinois, 1960.
- MONROE, S. Walter.**, (Editor), *Encyolopedia of Educational Research*, The MacMillean Company, New York, 1941, 1950, 1960 ve 1969.
- SELLTİZ. Claire, MARİE, Jahoda, DUTSCH, Mortaon and COOK, Stuart W.**, *Research Methods in Social Relations*, Holt, Rinehart and Winston, New York, 1959.
- TRIVERS, Robart M.W.**, *An Introduction to Educational Research*, The MacMillan Company, New York, 1964.
- WEBB, Eugene J., CAMPBELL, Donald T., SCHWARTZ, Richard D.**, *Unobirusive Research in the Social Sciences*, Rand McNally and Co., Chicago, 1971.
- YOUNG, Pauline V.**, *Scientific Social Surveys and Research*, Prentice-Hall of İndia Private Limited, New Delhi, 1968.