




ÖLÇME ARAÇLARINDA BULUNMASI GEREKEN NİTELİKLER

Dr. İbrahim Yıldırım



Ölçme araçlarından elde edilen bilgilerin sağlıklı bir şekilde amacına hizmet edebilmesi için ölçme aracının ya da elde edilen sonuçların 3 temel özelliğe sahip olması gereklidir.



- **Güvenirlik**
- **Geçerlik**
- **Kullanışlılık**

KULLANIŞLILIK

- Ölçme aracının hazırlama, uygulama ve puanlamasının zaman, emek ve maliyet bakımından **ekonomik** olması demektir.

GÜVENİRLİK

- Güvenirlilik ölçme sonuçlarının **tesadüfi hatalardan arınık olma** derecesidir.
- Güvenirlilik farklı yollarla bir katsayı **(korelasyon)** olarak hesaplanır. Bir testin iki formundan elde edilen puanlar arasındaki korelasyon diyebiliriz.
- Güvenirlilik **negatif olamaz**.
- Bu yüzden korelasyon ile hesaplanan güvenirlilik katsayısı **0 ile 1 arasında** değerler alır.
 - Güvenirlilik katsayısı 1'e yaklaştıkça güvenirliliğin yüksek olduğu yani ölçme sonuçlarına karışan tesadüfi hataların az olduğu anlamına gelir.
- Genelde eğitimde kullandığımız testlerin güvenirliliğinin **en az 0,70** olması beklenir.
- **Güvenirlilik diğer önemli özellik olan geçerlik için ön şarttır.**

- 
- 
- Güvenirlilik, yalnızca kullanılmış olan ölçme aracı tarafından değil, örneklem tarafından da etkilenmektedir.
 - **Bir test güvenilir veya güvenilmez değildir. Test ölçümleri veya eldeki veriler güvenilir veya güvenilmezdir.**

Güvenirliği etkileyen faktörler

➤ Ölçme aracı

- Ölçme yöntemi / aracı ve kullanılan soru sayısı
- Soruların homojenliği
- Yönerge
- Puanlama nesnelliği



➤ Testi alan birey ve gruba bağlı faktörler

- Bireyin sağlık durumu
- Grubun heterojen olması

➤ Uygulama koşulları ve zaman

Güvenirlığe ilişkin 4 kavram

- Duyarlılık
- Kararlılık
- Tutarlılık
- İç tutarlılık

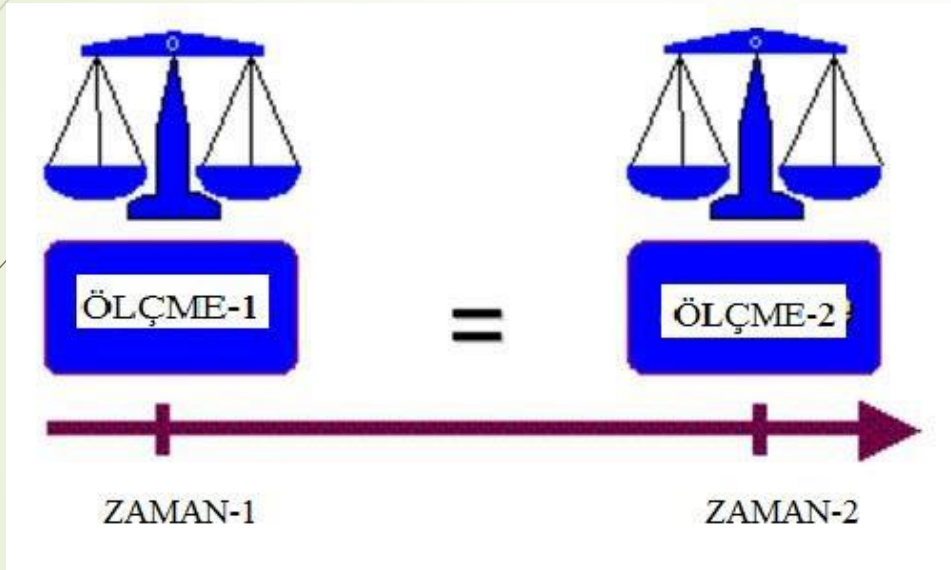
- 
- 
- Birimleri küçük olan ölçme aracı veya sonucu, birimi büyük olandan daha duyarlıdır (santimetre-metre) dolayısıyla daha güvenilirlerdir. Testin madde sayısı arttıkça **duyarlılık** artar.
 - Bir testin farklı 2 formundan elde edilen puanların korelasyonu yüksek ise bu testin **tutarlılığı** yüksektir dolayısıyla güvenilirlerdir.
 - Bir test aynı gruba farklı zamanda uygulanıyor ve benzer sonuçlar alınıyorsa **kararlılık** yüksektir dolayısıyla güvenilirlerdir.
 - Bir testin maddeleri aynı ya da benzer konuları ölçüyorsa bu testin **iç tutarlılığı** yüksektir dolayısıyla güvenilirliği yüksektir.

Güvenirlik hesaplama yöntemleri

GÜVENİRLİK HESAPLAMA YÖNTEMİ	GÜVENİRLİK TÜRÜ
---	Duyarlılık
Test-tekrar test	Kararlılık
Paralel formlar	Tutarlılık
Puanlayıcılar arası güvenirlik	Tutarlılık
Eşdeğer yarılar	İç tutarlılık
KR-20, KR-21, Cronbach Alfa	İç tutarlılık

Test tekrar test

- **Aynı testin aynı gruba farklı zamanlarda** iki kez uygulanmasına dayanan **kararlılık** belirleme yaklaşımıdır.
- Eğer test sonuçları güvenilirse aynı özellik 2 kez ölçüldüğünde aynı sonucu vermelidir.
- Bu durumda güvenirlik katsayısı, aynı bireylerin her iki uygulamadan aldıkları puan dağılımları arasındaki **korelasyon** katsayısıdır. Test-tekrar test güvenirliği [0 ile 1 arası] değerler alır.
- Test-tekrar test güvenirliği ne kadar yüksekse, test puanları, test dışındaki çevresel ve bireysel değişimlere o derece az duyarlı olacaktır.



Öğrenci	1. Test	2.Test
Tolga	85	88
Ersan	76	77
Necip	72	70
Veli	74	81
Oğuzhan	95	90
Olcay	94	94
Gökhan	90	88
Mustafa	68	77
Cenk	78	86
$r = 0,87$		

Paralel formlar

- Eşdeğer/paralel formlar güvenirliği aynı öğrenci grubuna **birbirinin paraleli olan iki testin** birlikte uygulanmasına dayalıdır.
- Ağırlığın 2 farklı terazi ile 1'er kez ölçümü
- Sınavlarda aynı davranışları ölçen fakat farklı sorulardan oluşan iki ayrı test hazırlayarak uygulama
- İki testten elde edilen puanlar arasındaki **korelasyon** katsayısı hesaplanır.
- Bulunan güvenirlik katsayısı +1'e yaklaştıkça eşdeğer formların birbirlerine göre tutarlı ölçmeler yaptığını gösterir bu nedenle bu katsayı **tutarlılık** anlamı taşır.
- **İki testin paralellik şartının sağlanması** için içerik, güçlük, madde sayısı, madde tipi, testlerin ortalaması gibi özellikler açısından eşit olması gerekir.

Puanlayıcılar arası güvenirlik

- Aynı testin sonuçlarının **2 veya daha fazla puanlayıcı** tarafından puanlarının **tutarlı** bulgular sunması beklentisinde olan yaklaşımdır.
- Eğitimde yazılı ve sözlü yoklamaların puanlanmasında, öğrenci gelişim dosyalarının (portfolyo) değerlendirilmesinde puanlayıcılar arası güvenirlik hesaplanabilir.
- Veriler kesikli ise **uyuşma yüzdesi** veya **kappa katsayısı**; sürekli ise **korelasyon katsayısı** belirleyici olarak kullanılabilir.



Yarışmacı	1.Jüri	2.Jüri	Uzlaşma
1	6	6	1
2	7	6	0
3	4	4	1
4	3	2	0
5	4	4	1
			3/5

Bu yarışmada jüriler arası uyuşma oranı 5'te 3 yani 0,6'dır. Bunu yüzde 60 uyuşma olarak da ifade edebiliriz.

Eşdeğer yarılar

- Bu yöntem **bir testin bir defa** uygulanmasına dayanır.
- Bir defa uygulanan test **uygulandıktan sonra eşit iki yarıya** ayrılarak puanlanır.
- Test maddeleri belli bir yöntemle iki yarıya bölünür
- İki yarıdan elde edilen puanlar arasındaki korelasyona bakılır.
- Testin tamamından elde edilen puanlar için güvenilirliği bulmak Spearman-Brown formülü sayesinde gerçekleştirilir.

Sperman Brown Formülü

Formül : $(2 * r) / (1 + r)$

r : eş yarılar arasında bulduğumuz korelasyon

Örnek:

Aşağıdaki tabloda 10 soruluk bir testin doğru cevapları 1, yanlış cevaplar da 0 olarak puanlanmıştır. Tek sayılı sorulardan elde edilen puana birinci yarı, çift sayılı sorulardan elde edilen puana da ikinci yarı adı verilmiştir.

Sorular	1	2	3	4	5	6	7	8	9	10
Şenem	1	0	0	1	1	1	1	1	1	1
Cenk	1	1	1	0	0	1	1	1	0	1
Berke	0	0	0	0	0	1	1	0	0	1
Bilge	1	0	1	1	0	1	1	0	1	1
Hakan	0	0	0	1	0	0	1	0	0	1

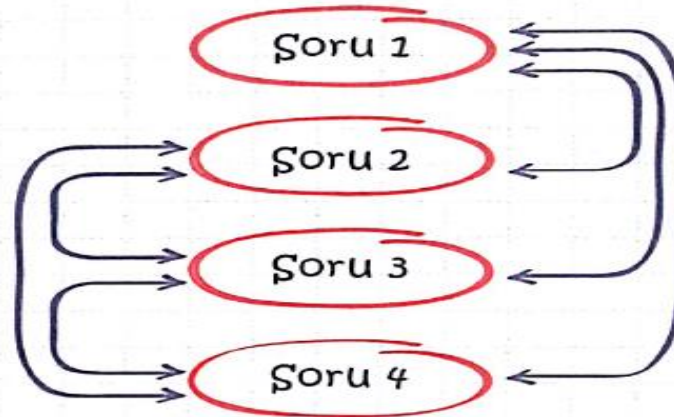
KR-20 ve KR-21

Kuder Richardson
Yöntemi (KR-20,
KR-21 Formülleri):




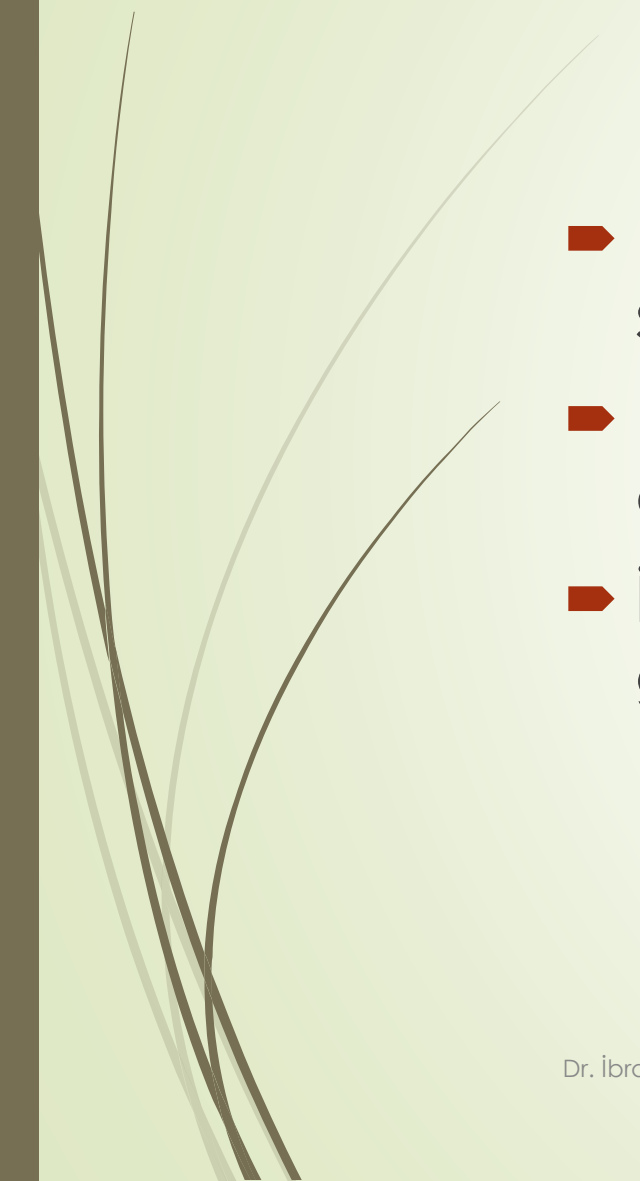
(1 test, 1 uygulama)

Buraya kadar incelenen güvenirlik kestirme yollarında testin tamamından ya da yarısından elde edilen puanlara dayalı iken KR-20 ve KR-21 testin maddelerine ve **bu maddeler arasındaki uyuma dayalı** bir güvenirlik kestirme yoludur. Testin maddeleri arasındaki uyuma dayalı olduğundan içtutarlılığa dayalı bir güvenirlik belirleme yoludur.



Testin olası tüm yarılarından elde edilecek güvenirliklerin ortalamasını verir.

Her sorular tek tek

- 
- 
- KR-20 ve KR-21 güvenirlik katsayıları testteki soruların 1-0 şeklinde kodlanabildiği durumlarda (başarı testi) kullanılır.
 - KR-21, KR-20'ye göre daha düşük değer almakla birlikte aynı anlamı vermektedir.
 - İkisi de iç tutarlılığı belirlerler. 1'e yakın olduğu ölçüde test güvenilirirdir.

Cronbach alfa

- KR20 ve KR21 gibi Alfa da **tek uygulamaya** dayanır.
- Çoğunlukla Likert tipi (katılıyorum, katılmıyorum, asla, bazen, her zaman) maddelerle duyuşsal özelliklerin ölçülmesinde (tutum, ilgi, kişilik vb.) sonuçların güvenilirliği Cronbach Alfa katsayısı ile belirlenir.
- Cronbach Alfa ile hesaplanan güvenilirlik katsayısı da **iç tutarlılık** anlamındadır.
- Alfa güvenilirlik katsayısının yüksek olması güvenilirliği artırır.

Güvenirliđi artırmak için öneriler

- Testteki soru sayısını artırmak
- Şans başarısını azaltmak (seçenek sayısını artırmak)
- Soruları net ve açık şekilde sormak
- Öğrenci düzeyine göre soruyu sormak (yazı büyüklüğü, seçilen cümleler)
- Teste kolay sorularla başlamak (motivasyonu ve güvenilirliđi artırır)
- Sınav koşullarının tüm öğrencilere eşit olması
- Tüm öğrencilere yeterli zaman verilmesi
- Sınav sorularının orta güçlükte olması (ne çok zor ne çok kolay olmalı)
- Ölçme sonuçlarının doğru puanlanıp doğru kaydedilmesi
- Sınavın en başına açıklayıcı bir yönergenin konulması

GEÇERLİK

- **Bir ölçme aracının ölçmek üzere hazırlandığı amacını ölçme derecesidir.**
- Bir ölçme aracının ölçmeyi amaçladığı özelliği, başka herhangi bir özellikle karıştırmadan doğru olarak ölçebilme derecesidir.
- Geçerlik her türlü hatadan (doğrudan etkilendiği sabit ve sistematik hatadır) etkilenir.

GEÇERLİK TÜRLERİ

Ölçüte Dayalı
Geçerlik

Yapı Geçerliği

Kapsam
Geçerliği

Uygunluk
Geçerliği

Yordama
Geçerliği

Geçerliği Etkileyen Faktörler

- Ölçme sonuçlarının güvenilirliği
- Ölçme yöntemi ve soru sayısı
- Uygulama koşulları
- Testin ve amacının uyumu
- Testin kapsamı
- Testin yapısal olarak uygunluğu

Kapsam geçerliđi

- Ölçme aracının, kapsamayı iddia ettiđi alanı veya öğeleri adil ve kapsamlı bir şekilde kapsayıp kapsamadığının incelenmesidir.
- Kapsam geçerliđi bir testin bu testle ölçölmek istenen davranışları ne derece kapsadığı ile ilgilidir.
- Kapsam geçerliđi belirtke tablosu ile ortaya konabilir.

Belirtke tablosu örneđi

Hafta	Kazanımlar	İlgili olduđu soru
1. ve 2. Haftalar	1-2-3	1/2
3. ve 4. Haftalar	4-5-6-7-8-9-10-11	2/3/4
5. Hafta	12-13	5/20
6. 7. ve 8. Haftalar	14-15	6/7/8/9/11
9. Hafta	16-17	7/10/11
10. ve 11. Haftalar	18-19-20	12/13/15/16/17/18
12. ve 13. Haftalar	21	14/15/16/17/18
Dr. İbrahim Yıldırım 14. Hafta	22-23-24-25	16/17/18/19

Yapı geçerliği

- Yapı; birbiriyle ilişkili olan ögelerin oluşturduğu homojen örtüye denir. Doğrudan gözlenemeyen (ör: zeka) psikolojik özellik anlamına da gelmektedir.
- Yapı geçerliği ölçme aracının teoride var olduğu düşünülen, gözle görülemeyen bu yapıları pratikte ölçebilme derecesidir (ör:Matematik).
- **Bir testin yapı geçerliğinin olabilmesi için sorularının birbiriyle yüksek ilişkili olması lazım.**
- Bu da soruların birbiriyle benzer özellikleri ölçtüğü yani aynı yapıyı temsil ettiği anlamına gelir.
- Mesela bir tarih testinin tüm sorularının tarih yapısını ölçmesi gibi. Eğer içinde matematik ölçen bir soru var ise yapı geçerliği bozulur.

Yapı geçerliğini belirlemek için kullanılan metotlar

- Başka testlerle korelasyon
- Grup farkları
- Uzman görüşü
- Pratik ölçülerle korelasyon
- Cevaplayıcıların cevap verme süreçlerinin incelenmesi
- Test puanlarını geliştirici çalışmalar
- Puanların kararlılığı
- Faktör analizi

Ölçüte Dayalı Geçerlik



- Bu yöntem geçerliği sağlamak için **başka bir test puanının ölçüt olarak** kullanılmasına dayanır.
- Böyle bir geçerlik çalışmasında geçerliği kanıtlamak için kullanılan ve geçerliği kanıtlanacak testin puanları ile karşılaştırılan puanlara ölçüt puan adı verilir.
- Geçerliliği kanıtlanacak testin puanları ile ölçüt puanlar arasında **korelasyon** hesaplanır. Bu korelasyon geçerlik katsayısı olarak kullanılır.
- Geçerlik yöntemleri arasında bir geçerlik katsayısı veren tek yöntem ölçüt dayanaklı geçerliktir; $[-1, 1]$ arası değerler alır.

Ölçüte Dayalı Geçerlik

a. Uygunluk geçerliği

b. Yordama geçerliği

- Bir ölçüte dayalı geçerlik, **ölçütün elde edildiği zamana** göre uygunluk geçerliği ve yordama geçerliği olmak üzere ikiye ayrılır.
- Ölçütten elde edilen puanlar, ölçme aracından elde edilen puanlarla **aynı zamanda veya daha önceden** elde edilmişse, bu puanlar arasındaki korelasyona dayalı geçerliğe **uygunluk geçerliği** adı verilir.
- Ölçüt durumundaki puanlar, geçerliği belirlenecek ölçme aracından elde edilen puanlardan **daha sonra elde edilmişse**, bu puanlar arasındaki korelasyona dayalı geçerliğe **yordama geçerliği** adı verilir.

- 
- 
- Örneğin, öğrencilerin lise mezuniyet ortalamaları için ölçüt geçerliğini incelerken:
 - Lisede okula yönelik tutum puanlarını kullanıyorsam uygunluk geçerliği;
 - Üniversite 1. sınıf sene sonu ortalama puanlarını kullanıyorsam yordama geçerliğini inceliyorum demektir.

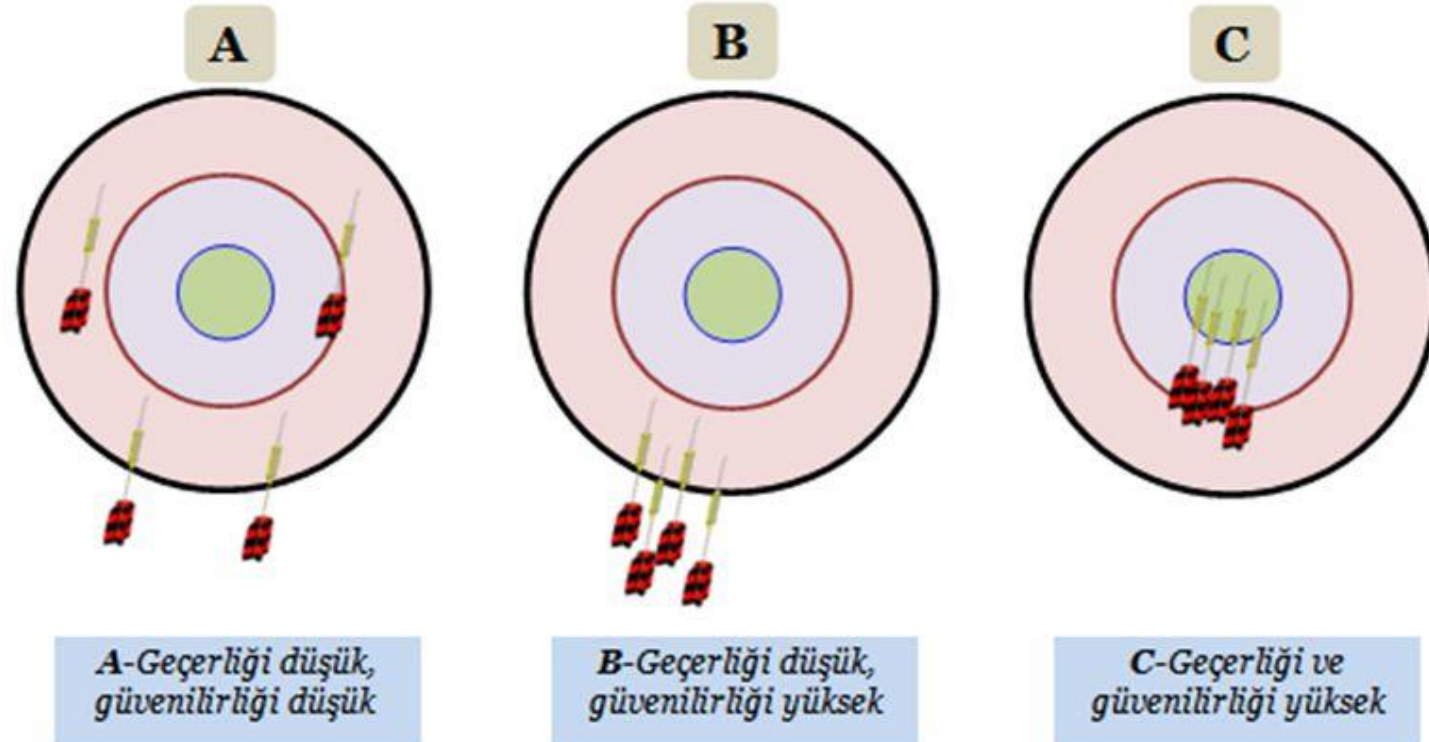
Geçerlik ve Yanlılık

- Test puanları öğrenciler hakkında önemli kararlar vermek için kullanılması nedeniyle olabildiğince geçerli olmasına dikkat edilmeli.
- Eğer bir test ya da test maddelerinden birisi belirli bir gruba avantajlı ise bu durum geçerliği etkiler.
- Eşitlik ilkesi bozulur.
- Test yanlılığı bir grup öğrencinin lehine veya aleyhine işleyen ve istenmeyen bir başka boyutun karışması anlamına gelir.
- KPSS Genel kültür sınavında futbol ile ilgili bir sorunun erkek öğrencileri daha avantajlı hale getirmesi gibi...

Güvenirlik ve Geçerlik İlişkisi

- Güvenirlik ölçme sonuçlarına ait; geçerlik ise ölçme aracında bulunması gereken özelliklerdir.
- Bir teste ait sonuçlar rapor edilirken ikisi bir arada verilir
- Güvenirlik $[0,1]$ arası değerler alırken geçerlik $[-1,1]$ arasında değerler alır.
- Bir ölçme aracından elde edilen sonuçların güvenilir olduğu o ölçme aracının geçerli olduğu anlamına gelmez.
- Sadece tesadüfi hatalardan etkilenen güvenilirlikten farklı olarak, bir testin geçerliği hem tesadüfi hem de sistematik hatalardan etkilenebilir.
- Bu nedenle bir test geçerli olmasa da sonuçları güvenilir olabilir.
- Güvenirlik geçerlik için gerekli fakat yeterli değildir.
- Geçerli olan her testin sonuçları bir derece güvenilir denebilir.
- Geçerli olmayan bir testin sonuçları güvenilir olsa da bir işe yaramaz.

Geçerlik-Güvenirlilik İlişkisi



NİTELDE GEÇERLİK VE GÜVENİLİRLİK

Nitel araştırmalarda geçerlik ve güvenirlik kavramları nitel araştırmaların doğasına uygun düşecek şekilde;

- inandırıcılık (iç geçerlik),
- aktarılabilirlik (dış geçerlik),
- tutarlılık (güvenirlik),
- teyit edilebilirlik (nesnellik)

kavramlarıyla anlam bulmaktadır.

İnandırıcılık (iç geçerlik)

- Verilerin nitelikli toplanması, aynı konuya ilişkin farklı veri kaynakları ile **veri çeşitlendirilmesi** böylece çıkarılan anlamın yani bulguların uygunluğunun değerlendirilmesi olarak tanımlanabilir.
- Bunun için araştırmacı, raporunu okuyucuya kendi izlediği yolun ve elde ettiği sonuçların tutarlılığını değerlendireceği açıklıkta sunmalıdır.

Aktarılabilirlik (dış geçerlik)

- **Benzer bağlamlarda (benzer ortam ve bireylerle) ilgili sonuçların geçerli olabileceğine** ilişkin anlayış geliştirme anlamında kullanılabilir.
- Ayrıntılı betimleme okuyucuyu bağlama götürdüğü hesaba katılarak amaçlı örnekleme ve ayrıntılı/derin betimleme yapılması aktarılabilirliğe olumlu katkı sağlamaktadır.

Tutarlılık (güvenirlik)

- Araştırma benzer bağlamlarda benzer özelliklere sahip bireylerle **tekrarlandığında** benzer sonuçlar oluşacağına ilişkin kanıtlar sunmaktır.

Teyit edilebilirlik (nesnellik)

- Bulguların araştırmacının önyargılarının veya olayı algılamasının bir yansıması değil **gerçekten araştırma verilerinin ürünü** olarak görülme derecesidir.