

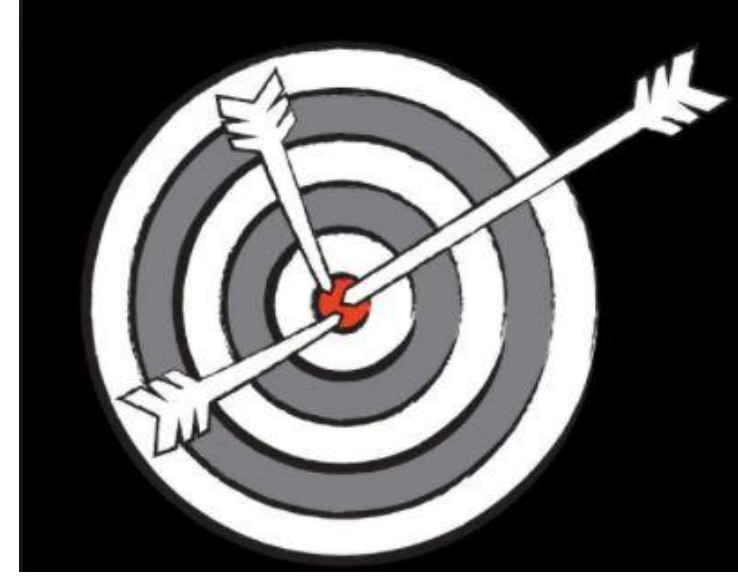
ÖLÇME ARAÇLARINDA ya da SONUÇLARINDA BULUNMASI GEREKEN NİTELİKLER

• 2. SUNUM

- 3 temel özellik
- Ölçmede hata ve hata türleri
- Güvenirlik
- Güvenirlik hesaplama yöntemleri
- Ölçmenin standart hatası
- Geçerlik
- Kullanışlılık

Ölçmede Gerekli 3 Özellik

- ▶ 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**
- ▶ Geçerlik ve güvenilirlik kavramları hata kavramı ile bağlantılı olduğu için bu konulara geçmeden önce ölçmede hata ve hata türlerini bilmekte fayda var.



Ölçmede Hata

- ▶ Gerek fiziksel bilimlerde gerekse sosyal bilimlerde ölçüm sonuçlarına mutlaka bir miktar hata karışır ve bu hatanın miktarı kesin olarak bilinemez.
- ▶ Eğitim ve psikolojide ölçülen özelliklerin doğası gereği hata olasılığı daha yüksektir.
- ▶ Hatayı ortadan kaldırmak mümkün olmasa da hata kaynaklarını belirleyip en aza indirmek mümkün olabilir.
- ▶ Ölçme sonuçlarına karışan hatalar, ölçmecinden, ölçme aracından, ölçme yönteminden, ölçmenin yapılacağı ortamdan ve üzerinde ölçme yapılan birey ya da nesnelerin bu etkenlerle etkileşiminden kaynaklanabilir.
- ▶ Ölçmede hata ölçülen özelliğin ***gerçek değeri ile gözlenen değeri arasındaki fark*** olarak tanımlanır.
- ▶ Ölçmede 3 çeşit hata bulunmaktadır; **sabit hata, sistematik hata ve tesadüfi hata.**

Sabit Hata

- ▶ Bir ölçmeden diğerine miktarı ve yönü değişmeyen hatalardır.
- ▶ Yapılan her ölçmeye aynı miktarda karışan hatalardır.
- ▶ Ölçmeye konu olan nesneler üzerinde yapılan bütün ölçmelerde yapılan hata miktarı birim olarak (kg, cm) eşit olmalıdır.
- ▶ Oran verilmesi durumunda hata türü sabit olmaz. ör: her öğrencinin notuna yüzde 10 eklenmesi.
- ▶ Sabit hatalarda hatanın kaynağı, yönü ve miktarı bilinir. Bu nedenle sabit hataları önlemek mümkündür.
- ▶ Ucu 2cm kırık olan 2m uzunluğundaki bir cetvel ile sınıftakilerin boyunun ölçülmesi (her seferinde 1 kez kullanılacak)
- ▶ Süresi az verilen 5 soruluk sınav

Sistematiik Hata

- ▶ Belli bir kurala göre ortaya çıkan hatalara sistematiik hatalar denir.
- ▶ Hatanın kurallı (sistemli) olması.
- ▶ Burada hatanın miktarı ölçülen büyüklüğe, ölçmeciye veya ölçme koşullarına ve belli bir duruma bağılı olarak değişir.
- ▶ Ölçmeden ölçmeye miktarı ve yönü değişen hata türüdür.
- ▶ Ucu kırık cetvelin her ölçüm için farklı sayıda kullanılması,
- ▶ Sedat Hoca'nın yaptığı bir sınavda Beşiktaş'lı öğrencilere 5'er puan fazla vermesi, (Yanlılık)
- ▶ Yazısı güzel olan öğrencilere 1'er puan fazla verilmesi,
- ▶ Her öğrenciye notunun %10'u kadar ekleme yapılması.
- ▶ Sistematiik hatada ölçme konularına karışan hatanın miktarı, yönü ve kaynağı bilinir.

Tesadüfi (Rastgele) Hata

- ▶ Ölçme sonuçlarına *gelişigüzel* karışan, kaynağı, yönü ve miktarı tam olarak bilinmeyen hatalardır.
- ▶ Ölçmeden ölçmeye miktarı ve yönü değişen hatalardır.
- ▶ Bazen ölçmeyi yapan kişinin yorgunluğu ya da dikkatsizliği, bazen ölçmenin yapıldığı ortamdan, bazen ölçme yapılan bireyden bazen de ölçme aracının özelliklerinden ölçmeye karışan hatalardır.
- ▶ Ölçme sonuçlarına karışan tesadüfi hatalar güvenilirliği düşürür.
- ▶ Bazen pozitif bazen negatif yönde olabilir.

Hata Kaynakları:

- ▶ Ölçmeyi yapan kişinin yorgun olması, öznelliği vb.
- ▶ Testin çok kısa ya da çok uzun olması, maddelerin iyi ifade edilememesi.
- ▶ Testin yapıldığı sınıfın sıcaklığı, ışık, havalandırma ve gürültü düzeyi.
- ▶ Ölçmenin yapıldığı grubun ölçülen özellik açısından homojen ya da heterojen bir yapıya sahip olması.

Korelasyon

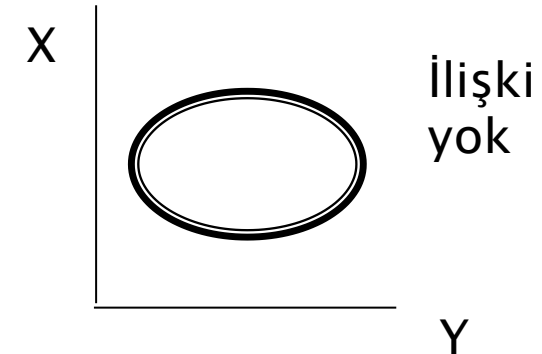
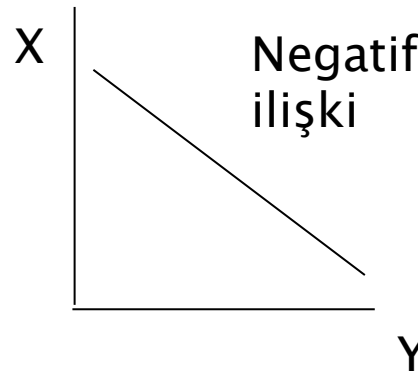
- ▶ Korelasyon (ilişki), iki değişkenin birlikte değişiminin bir ölçüsüdür.
- ▶ Boy uzunluğu ile kilo arasındaki ilişki, yaş ile boy arasındaki ilişki, çalışma saati ve sınav puanı arasındaki ilişki, hava sıcaklığı ve doğalgaz tüketimi arasındaki ilişki eğer sayısal veriler varsa korelasyon katsayısı cinsinden gösterilebilir.
- ▶ Korelasyon katsayısı matematiksel olarak -1 ile $+1$ arasında değerler alır.
- ▶ Korelasyonun **büyüklüğü** ($0-1$) iki değişken arasındaki ilişkinin güçlülüğünü gösterirken **işareti** ($+, -$) değişkenlerin aynı yönde ($+$) artıp azaldığını ya da zıt yönlerde ($-$) artış ve azalış gösterdiğini belirtir.
- ▶ Hava sıcaklığı ve doğalgaz tüketimi arasındaki ilişki **NEGATİF**
- ▶ Çalışma saati ve sınav puanı arasındaki ilişki **POZİTİF** olabilir.
- ▶ Eğer iki değişken arasında hiç ilişki yoksa korelasyon katsayısı sıfır çıkar. Eğer bu iki değişken birbiyle yüzde yüz oranında ilişkili ise korelasyon maksimum (1) değeri (mükemmel ilişki) alır.

Korelasyonun Gösterilmesi

- Korelasyon **matrisleri** ya da **grafikler** kullanarak gösterilir.

	Tarih	Matematik	Fizik	Coğrafya	Kimya
Tarih	1	-0,3	0,2	0,8	0,3
Matematik	-0,3	1	0,9	0,1	0,85
Fizik	0,2	0,9	1	0,2	0,96
Coğrafya	0,8	0,1	0,2	1	0,4
Kimya	0,3	0,85	0,96	0,4	1

Matematik ile Fizik arasındaki korelasyon değeri=0,9 iken Matematik ile Tarih arasında bu değer -0,3'tür.



1. Güvenirlik

- ▶ **Güvenirlik** ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesidir.
- ▶ Bir ölçme sonucu tesadüfi hatalardan ne kadar arınık olursa o kadar güvenilir olur.
- ▶ Güvenirlik farklı yollarla bir katsayı (korelasyon) olarak hesaplanır. Bir testin iki formundan elde edilen puanlar arasındaki korelasyon diyebiliriz.
- ▶ Güvenirlik negatif olamaz. Bu yüzden korelasyon ile hesaplanan güvenilirlik katsayısı 0 ile 1 arasında değerler alır.
- ▶ Güvenirlik katsayısı sıfıra yaklaştıkça güvenilirliğin düşük olduğu yani ölçme sonuçlarına karışan tesadüfi hataların çok olduğu anlamına gelir.
- ▶ Güvenirlik katsayısı 1'e yaklaştıkça güvenilirliğin yüksek olduğu yani ölçme sonuçlarına karışan tesadüfi hataların az olduğu anlamına gelir.
- ▶ Genelde eğitimde kullandığımız testlerin güvenilirliğinin en az 0,70 olması beklenir.
- ▶ **Güvenirlik** diğer önemli özellik olan **geçerlik** için ön şarttır.

Güvenirliğin Farklı Anlamları

- ▶ Duyarlılık
 - ▶ Kararlılık
 - ▶ Tutarlılık
 - ▶ İç tutarlılık
-
- ▶ Birimleri küçük olan ölçme aracı veya sonucu, birimi büyük olandan daha duyarlıdır (santimetre–metre) dolayısıyla daha güvenilirdir. Testin madde sayısı arttıkça duyarlılık artar.
 - ▶ Bir özelliği benzer koşullar altında aynı araçla birden fazla ölçüp tutarlı sonuçlar bulunuyorsa bu ölçme aracının güvenilirliği yüksek demektir.
 - ▶ Eğer bir testin 2 formundan elde edilen puanların korelasyonu yüksek ise bu testin tutarlılığı yüksektir dolayısıyla daha güvenilirdir.
 - ▶ Bir testin maddeleri aynı ya da benzer konuları ölçüyorsa bu testin iç tutarlılığı yüksektir dolayısıyla güvenilirliği yüksektir.

Güvenirlilik Türleri ve Belirleme Yöntemleri

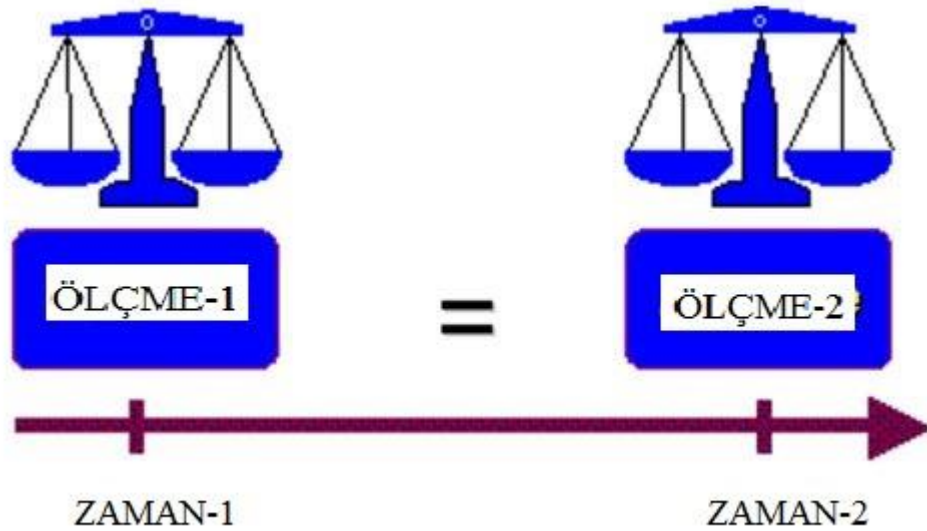
- Güvenirliğin farklı anlamları olduğu gibi farklı güvenirlik hesaplama yöntemleri de mevcuttur.

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

Test–tekrar Test

- ▶ Aynı testin aynı gruba farklı zamanlarda iki kez uygulanmasına dayanır.
- ▶ 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.

Test-tekrar Test



Öğrenci	1. Test	2. Test
Tolga	85	88
Ersan	76	77
Necip	72	70
Veli	74	81
Oğuzhan	95	90
Olçay	94	94
Gökhan	90	88
Mustafa	68	77
Cenk	78	86

$$r=0,87$$

Test-tekrar Test

- ▶ İki farklı zamandaki ölçmelerden farklı sonuçlar elde edilmesinin iki sebebi vardır birincisi ölçme aracı ikincisi ölçülen bireye ait değişikliklerdir.
- ▶ İki ölçme uygulamasının arasının çok kısa tutulması sorulan soruların ve yanıtların kolay hatırlanmasına, çok uzun tutulması ise, bireylerin ölçülen özellik açısından yeni kazanımlara neden olabilir.
- ▶ Psikolojide, zeka, kişilik, stres, kaygı, mutluluk vb. kavramların ölçümünde kullanılır başarı testlerinde kullanılması tavsiye edilmez.

Eş Değer (Paralel) Formlar

- ▶ Eşdeğer/paralel formlar güvenirliği aynı öğrenci grubuna birbirinin paraleli olan iki testin uygulanmasına dayalıdır.
- ▶ Ağırlığın 2 farklı terazi ile 1 er kez ölçümü
- ▶ Sınavlarda aynı davranışları ölçen fakat farklı sorulardan oluşan iki ayrı test hazırlanarak uygulanır
- ▶ 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.
- ▶ 2 form uygulanırken arada zaman bırakmaya gerek yoktur. Aynı anda uygulanabileceği gibi farklı zamanlarda da uygulanabilir. AVANTAJ
- ▶ İki testin paralellik şartının sağlanması için içerik, güçlük, madde sayısı, madde tipi, testlerin ortalaması gibi özellikler açısından eşit olması gerekir. DEZAVANTAJ

Eş Değer Formlar–Örnek

- Bir öğretmen matematik sınavı için 2 tane 10 soruluk test (A ve B) hazırlamıştır. Bu 2 testteki soruları hazırlarken içerik, kapsam, madde tipi vb. açısından eşit (ya da çok benzer) olmasına dikkat etmiştir. İlk olarak sınıftaki öğrencilerin yarısına A testini yarısına da B testini vermiştir. Öğrenciler ilk testi tamamladıktan sonra daha önce A testini alanlara B testini ve daha önce B testini alanlara da A testini dağıtarak her bir öğrencinin 2 testi de cevaplamasını sağlamıştır. Sınav bittikten sonra öğrencilerin 2 testten aldıkları puanların korelasyonunu hesaplayarak tutarlılık anlamındaki güvenirlik katsayısını hesaplamıştır.

Eşdeğer Yarılar

- ▶ Bu yöntem 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.
- ▶ Test, ikiye bölünerek korelasyon hesaplandığından testin sadece yarısının güvenilirliği belirlenmiş olur.
- ▶ Testin tamamından elde edilen puanlar için güvenilirliği bulmak Spearman–Brown formülü sayesinde gerçekleştirilir.

r = Eşdeğer yarılar yöntemiyle bulunan testin yarısına ait korelasyon ise “Testin bütünü” için korelasyon değerini bulmak için şu formül kullanılır:

$$r_{\text{TEST}} = 2r / (1 + r)$$

Eşdeğer Yarılar

- ▶ Diyelim ki eş değer yarılar yöntemi ile testin yarısından elde edilen korelasyon 0,6 olsun.
- ▶ Bu durumda testin bütünü için korelasyon değeri = $2 \times 0,6 / (1 + 0,6)$ ile hesaplanır yani $1,2 / 1,6 = 0,75$ olarak bulunur.
- ▶ Bu yöntemin uygulama aşamaları şöyledir:
 - 1) Tüm test bir defada bir gruba uygulanır.
 - 2) Puanlama yapılırken test ikiye ayrılır.
 - 3) Her bir öğrencinin yarılardan aldıkları puanlar belirlenir.
 - 4) Öğrencinin yarılardan aldıkları puanlar arasında korelasyon hesaplanır.
 - 5) Testin bütününe ilişkin korelasyon katsayısı $2r / (1 + r)$ ile hesaplanır. Hesaplanan bu değer iç tutarlılık anlamına gelir. Bulunan değer 1'e ne kadar yakın ise iç tutarlılık (güvenirlilik) o kadar yüksektir denir.

KR-20 ve KR-21 Yöntemleri

- ▶ KR-20 ve KR-21 güvenirliği testteki soruların birbirleriyle uyumunu, tutarlılığını gösteren güvenirlik hesaplama yöntemleridir.
- ▶ Testin bir kere uygulanması yeterlidir.
- ▶ KR-20 hesaplanırken testteki sorular 1 ve 0 (doğru-yanlış) şeklinde puanlanır ve madde güçlük indeksini içeren formül kullanarak hesaplanır.
- ▶ KR-21 hesaplanırken madde güçlük indeksi yerine test puanlarının ortalaması 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.

KR-20 ve KR-21 Yöntemleri

$$KR_{20} = \frac{K}{K-1} \left[1 - \frac{\sum pq}{\sigma^2 X} \right]$$

KR-20 Formülü

$$KR21 = \left(\frac{k}{k-1} \right) \left(1 - \frac{X(k-X)}{k s^2} \right)$$

KR-21 Formülü

Cronbach Alfa

- ▶ KR20 ve KR21 gibi Alfa da tek uygulamaya dayanır.
- ▶ KR-20 ve KR-21 güvenirlik katsayıları testteki soruların 1-0 şeklinde kodlanabildiği durumlarda (başarı testi) kullanılır.
- ▶ **Çoğunlukla Likert tipi (katılıyorum**
- ▶ **katılmıyorum, asla, bazen, her zaman) maddelerle duyuşsal özelliklerin ölçülmesinde (tutum, ilgi, kişilik vb.) sonuçların güvenirliği Cronbach Alfa katsayısı ile belirlenir.**
- ▶ Cronbach Alfa ile hesaplanan güvenirlik katsayısı da iç tutarlılık anlamındadır.
- ▶ Alfa güvenirlik katsayısının yüksek olması güvenirliği artırır.

Cronbach Alfa

$$\alpha = \left(\frac{K}{K-1} \right) \left(1 - \frac{\sum V_i}{V_T} \right)$$

Cronbach Alfa formülü

Puanlayıcılar Arası Güvenirlik

- ▶ Eğitimde yazılı ve sözlü yoklamaların puanlanmasında
- ▶ Performansa dayalı değerlendirmelerde
- ▶ Öğrenci gelişim dosyalarının (portfolyo) değerlendirilmesinde **puanlayıcılar arası güvenirlik** hesaplanabilir.

Birden çok puanlayıcının birbirinden bağımsız puanlama yapması durumunda, yapılan puanlamaların birbirleri ile ne kadar tutarlı olduklarını belirlemek için kullanılan tutarlık anlamında bir güvenirlik türüdür. Uyuşma yüzdesi ya da Kappa istatistiği kullanılarak hesaplanabilir.

Puanlayıcılar Arası Güvenirlik Örnek

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.

Ölçmenin Standart Hatası

- ▶ Ölçülen özelliğin gerçek değerine G , ölçme sonuçlarına $\bar{O}$ ve ölçmenin standart hatasına S_e diyecek olursak bu üçü arasındaki ilişki; $G = \bar{O} \pm S_e$

- ▶ Bu eşitlikte ölçmenin sonuçları bilinen değerdir.
- ▶ S_e (ölçmenin standart hata) da şu formül ile hesaplanır:
- ▶ S_e : Ölçmenin standart hatası
- ▶ S : Test puanlarının standart kayması (sapması)
- ▶ r : Test sonuçlarının güvenirliği

$$S_e = S\sqrt{1 - r}$$

Örn: Güvenirliği (r) 0,64 olan bir teste ait puanların standart sapması (S) 3 ise ölçmenin standart hatası (S_e) kaçtır?

Çözüm:
$$S_e = 3\sqrt{1 - 0,64} = 3\sqrt{0,36} = 1,8$$

Hata ile güvenirlik ters orantılıdır!!!

Gerçek Değerin Kestirilmesi

- ▶ Ölçme sonucu ve standart hata hesaplandıktan sonra gerçek değer $G = \bar{O} \pm S_e$ formülü kullanılarak aşağı/yukarı tahmin edilebilir.
- ▶ Diyelim ki bir öğrencinin test puanı 80 ve standart hata 10 olarak hesaplandı o zaman bu öğrencinin gerçek puanı 80-10 ve 80+10 arasında yani 70 ile 90 arasındadır diyebiliriz. Bu aralıklar hesaplanırken belli olasılık değerleri kullanılır (%68, %95, %99).
- ▶ Eğer +1SE ve -1SE kullanırsak %68 olasılıkla o aralıktadır diyebiliriz.
- ▶ %68 olasılık= $G = \bar{O} \pm 1S_e$
- ▶ %95 olasılık= $G = \bar{O} \pm 2S_e$
- ▶ %99 olasılık= $G = \bar{O} \pm 3S_e$

Gerçek Değerin Kestirilmesi–Örnek

- ▶ Ölçme hatası 3 olan bir testte, ölçülen test puanı 77 olan bir öğrencinin gerçek puanını %68, %95 ve %99 olasılıkla tahmin edersek gerçek puan hangi aralıklarda olur?
- ▶ %68 olasılıkla $77-(1 \times 3)$ ve $77+(1 \times 3)$ yani **74** ve **80** arasında
- ▶ %95 olasılıkla $77-(2 \times 3)$ ve $77+(2 \times 3)$ yani **71** ve **83** arasında
- ▶ %99 olasılıkla $77-(3 \times 3)$ ve $77+(3 \times 3)$ yani **68** ve **86** arasında yer alır.

Güvenirliği Artırma Yolları

- ▶ 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 güvenilirliği 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

2. Geçerlik

- ▶ Ölçme aracının ölçmek istediği özelliği ölçebilme derecesidir.
- ▶ Bir testin ölçmek istediği özelliği başka özelliklerle karıştırmadan (veya daha az karıştırarak) ölçebilmesi geçerli olduğu anlamına gelir.
- ▶ Yani bir ölçme aracı amacına hizmet ettiğinde geçerlidir.
- ▶ Ölçme aracının amacına uygunluğu olarak da tanımlanır.
- ▶ Yani bir testle ne amaçlanıyorsa onun gerçekleştirilmesi ölçüsünde geçerli olmasından bahsedilir.
- ▶ Geçerlik $[-1, 1]$ arasında değerler alır.

Geçerlik Türleri



Geçerliği ortaya koymada birden fazla geçerlik kanıtına ihtiyaç vardır!!!

Görünüş Geçerliği

- ▶ Bir testin neyi ölçüyor görüldüğü o testin ***görünüş geçerliği***dir.
- ▶ Ya da bir testin ölçmek istediği şeyi ölçüyor görünmesidir.
- ▶ Matematik bilgisi ölçmek için hazırlanan bir teste hızlı bir bakış attığımızda tüm soruların matematikle ilgili olması,
- ▶ İngilizce seviyesini ölçen bir sınava göz attığımızda tüm maddelerin İngilizce içerikte olması
görünüş geçerliği sağlandığının işareti olabilir.
- ▶ **Görünüş geçerliği teknik anlamda bir geçerlik türü olmayıp, testin ölçmek istediği özelliği ölçüyor görünmesi olarak algılanabilir. Fakat bir geçerlik yöntemi olarak sayan görüşler de mevcuttur.**

Görünüş Geçerliği

- ▶ İlköğretim düzeyinde geliştirilmiş bir test, yetişkinler üzerinde deneme uygulaması yapılırsa testi alanlar bu tür soruları cevaplamaya direnç gösterebilir.
- ▶ Bilgisayar mühendisliğindeki öğrencilerin matematiksel işlem becerileri ölçülürken “10 kg portakal 20 TL ise, 1 kg portakal kaç TL?” diye sormak görünüş geçerliliğini en aza indirebilir, ortadan kaldırabilir.

Kapsam Geçerliği

- ▶ Eğitimde çoğunlukla testler öğrencilerin belli bir alana ya da derse ilişkin davranışları (bilgileri) edinip edinmediğini anlamak için uygulanır.
- ▶ Bu durumda testi geliştiren kişi, testin, bu davranış ya da bilgileri yeteri kadar içerip içermediğini araştırması gerekir.
- ▶ Testin amaçlarının testte bulunan sorularla uyumlu olup olmadığı incelenmelidir.
- ▶ Kapsam geçerliği bir testin bu testle ölçülmek istenen davranışları *ne derece kapsadığı* ile ilgilidir.
- ▶ Bir testin kapsam geçerliği, ölçmek istediği alan kapsamını *ne kadar temsil ettiği* ile alakalıdır.
- ▶ Kısaca bir testin ölçmek istediği konudaki hedef–davranışları *ne kadar iyi örneklediği* kapsam geçerliği ile ilintilidir.

Kapsam Geçerliği Nasıl Test edilir?

- ▶ Kapsam geçerliği belirlenirken en çok *uzman görüşüne* başvurulur. (örn:uzmanların bir araya geldiği bir panel yapmak)
- ▶ Örnek: Bir matematik testinin kapsam geçerliği o testteki maddelerin matematik öğretmenlerine inceletilip onay alınması ile sağlanabilir.
- ▶ İkinci yaygın yöntem de *Belirtke Tablosu* hazırlanmasıdır. Belirtke tablosu bir boyutunda konuların bir boyutunda hedef–davranışların bulunduğu bir tablodur. Bu tablo sayesinde ölçme kapsamındaki tüm davranışların teste soru olarak alınması sağlanır.
- ▶ Üçüncü bir yöntem de aynı kapsamı ölçen bir testle karşılaştırmaktır. Bir testi ölçtüğü bilinen bir testle aynı kapsamı ölçtüğü düşünülen testler arasındaki korelasyona dayanır.

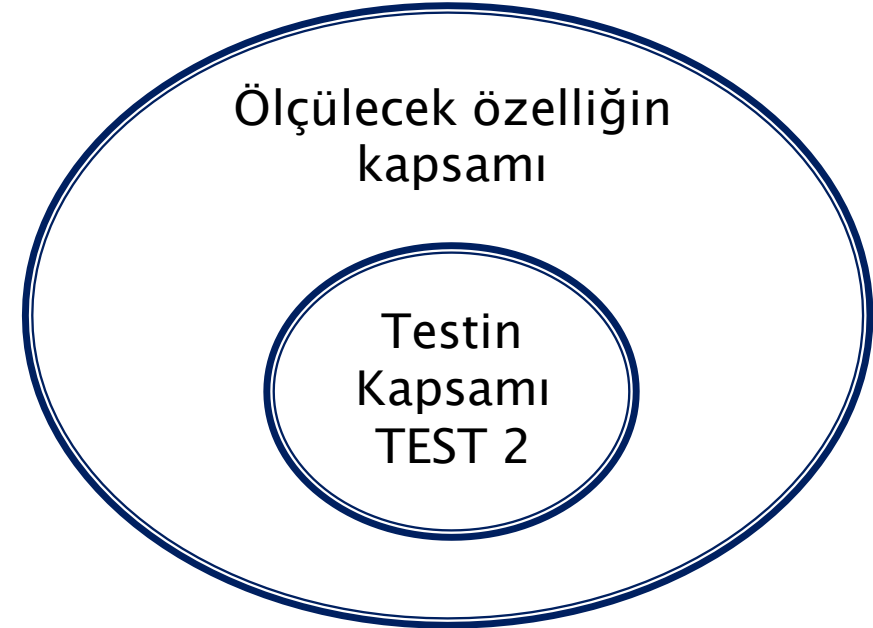
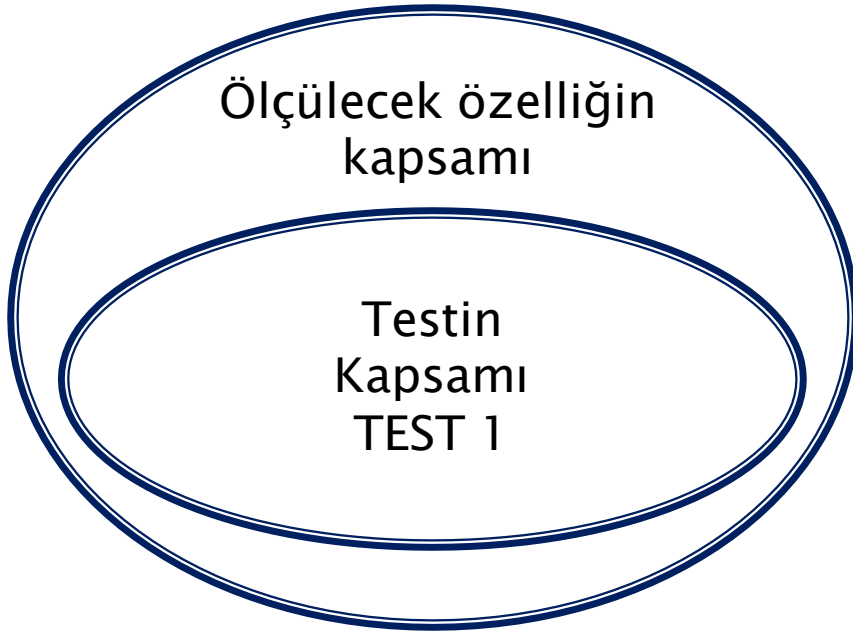
Belirtke Tablosu Örneđi

Bir ölçme değeriendirme dersine ait ara sınav sorularının dağılımı aşağıdaki gibi olabilir:

Konular	Bilgi	Kavrama	Uygulama	Analiz	Sentez	Değeriendirme	Toplam
Ölçmenin temel kavramları	2	2	2	1	-	1	8 (%32)
Geçerlik	1	2	-	1	2	-	6 (%24)
Güvenirlik	1	2	1	2	2	-	8 (%32)
Kullanışlılık	1	1	-	1	-	-	3 (%12)
Toplam (%)	5 (%20)	7 (%28)	3 (%12)	5 (%20)	4 (%16)	1 (%4)	25 (%100)

Kapsam Geçerliğinin Şematik Olarak Gösterilmesi

Hangi testin kapsam geçerliliği fazladır?



Kaynak: Nükhet Çıkrıkçı Demirtaşlı (2012). *Eğitimde Ölçme ve Değerlendirme*. Edge akademi, Ankara.

Ölçüt Dayanaklı Geçerlik

- ▶ Bu yöntem geçerliği sağlamak için başka bir test puanının ya da puanlarının 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.
- ▶ Bir ölçüte dayalı geçerlik, ölçütün elde edildiği zamana göre *uygunluk geçerliği* ve *yordama geçerliği* olmak üzere ikiye ayrılır.

Uygunluk (Zamandaş) Geçerliği

- ▶ 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.
- ▶ Örneğin YGS'nin geçerliğini belirlemek istersek ve bu amaçla öğrencilerin lise notlarını kullanırsak geçerlik uygunluk geçerliği olur. Burada geçerliği sağlanan puanlar YGS puanları ölçüt puanlar ise lise notlarıdır.
- ▶ 2 puan dağılımı arası korelasyona bakılır.
- ▶ Zamandaş, eş zamanlı, halihazır, hemhal kavramları uygunluk geçerliği için kullanılan diğer adlandırmalardır.

Yordama Geçerliđi

- ▶ Yordama bir tahmindir ve geleceđe yöneliktir!
- ▶ Bir testin öğrencinin gelecek performansını tahmin edebilme derecesidir.
- ▶ Ölçüt durumundaki puanlar, geçerliđi belirlenecek ölçme aracından elde edilen puanlardan daha sonra elde edilmişse, bu puanlar arasındaki korelasyona dayalı geçerliđe **yordama geçerliđi** adı verilir.
- ▶ YGSde yüksek puan alan öğrencilerin üniversitede başarılı olacağını söylemek bir tahmindir (yordama).
- ▶ Bunu ispatlamak için YGS puanları ile üniversite ortalamalarının korelasyonuna bakmak yeterlidir. Eğer korelasyon yüksek ve pozitif çıkarsa YGS üniversite başarısı için geçerli bir yordama yöntemidir denebilir.
- ▶ yordayıcı/yordanan
- ▶ Burada önemli olan ölçüt puanlarının geçerliđi kanıtlanacak puanlardan sonra toplanması.

Yordamak: Bilinen veya gözlenen durumlardan yola çıkarak bilinmeyen veya gözlenmeyen durumlar hakkında tahminde bulunmak(tdk.gov.tr)

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 Belirlemede Kullanılan Yöntemler

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

Yapı Geçerliğini Belirlemede Kullanılan Yöntemler

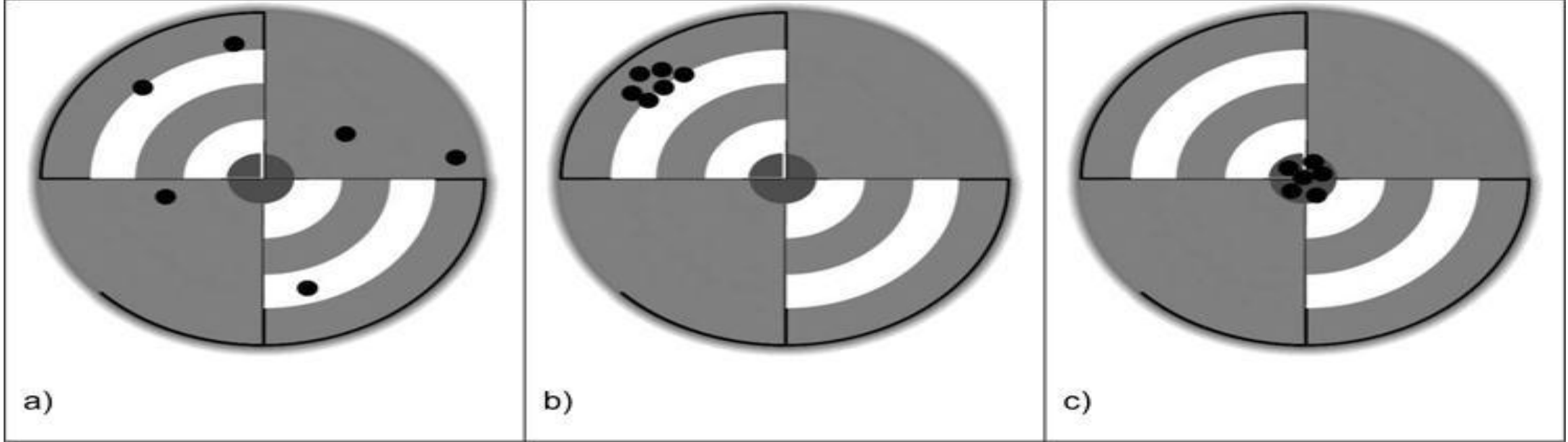
- ▶ Benzer ya da zıt testlerle (ya da daha pratik testlerle) olan korelasyonu bir testin yapı geçerliliği hakkında bilgi verir.
- ▶ Yapı geçerliği, ölçülmek istenen yapının farklı gruplarda incelenmesine (Grup farkları).
- ▶ Örneğin, sınav kaygısını ölçmek üzere geliştirilmiş bir ölçme aracı, daha önceden belirlenmiş sınav kaygısı yüksek ve normal bireylere uygulandığında, kaygısı yüksek olan bireyler testten yüksek puan, normal olanlar da düşük puanlar alıyorsa yapı geçerliği hakkında olumlu yorumlar yapılabilir. Bu yöntem farklı-zıt gruplar çalışmaları (contrasted group studies) olarak adlandırılmaktadır.
- ▶ Cevaplayıcıların testi cevaplarırken soruları sesli yanıtlamaları ve ne düşündüklerini söylemesi testin yapı geçerliği hakkında bilgi verir.
- ▶ Test-tekrar test (kararlılık) yüksek ise yapı geçerliği de yüksektir.
- ▶ Faktör analizi bir testin maddeleri arasındaki korelasyon değerlerine bağlı olarak o testin arkasında kaç tane faktör(yapı) olduğunu gösteren istatistiksel bir yöntemdir.

Güvenirlilik ve Geçerlik İlişkisi

- ▶ Güvenirlilik ölçme sonuçlarına ait; geçerlik ise ölçme aracında bulunması gereken özelliklerdir.
- ▶ Bir teste ait sonuçlar rapor edilirken ikisi bir arada verilir
- ▶ Güvenirlilik $[0,1]$ 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 ekilenen güvenirlilikten farklı olarak, bir testin geçerliği hem tesadüfi hem de sistematik hatalardan etkilenebilir.
- ▶ Bu nedenle bir test geçerli olmasa da sonuçları güvenilir olabilir.
- ▶ Güvenirlilik 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.

$$\text{Maksimum Geçerlik} = \sqrt{\text{Güvenirlilik}}$$

Güvenirlilik ve Geçerlik



- ▶ a) Güvenilmez ve geçersiz ölçüm
- ▶ b) Güvenilir ama geçersiz ölçüm,
- ▶ c) Güvenilir ve geçerli ölçüm.

Geçerliği Neler Düşürür?

- ▶ Güvenirliğin düşük olması geçerliği de düşürür.
- ▶ Sabit ve sistematik hatalar
- ▶ Her türden yanlılık
- ▶ Sınav öncesi öğrencilerin soruları görmesi
- ▶ Yazı güzelliği, kağıt düzeni gibi testle ölçülmek istenmeyen davranışlara puan verilmesi
- ▶ Kopya ve kopyaya neden olabilecek herşey
- ▶ Soruların ipucu ya da şansla cevaplanabilmesi
- ▶ Sınav süresinin uzun olması (kopya riski)
- ▶ Sınav süresinin çok kısa olması
- ▶ Soruların açık ve net ifade edilememesi gibi durumlar geçerliği düşürebilir...

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...

3. Kullanışlılık

- ▶ Ekonomiklik–Uygulanabilirlik–Puanlanabilirlik
 - ▶ Testin hazırlanması ve puanlanmasındaki zaman, para ve emek bakımından ekonomik olması
 - ▶ Ölçme araçlarının hazırlanma süresi
 - ▶ Testin uygulanma süresi
 - ▶ Uygulama kolaylıkları
 - ▶ Testin formatı
 - ▶ Puanlama kolaylıklarının tümü
- Kullanışlılık özelliğini vurgular.

Kaynakça

- ▶ Atılgan, Hakan. (2014). *Eğitimde ölçme ve değerlendirme. Ankara: Anı Yayıncılık.*
- ▶ Güler, Neşe. (2014). *Eğitimde ölçme ve değerlendirme. Ankara: Pegem Akademi Yayıncılık.*