

ÖLÇME ARAÇLARINDA BULUNMASI GEREKEN NİTELİKLER

• 2. SUNUM

- Ölçme aracındaki 3 temel özellik
- Ölçmede hata
- Hata türleri
- Güvenirlik
- Güvenirlik hesaplama yöntemleri
- Ölçmenin standart hatası
- Geçerlik
- Kullanışlılık

Ölçme aracında olması gereken 3 özellik

- ▶ Ölçme araçlarından elde edilen bilgilerin sağlıklı bir şekilde amacına hizmet edebilmesi için 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 nesnelere 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'lik bir cetvel ile sınıftakilerin boyunun ölçülmesi (her seferinde 1 kez kullanılacak)
- ▶ Süresi az verilen 5 soruluk sınav

Sistemik Hata

- ▶ Belli bir kurala göre ortaya çıkan hatalara sistemik hatalar denir.
- ▶ Hatanın kurallı (sistemli) olması.
- ▶ Burada hatanın miktarı ölçülen büyüklüğe, ölçmeceye veya ölçme koşullarına ve belli bir duruma bağlı 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.
- ▶ Sistemik 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

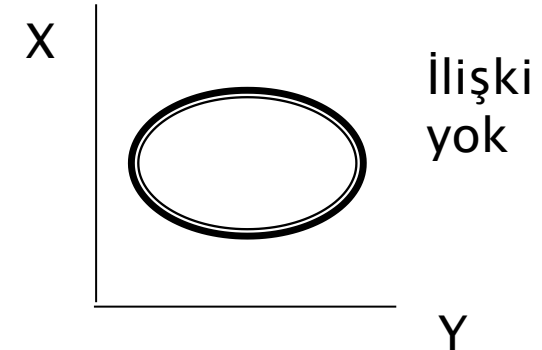
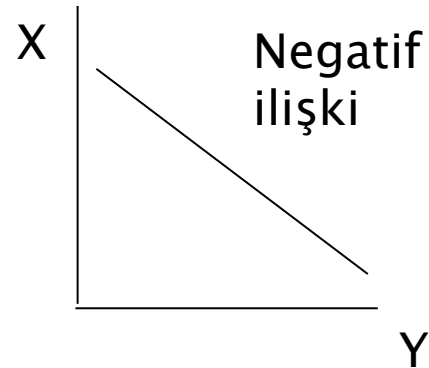
- ▶ Korelasyon (ilişki), iki değişkenin birlikte değişiminin bir ölçüsüdür.
- ▶ Boy uzunluğu ile kilo arasındaki ilişki, yaş ile boy arasındaki ilişki, çalışma saati ve sınav puanı arasındaki ilişki, hava sıcaklığı ve doğalgaz tüketimi arasındaki ilişki eğer sayısal veriler varsa korelasyon katsayısı cinsinden gösterilebilir.
- ▶ Korelasyon katsayısı matematiksel olarak -1 ile $+1$ arasında değerler alır.
- ▶ Korelasyonun **büüklüğü** ($0-1$) 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.
- ▶ Örneğin eğitimde kullandığımız testlerin güvenilirliğinin en az 0,70 olması gerekir.
- ▶ **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 güveniliridir.
- ▶ 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 güvenirlilik hesaplama yöntemleri de mevcuttur.

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

Test–tekrar test

- ▶ Aynı testin aynı gruba farklı zamanlarda iki kez uygulanmasına dayanır.
- ▶ Eğer test güvenilirse aynı şeyi 2 kez ölçtüğünde aynı sonucu vermelidir.
- ▶ Bu durumda güvenilirlik katsayısı, aynı bireylerin her iki uygulamadan aldıkları puan dağılımları arasındaki korelasyon katsayısıdır.
- ▶ Test–tekrar test güvenilirliği ne kadar yüksekse, test puanları, test dışındaki çevresel ve bireysel değişimlere o derece az duyarlı olacaktır.
- ▶ İki farklı zamandaki ölçmelerden farklı sonuçlar elde edilmesinin iki sebebi vardır birincisi ölçme aracı ikincisi ölçülen bireye ait değişiklikler.
- ▶ 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.
- ▶ İki testten elde edilen puanlar *sürekli değişkendir* ve 2 puan dağılımı arası korelasyon hesaplanır. Bu korelasyona Pearson Momentler Çarpımı Korelasyon Katsayısı (PMÇKK) denir. -1 ve 1 arası değerler alır fakat güvenilirlik negatif olamayacağından dolayı test–tekrar test güvenilirliği [0,1] değerler alır.
- ▶ Psikolojide, zeka, kişilik, stres, kaygı, mutluluk vb. kavramların ölçümünde kullanılır başarı testlerinde kullanılması tavsiye edilmez.

Kişi	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

$$r=0,87$$

Eş Değer (Paralel) Formlar

- ▶ Eşdeğer/paralel formlar güvenilirliği aynı öğrenci grubuna birbirinin paraleli olan iki testin uygulanmasına dayalıdır.
- ▶ Ağırlığın 2 kez aynı terazi ile ölçülmesi vs. 2 farklı terazi ile 1 er kez ölçümü
- ▶ Aynı davranışları ölçen fakat farklı sorulardan oluşan iki ayrı test hazırlanarak uygulanır
- ▶ İki testten elde edilen puanlar arasındaki korelasyon katsayısı hesaplanır.
- ▶ Bulunan güvenilirlik 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 güvenilirlik 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 bütününe için güvenilirliği bulmak için Spearman–Brown formülü ile düzeltilmesi gerekir.

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

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

Eşdeğer Yarılar

- ▶ Diyelim ki eş değer yarılar yöntemi ile testin yarısının güvenilirliği 0,6 olsun. Bu durumda testin bütününün güvenilirliği $= 2 \times 0,6 / (1 + 0,6)$ 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 yarılarından aldıkları puanlar belirlenir.
 - 4) Öğrencinin yarılarından 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 güvenilirliđi testteki soruların birbirleriyle uyumunu, tutarlılıđını gösteren güvenilirlik hesaplama yöntemleridir.
- ▶ Testin bir kere uygulanması yeterlidir.
- ▶ KR-20 hesaplanırken testteki sorular 1 ve 0 (dođru-yanlıř) řeklinde 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 deđer almakla birlikte aynı anlamı vermektedir.
- ▶ İkisi de iç tutarlılıđı belirlerler. 1'e yakın olduđu ölçüde test güvenilirlidir.

Cronbach Alfa

- ▶ KR20 ve KR21 gibi Alfa da tek uygulamaya dayanır.
- ▶ KR-20 ve KR-21 güvenilirlik katsayıları testteki soruların 1-0 şeklinde kodlanabildiği durumlarda (başarı testi) kullanılır.
- ▶ 1 ve 0 şeklinde puanlanamayan ve doğru cevabı olmayan duyuşsal özelliklerin ölçülmesinde (tutum, ilgi, kişilik vb.) güvenilirliği Cronbach Alfa katsayısı belirler.
- ▶ Cronbach Alfa ile hesaplanan güvenilirlik katsayısı da iç tutarlılık anlamındadır.
- ▶ Alfa güvenilirlik katsayısının yüksek olması güvenilirliği artırır.

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, puanlayıcıların birbirleri ile tutarlı puanlama yapıp yapmadıklarını belirlemek için kullanılan tutarlık anlamında bir güvenirlik türüdür.

Kappa istatistiği kullanılır.

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 : Testin güvenirliği

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

Örn: Güvenirliği (r) 0,64 olan bir testin 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 $G = \bar{O} \pm S_e$ 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 $77-1*3$ ve $77+1*3$ yani **74** ve **80** arasında
- ▶ %95 olasılıkla $77-2*3$ ve $77+2*3$ yani **71** ve **83** arasında
- ▶ %99 olasılıkla $77-3*3$ ve $77+3*3$ yani **68** ve **86** arasında yer alır.

Güvenirliđi 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 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

2. Geerlik

- ▶ Ölme aracının ölçmek istediđi özelliđi ölçebilme derecesidir.
- ▶ Bir testin ölçmek istediđi özelliđi başka özelliklerle karıştırmadan (veya daha az karıştıırarak) ölçebilmesi geerli olduđu anlamına gelir.
- ▶ Yani bir ölçme aracı amacına hizmet ettiđinde geerlidir.
- ▶ Ölme aracının amacına uygunluđu olarak da tanımlanır.
- ▶ Yani bir testle ne amaçlanıyorsa onun gerçekleştirilmesi ölçüsünde geerli olmasından bahsedilir.
- ▶ Geerlik $[-1, 1]$ arasında deđerler alır.

Geçerlik Türleri



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

Görünüş Geçerliđi

- ▶ Bir testin neyi ölçüyor göründüđü o testin *görünüş geçerliđidir*.
 - ▶ Ya da bir testin ölçmek istediđi 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 yazılı olması
- görünüş geçerliđi sağlandığının işareti olabilir.
- ▶ **Görünüş geçerliđi teknik anlamda bir geçerlik türü olmayıp, testin ölçmek istediđi özelliđi ölçüyor görünmesi olarak algılanabilir. Fakat bir geçerlik yöntemi olarak sayan görüşler de mevcuttur.**

Görünüş Geçerliđi

- ▶ İlköğretim düzeyinde geliştirilmiş bir test, yetişkinler üzerinde deneme uygulaması yapılırsa test 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 Geçerliđi

- ▶ Eđitimde çođunlukla testler öđrencilerin belli bir alana ya da derse iliřkin davranıřları (bilgileri) edinip edinmediđini anlamak için uygulanır.
- ▶ Bu durumda testi geliřtiren kiři, testin, bu davranıř ya da bilgileri yeteri kadar iđerip iđermediđini arařtırması gerekir.
- ▶ Testin amaçlarının testte bulunan sorularla uyumlu olup olmadıđı incelenmelidir.
- ▶ Kapsam geçerliđi bir testin bu testle ölçölmek istenen davranıřları *ne derece kapsadıđı* ile ilgilidir.
- ▶ Bir testin kapsam geçerliđi, ölçmek istediđi alan kapsamını *ne kadar temsil ettiđi* ile alakalıdır.
- ▶ Kısaca bir testin ölçmek istediđi konudaki hedef–davranıřları *ne kadar iyi örnekleđiđi* kapsam geçerliđidir.

Kapsam Geçerliđi Nasıl Test edilir?

- ▶ Kapsam geçerliđi belirlenirken en çok *uzman görüŖüne* başvurulur. (örn:uzmanların bir araya geldiđi bir panel yapmak)
- ▶ Örnek: Bir matematik testinin kapsam geçerliđi 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 bulunduđu bir tablodur. Bu tablo sayesinde ölçme kapsamındaki tüm davranışların teste soru olarak alınması sağlanır.
- ▶ Aynı kapsamı ölçen bir testle karşılaştırmak. Bir testi ölçtüđu bilinen bir testle aynı kapsamı ölçtüđu düşünölen testler arasındaki korelasyona dayanır.

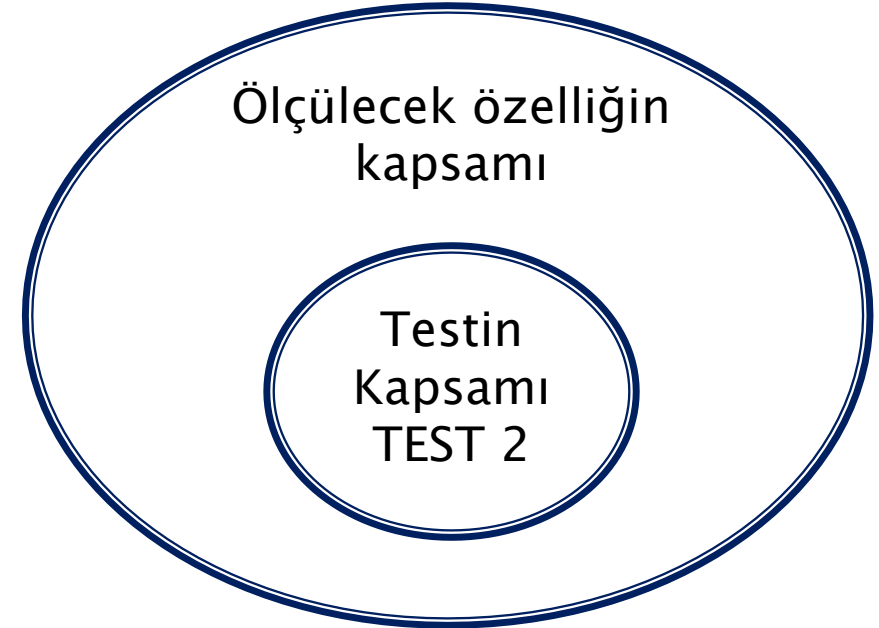
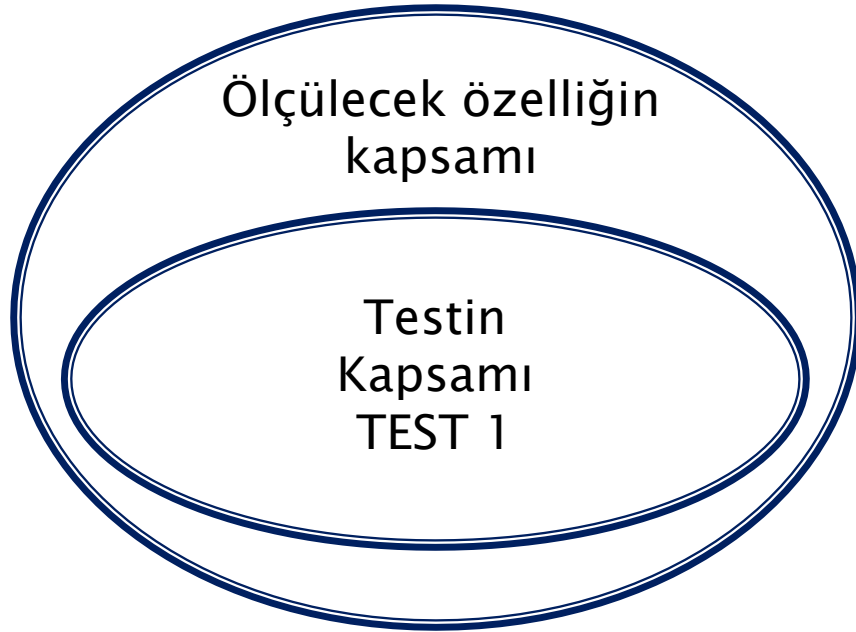
Belirtke Tablosu Örneđi

Bir ölçme deęerlendirme dersine ait ara sınav sorularının daęılımı ařađıdaki gibi olabilir:

Konular	Bilgi	Kavrama	Uygulama	Analiz	Sentez	Deęerlendirme	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 geçerliđinin Őematik olarak gösterilmesi

Hangi testin kapsam geçerliliđi fazladır?



Kaynak: Nükhet Çıkırıđı DemirtaŐlı (2012). *Eđitimde Ölçme ve Deđerlendirme*. 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 puanlar 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) Geçerliđi

- ▶ Ölçütten elde edilen puanlar, ölçme aracından elde edilen puanlarla aynı zamanda veya daha önceden elde edilmişse, bu puanlar arasındaki korelasyona dayalı geçerliğe uygunluk geçerliđi adı verilir.
- ▶ Örneđin YGS'nin geçerliđini belirlemek istersek ve bu amaçla öğrencilerin lise notlarını kullanırsak geçerlik uygunluk geçerliđi olur. Burada geçerliđi sağlanan puanlar YGS puanları ölçüt puanlar ise lise notlarıdır.
- ▶ 2 puan dağılımı arası korelasyona bakılır.
- ▶ Zamandaş, eş zamanlı, halihazır, hemhal kavramları uygunluk geçerliđi için kullanılan diğer adlandırmalardır.

Yordama Geçerliđi

- ▶ Yordama bir tahmindir ve geleceđe yöneliktir!
- ▶ 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 Geçerliđi

- ▶ Yapı; birbiriyle ilişkili olan ögelerin oluşturduđu homojen örtüye denir. Doğrudan gözlenemeyen (ör: zeka) psikolojik özellik anlamına da gelmektedir.
- ▶ Yapı geçerliđi ölçme aracının teoride var olduđu düşünölen, gözle görölemeyen bu yapıları pratikte ölçebilme derecesidir (ör:Matematik).
- ▶ Bir testin yapı geçerliđinin olabilmesi için sorularının birbiriyle yüksek ilişkili olması lazım.
- ▶ Bu da soruların birbiriyle benzer özellikleri ölçtüđu yani aynı yapıyı temsil ettiđi anlamına gelir.
- ▶ Mesela bir tarih testinin tüm sorularının tarih yapısını ölçmesi gibi. Eğer içinde matematik ölçen bir soru var ise yapı geçerliđi bozulur.

Yapı Geçerliđini 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

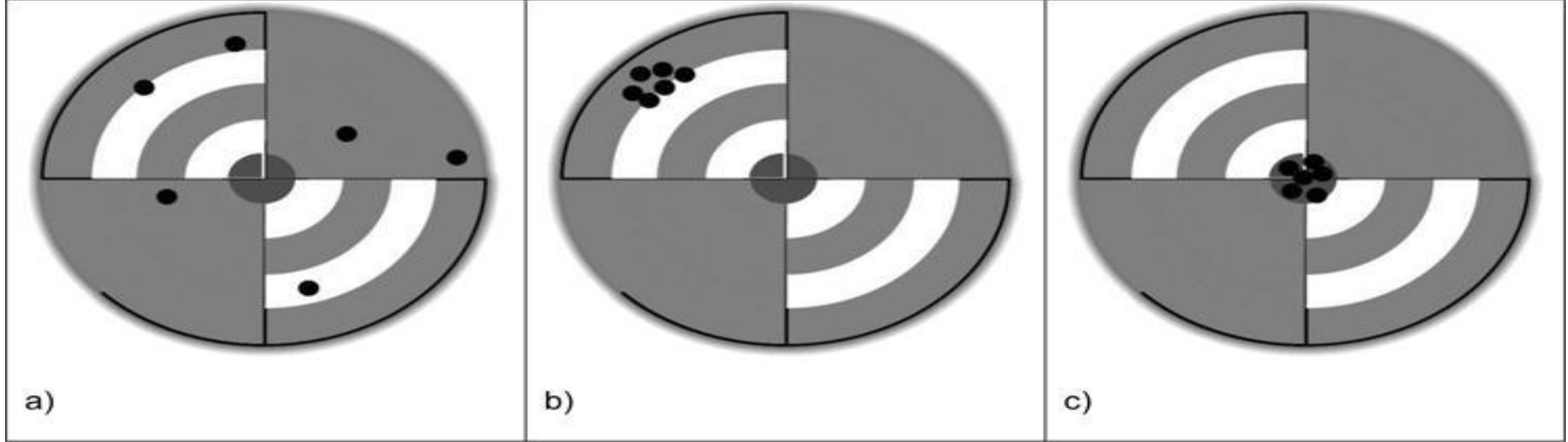
- ▶ Benzer ya da zıt testlerle (ya da daha pratik testlerle) olan korelasyonu bir testin yapı geçerliliği hakkında bilgi verir.
- ▶ Yapı geçerliği, ölçülmek istenen yapının farklı gruplarda incelenmesine (Grup farkları).
- ▶ Örneğin, sınav kaygısını ölçmek üzere geliştirilmiş bir ölçme aracı, daha önceden 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 ve geçerlik bir ölçme aracında bulunması gereken en önemli özelliklerdendir.
- ▶ Bir teste ait sonuçlar rapor edilirken 2'si bir arada verilir
- ▶ Güvenirlilik [0,1] arası değerler alırken geçerlik [-1,1] arasında değerler alır.
- ▶ Bir ölçme aracının güvenilir olduğu geçerli olduğu anlamına gelmez.
- ▶ Sadece teadüfi hatalardan ekilenen güvenirlilikten farklı olarak, bir testin geçerliği hem tesadüfi hem de sistematik hatalardan etkilenebilir.
- ▶ Bu nedenle bir test geçerli olmasa da güvenilir olabilir.
- ▶ Güvenirlilik geçerlik için *gerekli* fakat *yeterli değildir*.
- ▶ Geçerli olan her test güvenilir denebilir.
- ▶ Geçerli olmayan bir test 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.

Geçerliđi Neler Düşürür?

- ▶ Güvenirliđin düşük olması geçerliđi de düşürür.
- ▶ Sabit ve sistematik hatalar
- ▶ Her türden yanlılık
- ▶ Sınav öncesi öğrencilerin soruları görmesi
- ▶ Yazı güzelliđi, kađıt 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 geçerliđi 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 kolaylıklarınınin tümü
- Kullanışlılık özelliğini vurgular.

Kaynakça

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