

ÖLÇME ARAÇLARINDA BULUNMASI GEREKEN

NİTELİKLER

Güvenirlilik

Geçerlik

Kullanışlılık

GÜVENİRLİK

- Güvenirlik, ölçme aracının tesadüfi hatalardan arınık olma derecesidir.
- Güvenilir bir ölçme aracı, aynı özellekle ilgili olarak arka arkaya yapılan ölçmelerde yaklaşık olarak aynı sayısal sonucu verir.

Güvenilir bir ölçme aracıyla;

DUYARLI,

TUTARLI ve

İSTİKRARLI

ölçmeler yapılabilir.

Güvenirlik katsayısı (0.00) ile (1.00) arasında değer alır.

HATA KAYNAKLARI

- $X=T+e$
- X: Gözlenen herhangi bir test puanı
- T: Gerçek puan
- e: Pozitif ve ya negatif bir hata puanı

$$e = T - X$$

HATA KAYNAKLARI

Ölçme işlemine ve sonuçlarına karışan hatalar:

1. Ölçmeyi yapan kişiden kaynaklanan hatalar
2. Ölçme aracından kaynaklanan hatalar
3. Ölçülen özellikten - deęişkenden kaynaklanan hatalar
4. Ölçmenin yapıldığı ortamdan kaynaklanan hatalar
5. Ölçme Yönteminden kaynaklanan hatalar

HATA TÜRLERİ

- Sabit Hatalar
- Sistemli Hatalar
- Rastgele Hatalar

Sabit Hata:

Bir ölçmeden diğerine miktarı değişmeyen hatalara sabit hata denir.

Ölçme aracından kaynaklanan ve miktarı bilinen hata türüdür.

Bu hata türü, ölçme aracının bir ölçmede bir defa kullanılmasıyla yapılan hatadır.

ÖRNEK, ucu 2 cm kırık olan bir cetvelle bir kitabın eni ölçülmüş ve 20 cm gelmiş olsun. Bu ölçmede 2 cm'lik bir hata yapılmıştır, yani 18 cm olduğu halde 20 cm denilmiştir. Cetveldeki hata bilindiğinden 20 cm olan ölçme sonucundan 2 cm çıkarılarak hata düzeltilebilir.

Sabit Hata:

ÖRNEK, Yazılı sınav puanlaması sırasında yazısı güzel olanların tümüne fazladan 10 puan eklemek.

ÖRNEK, Puanlama anahtarında doğru yanıt olmayan seçeneđi doğru imiş gibi puanlama

- **Sistemetik Hata:**

Yapılan hata miktarı herkes için aynı olmayıp,
farklılık göstermektedir.

Sistematik hata yanlılık olarak da bilinir.

Sistematik hata; ölçülen büyüklüğe,
ölçmeciye ve ölçme koşullarına bağlı olarak miktarı değişen
hatadır.

*Bir terazinin bazı ağırlıkları 3 kg fazla tartarken bazılarını 2 kg az
tartması gibi.*

**Sabit hata veren bir ölçme aracının tekrar tekrar kullanımı da
sistematik hatayı doğurur.**

- **Sistemetik Hata:**

ÖRNEK, 20 cm olarak hazırlanan bir cetvel aslında 19 cm ise bu cetvelin 20 cm olarak ölçtüğü bir uzunluk için hata 1 cm iken 40 cm olarak ölçtüğü bir uzunluk için hata 2 cm olur.

ÖRNEK, Bir öğretmenin yazılı yoklama sınavlarını puanlarken yazı güzelliği, ifadelerin akıcılığı veya yazının okunaksızlığı gibi sebeplerden dolayı bazı öğrencilere fazla bazılarına az puan vermesi sistematik hataya örnektir.

ÖRNEK, Sınava girilen yerde trafik gürültüsüne bağlı olarak öğrencilerin farklı etkilenmeleri, alabileceklerinden farklı puanlar almasına neden olur. Birinin 5, diğerinin 10 ve bir diğerinin de 2 puan az almasına neden olabilir.

ÖRNEK, Öğretmenin puanlamayı gelişigüzel yapması

ÖRNEK, Puanlamanın farklı zamanlarda ve farklı psikolojik atmosferde yapılması.

Tesadüfi (Rasgele, Random) Hata:

- Kaynağı kesin olarak bilinmeyen,
miktarı kesin olarak tayin edilemeyen,
ancak istatistiksel yöntemlerle ve
belirli olasılıklarla kestirilebilen,
her ölçme sonucunda az veya çok mutlaka bulunan bir
hata türüdür.
- Güvenirlik kavramının tanımında geçen hatasızlık sözcüğü ile
random hata kastedilmektedir.
- **Güvenirliğin, ölçme sonuçlarınınin random hatalardan
arınıklık derecesi olarak tanımlanması
gelenekselleşmiştir**

Güvenirlik Kestirme Yöntemleri

Güvenirlik katsayısı; gerçek ölçümlerin varyansının, gözlenen puanların (gerçek ölçüm+hata) varyansına oranıdır.

- * TEST TEKRAR TEST YÖNTEMİ
- * PARALEL TEST (EŞDEĞER FORMLAR) YÖNTEMİ
- * TESTİ YARILAMA (EŞDEĞER YARILAR) YÖNTEMİ
- * KR 20 ve KR 21 FORMÜLLERİ
- * CRONBACH ALFA

Test Tekrar Test Yöntemi

1. Aynı test formunun, aralıklı olarak iki ayrı zamanda aynı gruba uygulanması ve iki uygulamadan elde edilen puanlar arasındaki korelasyonun bulunması (istikrarlılık, kararlılık, tutarlılık katsayısı).

Bir test aynı gruba belli bir zaman aralığıyla iki kez uygulanır.

Bireylerin birinci uygulamada aldıkları puanlarla ikinci uygulamada aldıkları puanlar arasındaki korelasyon bulunur.

Elde edilen korelasyon katsayısı testin güvenilirlik katsayısıdır.

Güvenirlik katsayıları hemen hemen her zaman 0.00 ile +1.00 arasında değişmektedir.

Test Tekrar Test Yöntemi

25 Eylül

Test A

1. asfiokama
2. dhjdahjsca
3. daksaks
4. saeaha

Puan= 82

15 Ekim

Test A

1. asfiokama
2. dhjdahjscla
3. daksaks
4. saeaha

Puan= 86

Şekil 8.1: Test Tekrar Test Yönteminin Şematik Gösterimi

Test Tekrar Test Yönteminin Hesaplanması

ÖĞRENCİ	I. UYGULAMA (X)	II. UYGULAMA (Y)	X^2	Y^2	XY
1	15	17	225	289	255
2	16	15	256	225	240
3	17	17	289	289	289
4	10	11	100	121	110
5	14	14	196	196	196
6	6	5	36	25	30
7	17	17	289	289	289
8	12	11	144	121	132
9	3	4	9	16	12
10	8	9	64	81	72
11	9	9	81	81	81
12	14	14	196	196	196
13	15	13	225	169	195
14	12	13	144	169	156
15	15	15	225	225	225
N=15	$\sum X = 183$	$\sum Y = 184$	$\sum X^2 = 2479$	$\sum Y^2 = 2492$	$\sum XY = 2478$

$$r_{xy} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{N} \right] \left[\sum Y^2 - \frac{(\sum Y)^2}{N} \right]}}$$

$$r_{xy} = \frac{2478 - \frac{183 \cdot 184}{15}}{\sqrt{\left[2479 - \frac{33489}{15} \right] \left[2492 - \frac{33856}{15} \right]}} = 0.95$$

EŐDEĐER FORMLAR / PARALEL TESTLER

- . İki eşdeđer testin ya da bir testin iki eşdeđer formunun aynı gruba aynı zamanda ya da belli bir zaman aralığıyla uygulanması ve iki eşdeđer testten elde edilen puanlar arasındaki korelasyonun bulunması (Eşdeđerlik katsayısı).
- Paralel testlerin her ikisi de aynı öğrenci grubuna uygulanır.
- Elde edilen puanlara ilişkin istatistiksel işlemler ve elde edilecek olan korelasyon katsayısının yorumu test tekrar test yöntemindeki aynısıdır.

EŞDEĞER YARILAMA

3. Bir gruba uygulanmış bir testin iki eşdeğer yarıya bölünmesi ve öğrencilerin iki eşdeğer yarıdan aldıkları puanlar arasındaki korelasyonun bulunması (İç tutarlılık katsayısı).

Uygulanmış olan test iki eşdeğer yarıya bölünerek öğrencilerin testin iki yarısından aldıkları puanlar arasındaki korelasyon hesaplanır ve daha sonra bu hesaplanan korelasyondan hareketle **Spearman–Brown** formülünden de yararlanarak testin bütünüün güvenilirliği kestirilir.

Bu yöntem, aslında **iç tutarlılık** katsayısı verir.

Spearman-Brown formülü

$$r = \frac{2 \cdot r}{1 + r}$$

Kuder-Richardson 20 ve 21 Formülleri

Testteki her bir maddeye doğru cevap veren öğrenci yüzdesi hesaplanmışsa testin güvenilirliği, Kuder-Richardson formülleriyle tahmin edilebilir.

Bu yolla hesaplanan güvenilirlik katsayısı da testin iç tutarlılığının bir ölçüsüdür.

Kuder-Richardson formülleri, testteki her maddenin aynı değişkeni ölçtüğü, yani testin ölçtüğü şeyin homojen olduğu sayıltısına dayanır.

Testin iki yarısı yerine testteki tüm maddeler arasındaki tutarlılığın ölçümünü verir ve **iç tutarlılık** katsayısı olarak adlandırılır.

Kuder-Richardson 20 ve 21 Formülleri

KR-21 madde analizi yapılmamış testlere uygulanır; ve testteki maddelerin madde güçlükleri eşit farzedilir.

Bu nedenle KR-21 formülü ile elde edilen katsayı güvenilirliğin alt sınırı olarak kabul edilir.

Bir başka deyişle KR-21 her zaman için KR-20'den küçüktür. (**KR-20 > KR-21**).

Bir testin KR-20 veya KR-21 güvenilirliği yüksek bulunmuşsa, bu testteki maddelerin aynı yeterliği ölçtüğü (testin tek boyutlu olduğu) ve test puanlarının tesadüfi hatalardan arınık olduğu sonucuna ulaşılabilir.

Kuder-Richardson 20 ve 21 Formülleri

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

K = Testin soru sayısı

p = Madde güçlüğü

$q = 1 - p$

S_x^2 = Testin varyansı

$$KR_{21} = \frac{K}{K-1} \left[1 - \frac{K\bar{X} - \bar{X}^2}{KS_x^2} \right]$$

K = Testin soru sayısı

p = Madde güçlüğü

S_x^2 = Testin varyansı

SORU: Ortalaması 35, standart kayması 7 olan 60 maddelik bir testin KR-21 güvenirlik katsayısını hesaplayınız.

$$KR_{21} = \frac{60}{60-1} \left[1 - \frac{60*35 - 35^2}{60*7^2} \right] =$$

Cronbach Alpha Güvenirlik Katsayısı

Alpha güvenirlik katsayısı tek bir uygulama gerektiren güvenirlik bulma tekniklerinden biridir.

Ağırlıklı puanlama veya dereceleme yöntemiyle puanlama uygulandığı durumlarda kullanılabilecek bir güvenirlik hesaplama tekniğidir.

Testi oluşturan maddelerin dereceleme ölçeğine göre puanlanması (1'den 5'e kadar veya 0'dan 4'e kadar gibi) genellikle psikomotor becerileri ölçmek amacıyla hazırlanan ölçme araçlarının puanlanmasında veya tutum amacıyla hazırlanan ölçme araçlarının puanlanmasında veya tutum ölçeklerinin puanlanmasında kullanılmakla birlikte kısa cevaplı testlerin güvenirliğinin kestirilmesinde de kullanılabilir.

Cronbach Alpha Güvenirlik Katsayısı

$$\alpha = \frac{K}{K-1} \left[1 - \frac{\sum s_j^2}{S_x^2} \right]$$

K = Testin soru sayısı

S_x^2 = Testin varyansı

s_j^2 = j maddesinin varyansı

GÜVENİRLİĞİ ARTIRMA YOLLARI

1. Bir sınavda kullanılan soru sayısı arttıkça, çoğu halde, o sınavdan elde edilen toplam puanın güvenilirliği de artar.
2. Bir sınavdan kullanılacak soruların açıkça anlaşılır ve kesinlikle cevaplanabilir olması, o sınavdan elde edilen puanın güvenilirliğini artırır.
3. Her sınav objektif yollarla puanlanabilmelidir. Cevap anahtarı hazırlama puanlamadaki objektifliği artırır.
4. Öğrencilerin sınava olan güdeleri ölçme sonuçlarının güvenilirliğini etkiler. Öğrencilerin sınava yeterli derecede güdülenmediği hallerde rastgele cevap verme durumları artacağı için sınavın güvenilirliği düşer.
5. Sınav süresi dengeli olmalıdır. Sınav süresinin gereğinden az veya çok verilmesi güvenilirliği etkiler.

GÜVENİRLİĞİ ARTIRMA YOLLARI

6. Benzeşik konuları yoklayan testler, birbiriyle ilişkisiz konuları yoklayanlara göre daha güvenilirdir. (*Testin Kapsamı*)
7. Test maddeleri ne çok zor ne de kolay olmalıdır.
8. Madde Ayırdedicilik gücü yüksek maddeler olmalıdır.
9. Öğrencilerin sınav deneyimleri de ölçme sonuçlarının güvenilirliğini etkiler.

ÖLÇMENİN STANDART HATASI

Ölçmede bireyin gözlemlenen puanları elde edilir. Gerçek puanlar ve gözlenen puanla ilgili ölçme hatası bilinemez.

Standart hata, standart kayma ve güvenirlik katsayısından yararlanılarak hesaplanır.

Sem: Ölçmenin standart hatası

s: Standart kayma

r : Güvenirlik katsayısı

Ölçmenin Standart Hatası (ÖSH)

$$SEM = sd \times \sqrt{(1 - r)}$$

SEM: Ölçmenin standart hatası

sd: Test puanlarının standart sapması

R: Güvenirlilik katsayısı

Ölçmenin standart hatası;

gerçek değerlerin güven aralığını belirlemede kullanılır:

ÖLÇMENİN STANDART HATASI

Hatasız ölçme olmaz.

Bu nedenle ölçme sonuçları ölçek üzerinde tek bir noktadan ziyade iki nokta arasındaki bir aralık olarak düşünölmelidir. Güven aralıkları standart hatanın eklenmesi ve çıkarılması ile belirlenir.

Testlerde güven aralıkları,

1 standart hata için %68

2 standart hata için %95

3 standart hata için %99 olarak belirlenmiştir.

Bir test puanınının güven aralığını %68 olasılıkla belirlemek için;

Üst sınır = Gözlenen puan + 1 Sem

Alt sınır = Gözlenen puan – 1 Sem arasındadır.

ÖLÇMENİN STANDART HATASI

%95 olasılıkla öğrencinin puanına 2 Sem,
%99 olasılıkla da 3 Sem eklenir ve çıkarılır.

Örnek: Standart sapması 6, güvenilirlik katsayısı 0,75 olan bir testten 50 alan bir öğrencinin puanına % 68, % 95 ve % 99 olasılıkla hata karışma oranı düşünüldüğünde puan aralıkları nelerdir?

$$SEM = sd \times \sqrt{(1 - r)}$$

% 68 olasılıkla:

$$SEM = 6 \times \sqrt{(1 - .75)} = 3$$

$$50 + 3 = 53$$

$50 - 3 = 47$ Puan aralığı 47-53 arasındadır.

ÖLÇMENİN STANDART HATASI

%95 olasılıkla;

$$50 + 2.3 = 56$$

$50 - 2.3 = 44$ Puan aralığı 44 – 56 arasındadır.

%99 olasılıkla;

$$50 + 3.3 = 59$$

$50 - 3.3 = 41$ Puan aralığı 41 – 59 arasındadır.

UYARI: Standart hata azaldıkça güvenilirlik artar.

Yani güvenilirlik 1,00 ise standart hata 0,00'dır.

ÖLÇMENİN STANDART HATASI

UYARI: Standart hata azaldıkça güvenilirlik artar.

Yani güvenilirlik 1,00 ise standart hata 0,00'dır.

UYARI: Güvenirliğin yüksek olması ve standart sapmanın küçük olması ölçme işleminin standart hatasının küçük olduğunu gösterir.

GEÇERLİK

Bir ölçme aracının, ölçülecek özellik(leri), ölçülmesi sözkonusu olmayan başka özelliklerle karıştırmadan, tam ve doğru olarak ölçebilmesi; amaca hizmet etme derecesidir.

Ölçme aracında bulunması gereken en önemli özelliktir.

Geçerlik katsayısı (- 1.00) ile (+ 1.00) arasında değerler alır.

Geçerliğin iki önemli ögesi vardır:

Uygunluk / İlgililik - Güvenirlilik

GEÇERLİK

* **Kapsam Geçerliđi**

* **Ölçüt Geçerliđi**

Yordama Geçerliđi

Uygunluk Geçerliđi

* **Yapı Geçerliđi**

* **Görünüş Geçerliđi**

KAPSAM GEÇERLİĞİ

- Bir bütün olarak testin ve testteki her bir maddenin amaca ne derece hizmet ettiği.
- Bir testin kapsam geçerliği; o testteki toplam maddelerin ölçülecek davranışları ve konu içeriğini örnekleme derecesine ve testteki her bir maddenin ölçmek istediği davranışı ne derece iyi ölçtüğüne bağlıdır.
- Bir testin kapsam geçerliği 2 ana yaklaşımla belirlenebilir :
 - * *Mantıksal ya da rasyonel yaklaşım*
 - * *İstatistiksel yaklaşım*

ÖLÇÜT GEÇERLİĞİ

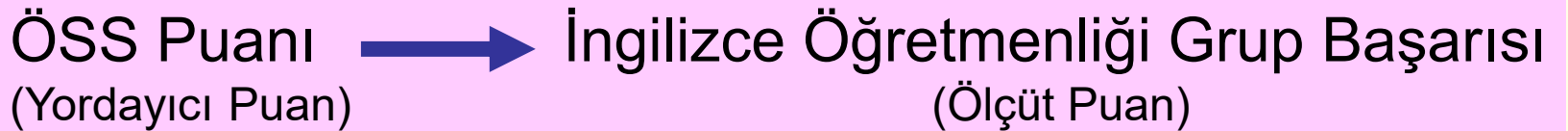
- Bireylerin bir işi gerçekleştirip gerçekleştirilemeyeceğine, ölçülen özelliikle ilgili gelecekteki başarısını belirlemeye veya bireyin bugünkü başarısının geçmişteki durumuyla ilişkisini belirlemeye yönelik olabilir.
- Ölçme aracından elde edilen puan, ölçüt durumundaki puanlarla karşılaştırılır.
 - * *Yordama Geçerliđi*
 - * *Uygunluk Geçerliđi*

YORDAMA GEÇERLİĞİ

- Yordama; istatistiksel teknikler kullanılarak ve bilinenlerden yararlanılarak bilinmeyen durumlar hakkında yapılan geleceğe yönelik tahminlerde bulunma işlemidir.
- Eğitimsel anlamda yordama; genellikle bir testten alınan puanlara dayanılarak, bireylerin belli bir programdaki ya da işteki başarılarını önceden kestirme işlemidir.

YORDAMA GEÇERLİĞİ

- Bir testin yordama geçerliği; o testten elde edilen puanlarla, testin yordamak için düzenlendiği değişkenin doğrudan ölçüsü olan ve daha sonradan elde edilen ölçüt arasındaki korelasyondur.
- Buradaki korelasyon katsayısı yordama geçerliği katsayısı olarak adlandırılır ve ne kadar büyükse o oranda yüksek yordama geçerliğine işaret eder.



* İki başarı arasındaki ilişki yüksekse,
ÖSS'nin yordama geçerliği yüksektir.

UYGULAMA GEÇERLİĞİ

- Bugün bilinenlerin geçmişteki bir duruma uygunluğunu bulmaya dayalıdır.

İngilizce Öğretmenliği Grup Başarısı
(Ölçüt Puan)



ÖSS Puanı
(Yordayıcı Puan)

İki başarı arasındaki ilişki yüksekse,
ÖSS'nin yordama geçerliği yüksektir.

YAPI GEÇERLİĞİ

- Yapı, birbiriyle ilgili olduğu düşünülen belli öğelerin ya da öğeler arasındaki ilişkilerin oluşturduğu bir örüntüdür.
- Bu anlamda, bir testin yapısını geçerleme süreci; temelde, testin maddelerine verilen cevaplar arasındaki ilişkilerin analizine dayanır.

GÖRÜNÜŞ GEÇERLİĞİ

- Görünüş geçerliği, bir testin gerçekten ne ölçtüğüyle değil, onun ne ölçüyor görüldüğüyle ilgilidir.
- Bir testin görünüş geçerliği, o testin ölçmek istediği şeyi ölçüyor görünmesidir.

GEÇERLİĞİ ARTIRMA YOLLARI

1. Soru belirsizliđi
2. Madde güçlüđü
3. Madde ayırdedicilik gücü
4. Puanlama yanlılıkları
5. Kopya
6. Aynı soruların kullanımı
7. Güvenilirlik

Güvenirlilik ve Geçerlik Arasındaki İlişki

- **1.** Güvenirlilik, geçerlik için bir ön şarttır. Bir ölçme aracının geçerli olabilmesi için güvenilir olması şarttır (Test güvenilir değilse, geçerliğine bakılamaz). Fakat bir ölçme aracının güvenilir olması için geçerlik şart değildir.
- **2.** Geçerlik, güvenilirliği de kapsayan daha geniş bir kavramdır. Eğer test geçerli ise genellikle güvenilirdir, ama her güvenilir test geçerli değildir.
- **3.** Geçerliğin yüksek olabilmesi için güvenilir olması gerekir. Fakat güvenilirliği yüksek olan bir testin geçerliği düşük olabilir.

Güvenirlilik ve Geçerlik Arasındaki İlişki

- **4.** Geçerlik her türlü hatadan (doğrudan etkilendiği sabit ve sistematik hatadır) etkilenmesine rağmen, güvenirlilik yalnızca tesadüfî hatadan etkilenir.
- **5.** Güvenirliği etkileyen her şey geçerliği de etkiler.
- **6.** Ölçme aracının en önemli özelliği geçerliğidir.

3. KULLANIŞLILIK

- Kullanışlılık; testin geliştirilmesi, çoğaltılması, uygulanması ve puanlanması açısından kolay ve ekonomik olmasıdır.
- Kullanışlılık; bir ölçme aracının kısa zamanda fazla para harcanmadan ve çok yorucu olmadan hazırlanabilmesi ve uygulanabilmesidir.

Kullanışlılık İçin Gereken Özellikler

- **1. Öğrenci, öğretmen ve gelecek açısından ekonomik olmalıdır** (fazla masraf gerektirmemelidir).
- **2. Hazırlanması ve uygulanması kolay olmalı ve kısa sürmelidir.** Bu bize zaman açısından tasarruf sağlar.
- **3. Kolay bir şekilde puanlanması ve puanların yorumlanması** gerekir.
- **4. Uygun bir yönergesinin olması** gerekir.
- **5. Ölçme aracı hazırlanırken uzmanlık gerektirmemesi** gerekir. Yani ölçme işlemini yapan kişiden, test hazırlama sürecinde istenen becerinin az olması gerekir.