

## b) Ölçme Aracının Güvenirliđi

**Güvenirlik;** bir ölçme aracıyla aynı koşullarda tekrarlanan ölçümlerde elde edilen ölçüm değerlerinin kararlılığının (tutarlılığının) bir göstergesidir (Öncü, 1994).

Örneđin; metre ile bir uzunluđun ölçülmesinde, farklı zamanlarda aynı sonuçlar elde edilebiliyorsa, güvenilirliđin yüksek olduđu söylenir. Öğretmen tarafından geliştirilen ve öğrencilerin Matematik becerilerini ölçmeyi amaçlayan bir test, aynı öğrenci grubuna 4 ay arayla uygulandıđında, ilk ve son test puanları birbirine ne kadar yakınsa, ölçme aracının güvenilirliđi de o kadar yüksek olur. Fakat ilk ölçümde sınıftaki öğrencilerin aritmetik ortalaması 75, ikincisinde 95 olursa, geliştirilen testin güvenilirlik düzeyi düşük olur.

Ölçme araçlarının güvenilirlik düzeyi, güvenilirlik analizi ile belirlenir. **Güvenirlik katsayısı “0-sıfır” ile “+1-bir” arasında değer alır.** Güvenirlik katsayısının değeri “+1”e ne kadar yakın ise ölçme aracının güvenilirlik düzeyinin yüksek, “0”a yakın olursa güvenilirlik düzeyinin düşük olduğu ifade edilir.

Yapılan ölçme işlemlerinde tesadüfî hata miktarı ne kadar çok olursa güvenilirlik düzeyi de o oranda düşük olur. Eğer ölçme işleminde **tesadüfî hata** yapılmamış ise güvenilirlik katsayısı “1”e yakın bir değer alacaktır. Fakat eğitim bilimlerinde yapılan ölçümlerde, kontrol altına alınamayan birçok değişken olduğundan, hatasız ölçüm yapmak neredeyse imkânsızdır. Bu nedenle sosyal bilimlerde güvenilirlik katsayısının 1’e çok yakın olması mümkün görülmemektedir.

Örneğin; bir öğretmen yeni geliřtirdiđi akademik başarı testini farklı zamanlarda aynı öğrencilere iki defa uyguladıđında, ilk uygulamadan elde ettiđi puanlar ile ikinci uygulamadan elde edeceđi puan arasında farklılık olma ihtimali oldukça yüksektir. Yani sınıftaki bütün öğrencilerin ilk uygulamada aldıkları puanları ikinci uygulamada da almaları çok zor bir ihtimaldir. Bu gibi nedenlerden dolayı, sosyal bilimlerde 0.65 deđerinden yüksek güvenilirliđe sahip ölçme araçlarının kullanılabilceđi, bilim adamlarınca desteklenmektedir.

Fakat pozitif bilimlerde (fizik, kimya, biyoloji, matematik) kullanılan ölçme araçlarının güvenilirlik düzeyinin 0.99'dan daha büyük olması gerekmektedir. Çünkü fen bilimlerinde kullanılan ölçme araçları oranlı ölçeklerdir ve daha gelişmişlerdir. Ayrıca bu araçlarla mutlak ölçümler yapıldıđından, güvenilirliđinin oldukça yüksek olması gerekmektedir.

Güvenilir ölçme araçlarında, gözlenen ölçümle gerçek ölçüm sonuçları birbirine yakındır. **Ölçme işleminde tesadüfi hata yapılmamışsa güvenilirlik katsayısı 1'e yakın olur.**

Fen bilimlerinde yapılan çalışmalarda güvenilirlik katsayısı yüksek olan ölçme araçları tercih edilirken eğitim bilimlerinde 0,65 değerinden yüksek güvenirlğe sahip olan araçların kullanılması uygundur (Büyüköztürk, 2004).

# Güvenirligi belirlemede iki temel yöntem vardır:

1. Birbirini izleyen ölçmelerde, bireyin grup içindeki **sirasının değişmezliğinin** belirlenmesi;

2. Birbirini izleyen ölçümlerde ölçme hatalarının büyüklüğü, yani kullanılan ölçeğin tekrarlı ölçümlerde aynı denekten yaklaşık olarak aynı ölçüm değerini elde etmesi ve ölçmenin **standart hatasının** düşük olması ile ilgilidir.



# **GÜVENİRLİĞİ BELİRLEME YÖNTEMLERİ:**

## 1. Test-tekrar Test Yöntemi:

◦ Bir ölçme aracının aynı gruba veya bu gruba denk bir başka gruba iki farklı zamanda uygulanmasıdır. Bu uygulamalarda elde edilen puanlar birbirine ne kadar yakın ise ölçme aracının güvenilirliği de o kadar yüksektir.

İki uygulamadan elde edilen ölçüm değerlerinin korelasyon katsayısı, ölçekle elde edilen sonuçların güvenilirlik katsayısını verir.

## **2. Eşdeğer Formlar Yöntemi:**

Paralel formlar yönteminde aynı davranış kalıplarını temsil edebilecek farklı maddeler örneklenerek iki eşdeğer form oluşturulur.

İki formun eşdeğer olabilmesi için, formların kapsamlarının, yapısının, zorluk derecesinin, talimatlarının, puanlama biçiminin, madde sayısının ve yorumlama türünün aynı olması gerekir.



### 3. Testi Yarılama Yöntemi:

Yöntem, formu iki eş parçaya bölerek, iki yarının deneklere aynı anda uygulanması sonrası, deneklerin yarılarından aldıkları puanlar arasındaki korelasyon ile güvenilirlik tahmini yapılmasını sağlar.

Testi iki eşit parçaya ayırmak için şu yollar izlenebilir:

- Öğrencilerin sayısına göre cevap kağıtları iki eşit parçaya bölünür ve bu parçalar karşılaştırılır
- Her öğrencinin tek sayılı sorularıyla çift sayılı sorularının puanları karşılaştırılır
- Testteki soruların ilk yarısı ile son yarısının puanları karşılaştırılır.

## 4. Kuder-Richardson Yöntemi:

Yöntem tüm maddelerin birbirleriyle ve ölçeğin tamamıyla iç tutarlılığını tahmin etme amacı üzerine kuruludur. Bu nedenle yöntem, ölçekteki tüm maddelerin aynı değişkeni ölçtüğü varsayımına dayanır.

Yöntemin uygulanmasında veri dizisi, ölçekteki maddelerden alınan cevaplar istenilen özelliği taşıyorsa “1” puan, istenilen özelliği taşımıyorsa veya boş bırakılmışsa “0” puan verilerek oluşturulur.

Bu yöntemle, iç tutarlılığa yönelik güvenilirlik kestiriminde bulunmada belirli kriterler dikkate alınarak Kuder-Richardson 20 veya 21 formüllerinden uygun olanı kullanılır.

KR-20 ve KR-21 formüllerinin kullanılabilmesi için iki temel özelliğe daha dikkat edilmektedir:

a) Testte yer alan her bir maddenin öğrencilerin en az % 90'ı tarafından doğru veya yanlış cevaplandırılmış olması,

b) Testteki maddelerinin tamamının aynı özelliği ölçüyor olması gereklidir.

## 5. Cronbach Alpha Yöntemi:

Cronbach tarafından geliştirilen ve 1951 yılında yayın olarak sunulan alfa katsayısı yöntemi, maddeler doğru-yanlış olacak şekilde puanlanmadığında, 1-3, 1-4, 1-5 gibi puanlandığında, kullanılması uygun olan bir iç tutarlılık tahmin yöntemidir.

Cronbach alfa katsayısı, ölçekte yer alan  $n$  tane maddenin varyansları toplamının genel varyansa oranlanması ile bulunan bir ağırlıklı standart değişim ortalamasıdır. Bu yöntemde, ölçekteki bütün olası iki kısmın karşılaştırılması yapılmaktadır. Yani testteki maddeler iki eşit parçaya ayrılır ve bütün eşit yarılar karşılaştırılır.

**Tutum ölçeklerini güvenilirliğini belirlemede en fazla kullanılan yöntemdir.**

Bu yöntem, KR-20 ve KR-21'in daha genel biçimidir. Bu yöntemin temel mantığı, ölçme aracındaki maddelerin bütün olası yarılarının karşılaştırılmasına dayanmaktadır.

Örneğin; 20 maddelik bir testte ilk olarak ilk 10 (1-2-3-4-5-6-7-8-9-10) soruyla, sonraki 10 (11-12-13-14-15-16-17-18-19-20) sorunun puanları karşılaştırılır. Daha sonra ilk 9 (1-2-3-4-5-6-7-8-9) soru ve 11. soru ile sonraki 9 (12-13-14-15-16-17-18-19-20) soru ile 10. soruların puanları karşılaştırılır.

# VERİLERİN STANDART HATASI

Bir veri dizisinde puanların her biri arasındaki ortalama farka denir. Başka bir tanıma göre, bir ölçümde yapılabilecek en küçük hatadır. “Se” ile gösterilir.

Örneğin, A grubunun standart hatası 0,5 B şubesindeki 0,7 ise, A grubunda öğrenim gören öğrenciler B şubesindeki öğrencilere göre daha homojen özelliklere sahiptirler.

İki tane temel formülü vardır:

$$a) Se = \frac{S}{\sqrt{N}}$$

$$b) Se = S\sqrt{1-r}$$

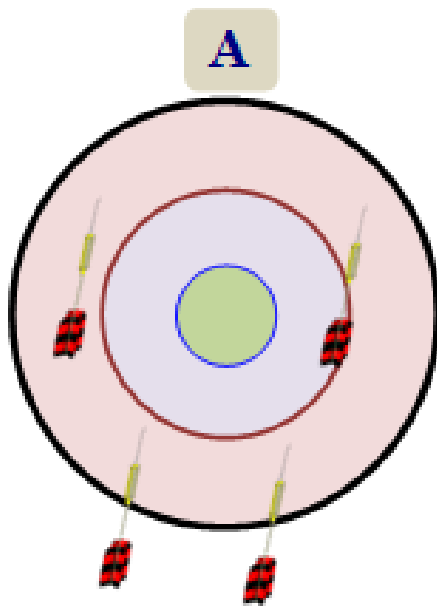
$Se$	=Standart hata
$S$	=Standart sapma
$N$	=Örnekleme sayısı
$r$	=Güvenirlilik katsayısı

Bu formüllerle ilgili olarak aşağıdaki genellemelere gidilebilir:

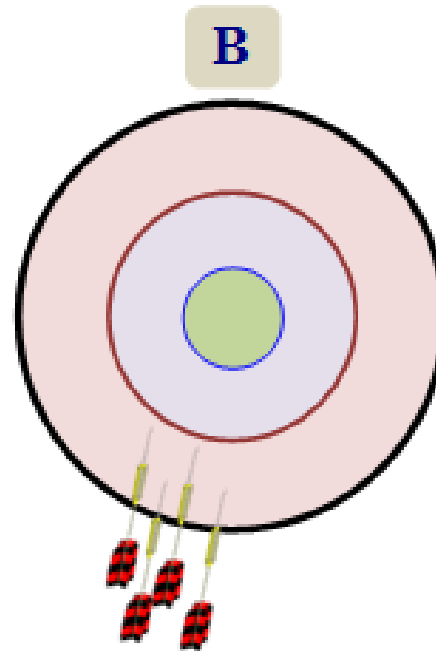
- 1. Standart sapma büyüdükçe standart hata da büyür.**
- 2. Örneklem sayısı arttıkça standart hata küçülür.**
- 3. Güvenirlilik arttıkça standart hata azalır.**

# Ölçme Aracının Geçerliği ve Güvenirliği Arasındaki İlişki

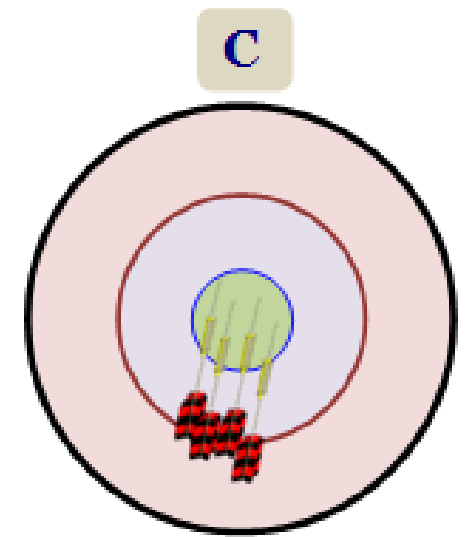
Bir ölçme aracının geçerliği ve güvenirligi arasındaki ilişki aşağıdaki şekilde gösterilebilir:



*A-Geçerliği düşük,  
güvenirligi düşük*



*B-Geçerliği düşük,  
güvenirligi yüksek*



*C-Geçerliği ve  
güvenirligi yüksek*



Bu bilgilere göre ařağıdaki genellemeleri çıkarmak mümkündür:

- Güvenirliğı düşük ölçme sonuçlarının geçerliğı de düşük olur (**A řekli**).
- Güvenirliğı yüksek ölçme sonuçlarının geçerliğı yüksek olmayabilir (**B řekli**).
- Geçerliğı yüksek ölçme araçlarının güvenirliğı de yüksek olur (**C řekli**).

# GEÇERLİK VE GÜVENİRLİĞİ ETKİLEYEN FAKTÖRLER

- *Test türü*
- *Soru sayısı*
- *Testin düzeyi*
- *Cevaplama süresi*
- *Testin puanlanması*
- *Testin anlaşılır olması*
- *Ortam*

## PUANLAYICI GÜVENİRLİĞİ

Güvenilir kişiler, olaylara “*nesnel*”, “*objektif*” ve “*tarafsız*” bakabilen kişilerdir. Zaten ölçme işleminde, sonuçları hiç değiştirmeden, olduğu gibi sunmak ölçmecinin temel görevidir.

Ölçmecinin güvenilirliğini belirlemenin bazı yöntemleri vardır. Bunlardan birincisi; elde edilen verilerin aynı kişi tarafından farklı zamanlarda tekrar puanlanmasıdır.

İkinci yöntem ise, elde edilen verilerin farklı kişiler (en az iki kişi) tarafından puanlanmasıdır. Bu kişilerin puanları birbiri ile tutarlı ise, ölçme işleminde objektif, yani güvenilir davranıldığı kabul edilir.

## c) Ölçme Aracının Kullanışlılığı

Kullanışlılık, bir ölçme aracında aranan üçüncü temel özelliktir. **Kullanışlılık**; bir ölçme aracının geliştirilmesinin, çoğaltılmasının, uygulanmasının, cevaplanmasının ve puanlanmasının kolay olmasıdır. Ayrıca bir testin maliyeti ne kadar düşük olursa, kullanışlılığı o ölçüde yüksek olur.

Ölçme aracının kullanışlılığı ile ilgili genel özellikler şunlardır:

- **Hazırlanması:** Öğretmenin ölçme aracını hazırlarken zorlanmaması ve kısa sürede tamamlaması,
- **Uygulanması:** Öğretmenin ölçme aracını uygularken güçlük çekmemesi ve kısa sürede uygulaması,
- **Cevaplanması:** Öğrencilerin ölçme aracını cevaplarken zorluk çekmemeleri ve kısa sürede cevaplamaları,
- **Puanlanması:** Öğretmenin ölçme aracını puanlarken zorlanmaması ve kısa sürede puanlaması,
- **Maliyetin düşüklüğü:** Ölçme aracını okula maddi yönden yük getirmemesidir.