



# UZMAN ÖĞRETMENLİK SINAVI

## 2. MODÜL ÖZETİ

### ÖLÇME VE DEĞERLENDİRME

**Erhan BOYRAZ**

**Sosyal Bilgiler Öğretmeni**

# 1. ÖLÇME VE DEĞERLENDİRMENİN TEMEL KAVRAMLARI

## 1.1. EĞİTİM SİSTEMİNDE ÖLÇME VE DEĞERLENDİRMENİN YERİ:

Eğitim bir sistemdir çünkü Birincisi gerçekleştirilmek istenen hedefler vardır. Bunlar:

**1. Uzak hedefler:** Devletlerin hedefleridir, bazı ülkelerde çerçevesi anayasa ile güvenceye alınmıştır.

**2. Genel hedefler:** Daha işevuruk hedeflerdir. Kurumların (MEB, MEB genel müdürlükleri, il millî eğitim müdürlükleri, okul yönetimleri vb.) hedefleri vb.

**3. Özel hedefler:** Dersin hedefleridir, bir dersin kazanımlarını kapsar.

Eğitimin bir sistem olmasının ikinci nedeni, bu hedefleri gerçekleştirmek için farklı öğelerden oluşmasıdır:

**1. Girdi:** Eğitim-öğretim için gerekli her şey eğitim sisteminin girdisini oluşturur. Örneğin öğrenci, öğretmen, okul, çalışanlar, yönetim, eğitim programı vb.

**2. Süreç:** Girdilerin hedefler doğrultusunda bir araya getirildiği, dersin hedeflerinin öğrenciye kazandırıldığı eğitim durumlarını kapsar. Süreç, öğretmenler, okul yönetimi, il yöneticileri ülke genelinde belirli aralıklarla kontrol edilmektedir. Kontrollerin sonucu olumlu ise süreç devam etmekte değilse sürece müdahale edilmektedir.

**3. Çıktı:** Öğrencilerin sahip olduğu niteliklerdir. Eğitim sistemi için çıktılar, dar anlamda öğrencinin bir dersin sonunda ulaştığı hedefler ve kazanımları ifade ederken geniş anlamda öğrencinin öğretim sürecinin başlangıcından sonuna kadar sahip olduğu bilgi beceri ve yetkinlik anlarını ifade etmektedir.

**NOT:** Niteliğin ölçülmesi çok zor ancak nicel veriler açısından değerlendirdiğimizde bir ülkede ortalama eğitim bir birim arttığında, ekonomik büyüme %0,7 artmakta.

Dünya Ekonomik Forumuna göre 2025 yılında çalışanların sahip olması gereken 10 temel beceri:

- Analitik düşünme ve yenilik/yaratıcılık
- Etkin öğrenme ve öğrenme stratejileri
- Karmaşık problem çözme
- Eleştirel düşünme ve analiz
- Yaratıcılık, orijinallik
- Liderlik ve sosyal etki
- Teknoloji kullanımı

## ÖLÇME, ÖLÇÜT VE DEĞERLENDİRMENİN KAVRAMSAL TEMELLERİ:

**1- Ölçme:** Öğrencinin belirli bir özelliğini gözleyerek o özelliğe sayı, sembol ya da sıfat/kategori adı verme işlemidir.

Bir eğitim sisteminde ölçme ve değerlendirme olmaksızın eğitim sisteminin geçerli ve güvenilir kontrolü sağlanamaz.

**2- Ölçüt:** Ölçülen özellik hakkında karar alabilmek /yargıya varabilmek / değerlendirme yapabilmek için dayanak alınan referans noktası ya da referans aralığıdır.

**Değerlendirme:** Bir ölçme sonucunu (ölçüm) en az bir ölçüte vurarak ölçülen nitelik hakkında karar verme / yargıda bulunma işlemidir. Değerlendirmenin basamakları

**a. Ölçme ⇒ Ölçme sonucu (ölçüm), b. Ölçüt ve c. Karardır**

Ölçüt, ölçme ve değerlendirme arasında köprü görevi görür. Ölçüt değişirse değerlendirme de değişir.

Örneğin

Ölçme	Ölçüt	Değerlendirme
Irmak Yabancı Dil Bilgisi Seviye Tespit Sınavından (YDS) 50 almıştır.	50	Yüksek lisansa başvuru yapabilir.
Irmak Yabancı Dil Bilgisi Seviye Tespit Sınavından (YDS) 55 almıştır.	55	Doktora başvuru yapamaz.

## ÖLÇME TÜRLERİ

### 1. Doğrudan Ölçme (Temel Ölçme):

Ölçmeye konu olan özelliğin doğrudan gözlenerek ölçümün elde edildiği ölçme türüdür. Genellikle beş duyu organı ile algılanan özellikler doğrudan gözlenebilirler. Örneğin *bir sınıftaki öğrenci sayısı, parktaki salıncak sayısı, öğrencilerin cinsiyetleri, boyları, kalem tutma becerisi, pas verme becerisi vb.* özellikler doğrudan gözlenebilir.

### 2. Dolaylı Ölçme (Göstergeyle Ölçme):

Bazı özellikler doğrudan gözlenemez. Bu özellikler ancak onun göstergesi olduğu bilinen ya da kabul edilen davranışlar aracılığı ile gözlenerek ölçülebilir. Örneğin *öğrencilerin dört işlem becerisi, derse yönelik tutumları, genel muhakeme düzeyleri, öz düzenleme becerileri, sınav kaygıları, motivasyon düzeyleri vb.* özellikler ancak dolaylı olarak gözlenebilir.

### 3. Türetilmiş Ölçme:

Ölçülmek istenilen özellik kendisinden farklı iki ya da daha fazla özelliğin arasındaki matematiksel bir bağıntı (dört işlem) yardımıyla belirleniyorsa bu ölçme türüne “türetilmiş ölçme” adı verilir. Örneğin *hız (yol / zaman), yoğunluk (kütle / hacim), zekâ bölümü (zekâ yaşı / takvim yaşı x 100) vb.* özellikler türetilmiş ölçmeye örnek olarak verilebilir.

## ÖLÇÜT TÜRLERİ

### 1- Mutlak Ölçüt (Kriter Referanslı Ölçüt):

Eğer ölçüt ölçme işleminin yapıldığı grubun özelliklerinden bağımsız olarak belirleniyorsa “mutlak”tır. Mutlak ölçüt genellikle ölçme işlemi öncesinde ilan edilir. Örneğin *dersten geçme notu 50’dir.*

### 2- Bağıl Ölçüt (Norm Referanslı Ölçüt):

Eğer ölçüt ölçme işleminin yapıldığı grubun belirli bir özelliğine / normuna dayalı olarak belirleniyorsa “bağıl”dır. Bağıl ölçüt ancak ölçme işlemi sonrasında belirlenebilir. Örneğin *aritmetik ortalamaya / ortancaya / moda denk ve üzerinde puan alan geçer ya*

*da akademik başarısı yüksek ilk on öğrenci proje için seçilecek gibi..*

## AMACA GÖRE DEĞERLENDİRME TÜRLERİ

### 1-Tanıma-Yerleştirmeye Yönelik Değerlendirme (Diyagnostik Değerlendirme)

**Sınıf içi ölçme ve değerlendirme:** Öğretim sürecinin başında, öğrencilerin önceki kritik nitelikteki öğrenmelerini belirlemek amacıyla yapılır. Diğer bir ifadeyle amaç öğrencinin ders bağlamındaki hazırbulunuşluluk düzeyini belirlemektir. Amaç not vermek değildir. Bu amaçla uygulanan ölçme araçlarına da “hazırbulunuşluluk testi” adı verilir.

**Okul / kurum geneli ölçme ve değerlendirme:** Yine öğretim sürecinin başında, öğrenciyi tanımak ve onu uygun olan programa / kura / sınıfa yerleştirmek amacıyla yapılır. Amaçlanan, not vermek değildir. Hedeflenen, öğrencinin niteliklerine uygun grupların oluşturulmasıdır. Zaman, para ve emek tasarrufu sağlar. Muafiyet sınavları ve seviye tespit sınavları buna örnek gösterilebilir.

### 2-Biçimlendirme-Yetiştirmeye Yönelik Değerlendirme (Formatif Değerlendirme)

Belirli bir konu / ünite vb. sonunda öğrencilerin öğretime konu olan davranışların ne kadarını kazandığını, diğer bir deyişle ünitedeki öğrenme eksiklerini belirlemek amacıyla yapılır. Amaç yine not vermek değildir. Bu değerlendirme türünde öğrenciye o ünitedeki öğrenme eksikleri bildirilir, bir sonraki kapsama geçmeden öğrenme eksikleri giderilir ya da öğrenme eksiklerinin giderilip giderilmediği denetlenir. Biçimlendirme-yetiştirme amacıyla yapılacak değerlendirmede kullanılan testlere alanda “izleme testi” ya da “tarama testi” denir. Not verilmeden yapılan kısa sınavlar (quiz), ünite tarama testleri bu amaçla yapılan ölçme ve değerlendirmeye örnek olarak verilebilir.

### 3- Değer Biçmeye / Düzey Belirlemeye Yönelik Değerlendirme (Summatif Değerlendirme)

Belirli bir öğretim sürecinin sonunda, not vermek amacıyla yapılan değerlendirmedir. Değer biçmeye yönelik değerlendirme, öğrencinin belirli bir öğretim süreci sonundaki durum tespiti için yapılır çünkü artık hedeflenen öğretim sürecinin sonuna gelinmiştir. Bu

amaçla uygulanan testlere “alanda erişim testi” denir. Bitirme sınavları, sertifika sınavları, üniversitedeki vize ve finaller, eğitim-öğretim kurumlarında not vermek amacıyla uygulanan yazılı yoklamalar bu ölçme ve değerlendirme türüne örnek olarak verilebilir.

#### 4- Rehberlik Amaçlı Değerlendirme

Rehberlik amaçlı değerlendirme kabaca iki bağlamda ele alınabilir. Bunlar özel eğitim ve mesleki rehberlik hizmetleridir.

**Özel eğitim:** Belirli alanlarda özel gereksinimi olan çocuklara uygun eğitim ortamları ve/veya programı uygulamak amacıyla ölçme ve değerlendirme süreçleri işe koşulmaktadır. Özel gereksinim kavramı zihinsel gerilikten öğrenme güçlüğüne, otizmden diğer spesifik gerilik alanlarına, üstün zekâdan işitme/görme vb. engel alanlarına kadar çok geniş bir yelpazede ele alınır. Özel gereksinimi olan çocuklara uygun eğitim olanaklarının sunulabilmesi için çocuğun ihtiyacı olan alanlar, eğitsel tanılama amacıyla genellikle standart testler veya ölçüt bağımlı testler (beceri öğretimde) ile belirlenmeye çalışılır. Bu tür testlerin sonuçları doğrultusunda, tıbbi ve eğitsel / gelişimsel olarak tanı alan çocuklar için bireysel eğitim programları (BEP) hazırlanır. Eğitsel değerlendirme bağlamında RAM’lar bu konudaki tek yetkilidir.

**Mesleki rehberlik:** Öğrencilerin alan seçmelerinde ve özellikle bir üst öğretim kurumuna yönlendirilmesinde, özelliklerine uygun alan / meslek seçmeleri hem öğrenci hem de ülkenin insan gücü kaynağının doğru planlanması açısından çok önemlidir. Bu amaçla öğrencilerin kendi yetenekleri, kişilik özellikleri, ilgileri vb. ile özelliklerini belirlemek amacıyla standart testlerden yararlanılmaktadır. Böylece öğrencinin kendi özellikleri hakkındaki farkındalıklarının artırılması ve dolayısıyla buna uygun alan / meslek seçmeleri sağlanmış olur.

#### 5- Program Değerlendirme

Eğitim programları bileşenlerinin yapısı ve niteliğine ilişkin olarak sistematik bilgi toplama ve değerlendirme sürecidir. Program değerlendirme eğitim programlarının planlanması, mevcut programlar ve/veya ürünlerin etkililiğinin değerlendirilmesi ve eğitim programlarının ve/veya ürünlerin geliştirilmesi amacıyla yapılır.

## **BİR ÖLÇME ARACINDA BULUNMASI GEREKEN PSİKOMETRİK NİTELİKLER**

**Geçerlilik** ölçmek istenilen özelliğin, başka özelliklerle karıştırılmadan, doğru ve tam olarak ölçülebilmesidir. Diğer bir ifadeyle ölçme aracından elde edilen puanın amaca hizmet etme derecesidir. **Güvenilirlik** ölçme işleminden elde edilen puanların tesadüfi hatalardan arınık olma derecesidir. Diğer bir ifadeyle puanların kararlı, tutarlı ve duyarlı olmasıdır. **Kullanışlılık** ise ölçme aracının geliştirilmesinin, uygulanmasının ve puanlanmasının kolay ve ekonomik (zaman, para, emek, araç gereç vb. açıdan) olması ile ilgilidir.

## **HATA ÇEŞİTLERİ**

**Sabit Hata:** Miktarı ölçmeden ölçmeye değişmeyen, diğer bir deyişle her ölçme işlemine aynı miktarda karışan hatalardır. Örneğin marketteki terazi, üzerinde herhangi bir nesne yokken terazi -120 g gösteriyorsa ne tartılırsa tartılsın 120 g eksik ölçülecektir. Bir öğretmen, sınavında herkese 10 puan fazla veriyorsa yine karışan hata sabit olacaktır.

**Sistemik Hata:** Ölçülen büyüklüğe, öğretmene ya da ölçme koşullarına göre miktarı değişen hatalardır. Örneğin marketteki terazi, her bir kilogramda 120 g eksik tartıyorsa üzerine konulan nesnenin ağırlığı arttıkça hata miktarı da artacaktır. Bir öğretmen, sınavında yazısı kötü olandan puan kırıyorsa yine karışan hata sistematik olacaktır.

**Tesadüfi (Rastlantısal) Hata:** Şansla ortaya çıkan ne yönde ve ne ölçüde karıştığı genellikle bilinemeyen hatalardır. Ölçmelere tek yönlü olarak karışmaz, ölçme sonuçlarına bazen pozitif bazense negatif yönde etki eder. Sabit ve tesadüfi olmayan hatalar, tesadüfi değişken olma özelliğine sahiptir. Bu nedenle psikometride hata kuramı tesadüfi hatalar üstüne kurulmuştur. Tesadüfi hataların genellikle dört kaynağı olduğu kabul edilir. Bunlar:

- Ölçme işlemini yapan kişiden / öğretmenden kaynaklanan hata
- Ölçme aracından kaynaklanan hata
- Bireyden / öğrenciden kaynaklanan hata
- Fiziksel ortamdan kaynaklanan hata

## GEÇERLİLİK, GÜVENİLİRLİK VE HATA İLİŞKİSİ

Geçerlilik tüm hata kaynaklarından etkilenirken klasik test kuramına göre güvenilirlik yalnızca tesadüfi hatalardan etkilenir.



Bu nedenle güvenilirlik, geçerlilik için bir ön şart ancak yeterli şart değildir. Diğer bir ifade ile bir testin güvenilir olması onun geçerli olacağı anlamına gelmez ancak bir test geçerli ise büyük olasılıkla güvenilirir.

## KORELASYON:

Korelasyon (co-relation), en az iki değişken arasında karşılıklı bir ilişki bulunup bulunmadığı, eğer ilişki varsa bu ilişkinin yönü ve miktarı hakkında bilgi veren istatistik bir tekniktir. “r” ile sembolize edilir. Örneğin

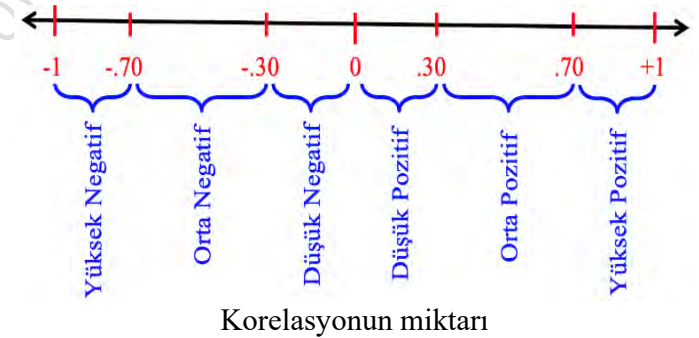
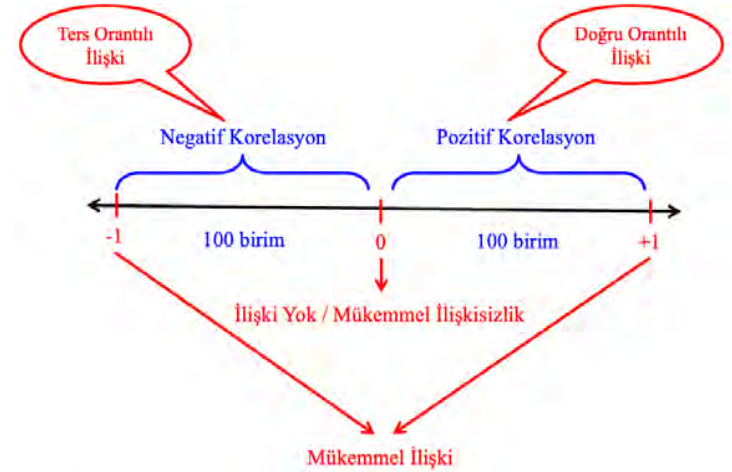
- Öğrencilerin derse ilişkin tutumları ile ders başarıları arasında bir ilişki var mıdır?
- Öğretmenlerin ders saati yükleri ile iş doyumları arasında bir ilişki var mıdır?
- Saç uzunluğu ile zekâ arasında bir ilişki var mıdır?

Korelasyon -1 ile 1 arasında değer alır. Bu iki değer arasında matematiksel olarak sonsuz birim vardır ancak kullanışlılık açısından 100 birim negatif korelasyonda, 100 birim pozitif korelasyonda, sıfır ile birlikte toplam 201 birimlik bir skala üzerinden değişkenler arasındaki ilişkinin yönü ve miktarı değerlendirilir.

Pozitif korelasyon iki değişken arasında doğru orantılı ilişki anlamına gelir. Örneğin ders çalışma süresi ile sınav notu ya da gelir ile tüketim arasında pozitif bir korelasyonun elde edilmesi beklenir. Negatif korelasyon, iki değişken arasında ters orantılı ilişki anlamına gelir. Örneğin ders süresi ile dikkat ya da

yükseklik ile sıcaklık arasında negatif bir korelasyonun elde edilmesi beklenir.

$\pm 1$  mükemmel korelasyon anlamına gelir ancak 1 mükemmel doğru orantılı ilişki anlamına gelirken, -1 ise mükemmel ters orantılı ilişki anlamına gelir. Diğer taraftan sıfır korelasyon iki değişken arasında sistematik bir ilişkinin olmadığı anlamına gelir. Bu duruma mükemmel ilişkisizlik de denir. Örneğin öğretmenlerin boyları ile aylık ücretleri arasındaki korelasyonun sıfır olması beklenen bir durumdur.



**Not 1:** Korelasyon mutlak değer olarak değerlendirilmelidir. Bir korelasyon katsayısının negatif ya da pozitif olması büyüklük-küçüklük belirtmez, yön bildirir.

**Not 2:** Korelasyon katsayısı ile neden-sonuç ilişkisi kurulamaz. Değişkenler arasında doğru ya da ters orantılı bir ilişki olması, söz konusu değişkenler arasında bir neden-sonuç ilişkisinin varlığı anlamına gelmez.



## GEÇERLİLİK SORGULAMA YÖNTEMLERİ:

### 1- KAPSAM GEÇERLİLİĞİ:

Kapsam geçerliliği özellikle başarı testlerinde aranan bir geçerlilik sorgulamasıdır. Eğitimde öğretmen bir program dâhilinde önceden belirlenen davranışları kazandırmayı amaçlar. Belirli bir zaman sonra davranışların ne kadarının kazanıldığını belirlemek isteyen öğretmen, geliştirdiği testte öğretime konu olan içeriğin testte ne ölçüde temsil edildiğini belirlemek durumundadır. Bu anlamda kapsam geçerliliği bir testin ölçülmek istenen davranışları ne derece kapsadığıyla ilgilidir.

### 2- ÖLÇÜT DAYANAKLI GEÇERLİLİK:

Ölçme aracından elde edilen puanların ölçüt bir puanla (testin tahmin etmeye çalıştığı ve geçerliliği yüksek bir puan) karşılaştırılarak geliştirilen ölçme aracının geçerliliğine ilişkin nitelendirme yapılır.

➤ **YORDAMA GEÇERLİLİĞİ:** Yordama, tahmin demektir ancak her tahmin yordama değildir. Bir tahminin yordama olabilmesi için elde geçerli ve güvenilir bir veri olması ve bu verinin sınanabilir, sayısal nitelikte, belirli analizlere tabi tutuluyor olması gerekiyor. Diğer bir deyişle yordama, eldeki bu nitelikteki veriden yola çıkarak geleceğe, henüz gerçekleşmemiş bir olguya ilişkin yapılan tahmindir.

➤ **ZAMANDAŞ GEÇERLİLİK:** Bu geçerlilik türü alanyazında hâlihazır geçerlilik, benzer ölçekler geçerliliği, uygunluk geçerliliği adı ile de anılmaktadır. Zamandaş geçerlilik sorgulamalarında ölçüt puan eş zamanlı olarak elde edilebilir. Geliştirilen ölçme aracı ile ilişkili olabilecek nitelikleri ölçen ve geçerliliği yüksek bir aracın puanı ölçüt puan olarak ele alınabilir. Daha sonra geliştirilen ölçme aracı ve ölçüt araç aynı zamanda uygulanıp iki araçtan elde edilen puanlar arasındaki ilişki incelenir.

➤ **3- YAPI GEÇERLİLİĞİ:** Yapı geçerliliği, bir testin dayandığı kuramsal temelleri ne derece iyi örneklediğiyle ilgilidir. İnsan davranışları ve özellikleri çoğunlukla soyut bir yapıya sahiptir. Zekâ, sevgi, merak, sosyal uyum ve ruh sağlığı gibi soyut

kavramları psikolojik testlerle ölçebilmek için önce bu kavramların içeriğinin bilinmesi gerekir. Birey ne tür davranışlar gösterdiği zaman hangi niteliklerin varlığına ya da yokluğuna karar verilebileceğine ilişkin ölçütler geliştirilmesi gerekmektedir. Yapı geçerliliği bir yandan testin ölçtüğü niteliklerin neler olduğunu araştırma, diğer yandan testi alan kişilerin elde ettikleri puanların ne anlama geldiğini açıklama çabalarıyla ilgilidir.



## GÜVENİLİRLİK

Güvenilir bir ölçme aracı, aynı özelliklerle ilgili olarak arka arkaya yapılan ölçmelerde yaklaşık olarak aynı sayısal sonucu verir; diğer bir ifadeyle bir test, aynı gruba iki ya da üç kez uygulandığında gruptaki her bir kişi bütün uygulamalarda yaklaşık olarak aynı puanı almalıdır. Bir testin ölçmek istediği özelliği ölçebilmesi için o testin söz konusu olan özelliği kararlı olarak ölçmesi gerekir. Güvenirlik çalışmalarının odak noktası şudur: "Eğer kişi iki defa teste tabi tutulursa iki testten aldığı puanlar birbirine benzer midir ya da birbirine ne kadar yakındır?" Bu noktada gözlenen puan, gerçek puan, ölçmenin standart hatası ve güvenilirlik katsayısı kavramlarına değinmek gerekmektedir.



Güvenilirlik, genellikle birden çok uygulamaya dayalı yöntemler ve tek uygulamaya dayalı yöntemler

başlıkları altında ele alınır. Birden çok uygulamaya dayalı yöntemler altında test-tekrar test ve eşdeğer (paralel) testler yöntemleri, tek uygulamaya dayalı yöntemler altında ise eşdeğer yarılar, KR-20, KR-21, Cronbach alfa, Hoyt'un varyans analizi, McDonald omega vb. yöntemler yer almaktadır.

➤ **TEST-TEKRAR TEST YÖNTEMİ:** Bu yöntem ile test güvenilirliğini test etmek için bir test, aynı gruba, belli bir zaman aralığıyla iki kez uygulanır. Daha sonra bireylerin birinci uygulamadan aldıkları puanlarla ikinci uygulamadan aldıkları puanlar arasındaki korelasyon hesaplanır. Elde edilen korelasyon katsayısına **kararlılık (devamlılık / istikrarlılık) katsayısı** adı verilir.

➤ **TEST YARILAMA (EŞDEĞER YARILAR / İKİ YARI GÜVENİLİRLİĞİ) YÖNTEMİ:** Bu yöntemle güvenilirliği tahmin etmede uygulanmış bir test iki eşdeğer yarıya bölünür ve bireylerin iki yarıdan aldıkları puanlar arasındaki tutarlılık incelenir. Buradaki en temel sorunlardan biri testin iki eşdeğer yarıya nasıl bölünmesi gerektiği ile ilgilidir. En sık başvurulanan yöntemler: a) ilk yarı ve son yarı b) tek ve çift ve c) rastlantısalıdır. İlk ve son yarı yöntemi her test için uygun değildir

➤ **KUDER-RICHARDSON 20 VE 21 YÖNTEMLERİ:** Kuder-Richarson 20 ve 21 ya da kısaca KR-20 ve KR-21, testin kendi içinde tutarlı bir bütün oluşturup oluşturmadığı ile ilgili bilgi verir. Bu nedenle her iki yöntemle de testin iç tutarlılığı değerlendirildiğinden bu yöntemlerden elde edilen katsayıya **iç tutarlılık katsayısı** adı verilir.

KR-20 ve KR-21 ile güvenilirlik kestirimi, puanlamanın kategorik olduğu ölçme araçları için uygundur. Puanlamanın kategorik olması, doğru yanıtlara 1 puan, yanlış ve boş bırakılan maddelere 0 puan vererek puanlamanın yapıldığı ya da maddenin yanıtının iki seçenekli verildiği durumlardır. Eğer testteki maddeler farklı ağırlıklarla puanlanmışsa ya da test puanları şans başarısı için düzeltilmişse bu formüller kullanılmaz.

➤ **CRONBACH ALFA YÖNTEMİ:** Eğer ölçme aracının puanlaması çok kategorili ise KR-20 ile aynı mantık üzerine kurulu Cronbach alfa hesaplanır. Elde edilen katsayının ismi yine **iç tutarlılık katsayısıdır**.

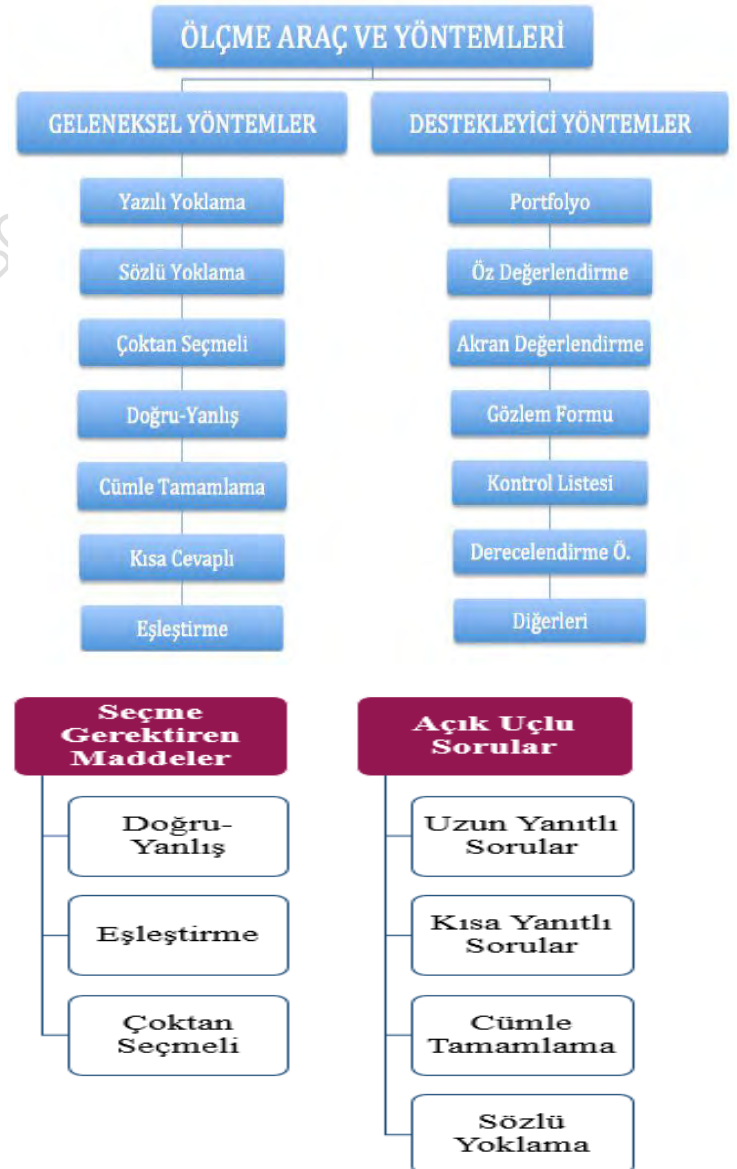
Derecelendirme ölçekleri puanlamanın çok kategorili olduğu araçlara örnek verilebilir.

## TEST GELİŞTİRME VE MADDE / SORU TÜRLERİ

### TEST TÜRLERİ:

Testi Alan Kişi Sayısına Göre	Testin Uygulanış Süresine Göre	Testin Ölçtüğü Niteliğe Göre	Değerlendirme Yaklaşımına Göre	Hazırlanış Biçimine Göre	Veri Toplama Tekniğine Göre
1. Bireysel	1. Süreli	1. Hız testleri	1. Objektif	1. Standart	1. Performans
2. Grup	2. Süresiz	2. Güç Testleri	2. Subjektif	2. Öğretmen yapımı	2. Kağıt-Kalem

### BAŞARININ ÖLÇÜLMESİNDE YÖNTEMLER



## GELENEKSEL ÖLÇME YÖNTEMLERİ

**1- DOĞRU-YANLIŞ MADDELERİ:** Cevaplayıcının verilen ifadelerin doğru mu, yanlış mı olduğunu belirlemesinin istendiği madde türüdür. Doğruyanlış maddelerinde verilen bir tür “önerme”dir.

### 2- EŞLEŞTİRME MADDELERİ:

Eşleştirme maddeleri, iki grup hâlinde verilen ve birbirleriyle ilgili olan bilgi öğelerinin belli bir açıklamaya göre eşleştirilmesini gerektirir.

### 3- ÇOKTAN SEÇMELİ MADDELER:

### 4- ESSAY (YAZILI YOKLAMA):

Cevaplayıcıların sorulara cevap verme konusunda sınırsız bir özgürlüğü vardır. Sınırsız cevap özgürlüğünün hem avantajı hem de dezavantajı bulunmaktadır.

Avantajı: Öğrencinin yalnızca sahip olduğu bilgiyi değil, aynı zamanda zihninin nasıl işlediğini anlama olanağı sağlar.

Dezavantajı: Öğrenci sınırlı bir bilgiye sahipse soruyla ilgisi olmayan, aklına estiği gibi uzun cevaplar yazabilir.

### 5- KISA CEVAPLI SORULAR:

Kısa cevaplı testler; öğrencinin bir sözcük, bir rakam, bir tarih ya da en çok bir cümle ile cevaplandırabileceği sorulardan oluşur. Kısa cevaplı testler, öğrencinin aklının estiği cevabı yazmasını büyük ölçüde engelleyecek bir özelliğe sahiptir. Bundan ötürü maddeleri cevaplandıran bir cevaplayıcının yazacağı cevaplar çok kısadır.

### 6- CÜMLE TAMAMLAMA SORULARI:

Cümle tamamlama, kısa cevaplı / sınırlı cevap gerektiren soru grubudur. Bu tür sorularda öğretmen önemli gördüğü bir cümleyi alır, kritik gördüğü bir ya da birden Ögfazla yeri çıkarır ve yerine bir boşluk koyar. Öğrenciden o cümleyi anlamlı bir biçimde tamamlaması istenir. Cümle tamamlama soruları boşluk doldurma olarak da geçmektedir.

### 7- SÖZLÜ YOKLAMA:

Sözlü sınavlar çok iyi organize edilse / kurgulansa dahi ciddi bir ölçme hatası barındırma potansiyeline sahiptir. Ancak bütün bu potansiyel sınırlılıklarına rağmen sözlü yoklamalar özellikle dile dayalı becerilerin ölçülmesi için geleneksel yöntemler içindeki tek türdür. Örneğin *sözlü anlatım, diksiyon, yabancı dil dersinde konuşma, solfej ve şarkı söyleme vb.*

## DESTEKLEYİCİ DEĞERLENDİRME YAKLAŞIMLARI

### 1- PORTFOLYO:

Öğrencinin bir bütün olarak gelişim ve öğrenme süreci ile ürünlerini gösteren, aynı zamanda değerlendirilmesini de sağlayan sistemli ve amaçlı olarak oluşturulmuş dosyalardır. Portfolyolar ile amaçlanan, öğrenciyi çok yönlü olarak ve öğrencinin gelişimini bir süreç olarak izlemek ve değerlendirmektir. Portfolyolar ile öğrencinin öz geçmişi, ilgileri, tutumları, becerileri, yetenekleri, öğrenme biçimi, sosyal özellikleri vb. özelliklerinin tanınması amaçlanır. Sürecin paydaşları; öğrenci, öğretmen, arkadaş, aile ve iletişimde olduğu diğer kişilerdir. Öğrencinin yalnızca okulda değil; aile ortamında, oyun ortamında ve dış çevrede tanınması amaçlanır.

#### Portfolyo Oluşturma Süreci



### 2- PERFORMANS DEĞERLENDİRME



#### ❖ Öğrenci Değerlendirmeleri

- **ÖZ DEĞERLENDİRME:** Öğrencinin belirli bir konuda (örneğin bir ürünü ortaya koymada gösterdiği performans vb.) kendi kendisini değerlendirmesine denir.
- **AKRAN DEĞERLENDİRME:** Öğrencinin ortaya koyduğu performansa ilişkin arkadaşlarının değerlendirmesine denir.

#### ❖ Öğretmen değerlendirmeleri

- **RUBRİK (DERCELİ PUANLAMA ANAHTARI):** Performans ürünlerinin değerlendirilmesinde en sık kullanılan araçlardan biridir. Amaç, öğretmen tarafından ürünün genel izlenimle puanlamasındaki



öznelliğini azaltmaktır. Rubrikler ikiye ayrılır: a. Bütüncül (Holistik) rubrik b. Analitik rubrik

- **KONTROL LİSTELERİ:** Gözlenen performans ürününün ölçütlere uygunluğunu “evet-hayır”, “var-yok”, “gösterdi-göstermedi” vb. bir biçimde kategorik (1-0) olarak puanlama amacıyla kullanılan araçlardır.

- **DERECELEME ÖLÇEKLERİ:** Bu araçların kullanımında performansa dayalı işlemler ilk baştan sonuna kadar listelenir ve davranışın karşısına davranışın gösterilme derecesi en az üçlü [*örneğin tam gösterildi (3), kısmen gösterildi (2) ve gösterilmedi (1)*] bir biçimde derecelendirilir.

- **GÖZLEM FORMLARI:** Öğrenme çıktılarının somut olarak gözlenebildiği bazı alanlarda bu yöntem oldukça kullanışlıdır.

### GELENEKSEL VE DESTEKLEYİCİ YÖNTEMLERİN KARŞILAŞTIRILMASI

GELENEKSEL YÖNTEMLER	DESTEKLEYİCİ YÖNTEMLER
Ürün değerlendirilir.	Süreç ve ürün birlikte değerlendirilir.
Öğrencinin ulaştığı noktanın tespiti önemlidir.	Ne öğrendikleri yanında, öğrendiklerini nasıl kullandıklarıyla ilgilenilir.
Essay dışında genellikle üst düzey düşünme becerilerinin değerlendirilmesinde yetersizdir.	Üst düzey bilişsel düşünme becerilerine odaklanır.
Değerlendirme öğrenmeden ayrıdır.	Değerlendirme öğrenmeyle bütünlüştür.
Bireyden ziyade gruba odaklıdır.	Odak noktası gruptan ziyade bireydir.
Başarının bireyler arası değerlendirilmesine odaklıdır.	Öğrencinin bireysel olarak gelişimine odaklıdır.
Geçerlilik ve güvenilirlik kontrolü daha kolaydır.	Geçerlilik ve güvenilirlik problemi olabilir.

## EĞİTİM İZLEME ARAŞTIRMALARI: TARİHİ, ÖNEMİ VE TÜRKİYE'DEN SONUÇLAR

### Uluslararası İzleme Çalışmaları

➤ birçok ulusal izleme araştırmasına temel teşkil eden ilk çalışmalardan biri 1964 yılında gerçekleştirilen uluslararası matematik izleme çalışmasıdır.

TIMSS'in de geliştiricisi olan kurum (IEA) tarafından gerçekleştirilen bu çalışma 1960'larda matematik üzerinde başlamıştır.

➤ 1960'lı yıllarda matematik alanı ile başlayan çalışmaların bugün kapsamı büyük ölçüde genişlemiştir. Örneğin PISA ile beraber değerlendirilen alanlar artmış; matematik, fen ve okuma becerileri üç ana alan olarak belirlenmiştir

### Günümüzde En Yoğun Katılım Gösterilen Çalışmalar

PISA, Ekonomik Kalkınma ve İşbirliği Örgütü (OECD) tarafından geliştirilen ve üç yıllık periyotlarla uygulanan izleme çalışmasıdır. OECD, Türkiye'nin kurucu üyesi olduğu bir örgüt olup temel amacı tüm üye ülkeler arasındaki ekonomik iş birliklerini ve kalkınmayı destekleyecek programlar oluşturarak birlikte kalkınmayı sağlamaktır. Bu bağlamda OECD, eğitimi, ekonomik kalkınmaya katkı sunan ve çağdaş vatandaşlık becerilerine sahip insan kaynağının geliştirilmesinde en önemli araç olarak görmektedir.

Bir diğer izleme çalışması olan TIMSS ise Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (IEA) tarafından geliştirilmekte ve dört yıllık periyotlarla uygulanmaktadır. Türkiye uzun süredir hem PISA'ya hem de TIMSS'e katılım göstermektedir. Son olarak PIRLS de yine IEA tarafından beş yıllık periyotlarla gerçekleştirilen bir izleme çalışmasıdır. Türkiye uzun bir aradan sonra 2021'de tekrar PIRLS'e katılmış ancak bu uygulamanın sonuçları 2022 yılının sonunda açıklanacaktır.

## PISA KAPSAMI VE SONUÇLARI

OECD, ekonomik kalkınmanın ve sürdürülebilirliğinin sağlanmasındaki en önemli kaynak olan beşerî sermayeyi üreten faktör olarak eğitimi konumlandırmaktadır. Eğitimin ekonomik kalkınma üzerindeki belirleyici rolü dolayısıyla eğitim niteliğini detaylı olarak ele alan ve üç yıllık periyotlarla uygulanan PISA çalışmasını gerçekleştirmektedir. En güncel PISA uygulaması olan 2018 döngüsüne 79 ülke/ekonomi katılım göstermiştir. Dünyanın dört bir yanından farklı coğrafyalardan, farklı kültürlerden, farklı dinamiklere sahip eğitim sistemleri olan ülkeler yaygın bir şekilde PISA'ya katılım göstermektedir. PISA, 15 yaş grubunu hedef grup olarak almaktadır. Bunun nedeni birçok Avrupa ve OECD ülkesinde 15 yaşa tekabül eden dönemin yaklaşık olarak ortaokulun sonuna denk gelmesi ve bu dönemde zorunlu eğitimin tamamlanmasıdır. Dolayısıyla PISA, zorunlu eğitim dönemini bitiren öğrencilerin henüz iş gücü piyasasına katılmadan ya da eğitimlerine devam etmeden matematik, fen ve okuma becerileri alanlarında hangi yeterliklere sahip olduğuna dair bilgi vermektedir.

Türkiye, matematik ve fen okuryazarlığında son 15 yılda, katıldığı tüm PISA uygulamalarından en yüksek ortalama puana ulaşmış; okuma becerilerinde ise performansını önemli ölçüde artırmıştır. Elbette hedef OECD ortalamasına ulaşmak ve OECD ortalamasını geçmektir ama öğrencilerimizin puanlarını artırmış olması olumlu bir gelişmedir.

## TIMSS KAPSAMI VE SONUÇLARI

TIMSS, IEA tarafından dörder yıllık döngüler şeklinde gerçekleştirilmektedir. PISA ile TIMSS arasındaki temel fark, PISA'da zorunlu eğitim dönemini tamamlayan öğrencilerin iş gücü piyasası tarafından ve ekonomik kalkınma açısından istenen, uygulamaya dönük okuryazarlık becerilerinin hangi seviyede olduğu değerlendirilmekte iken TIMSS'te eğitim programı (curriculum based) ile ilişkili becerilerin değerlendirilmesidir. TIMSS'te 4 ve 8. sınıf düzeyindeki öğrencilerin ilgili dönemin eğitim programlarından beklenen özelliklere hangi düzeyde sahip olduklarına dair matematik ve fen alanlarında ayrı ayrı uygulamalar yapılmakta ve ilgili eğitim programlarında öğrencilerin okul başarısına dair daha doğrudan çıktılar elde edilmektedir.

TIMSS, ilk defa 1995 yılında uygulanmaya başlanmış olup 2019'da araştırmanın yedinci döngüsü tamamlanmıştır. Türkiye TIMSS'e 1999 yılında ilk defa 8. sınıf düzeyinde katılım gösterirken 2011 yılından bu yana son üç döngüye hem 8. sınıf hem de 4. sınıf düzeyinde katılmıştır.

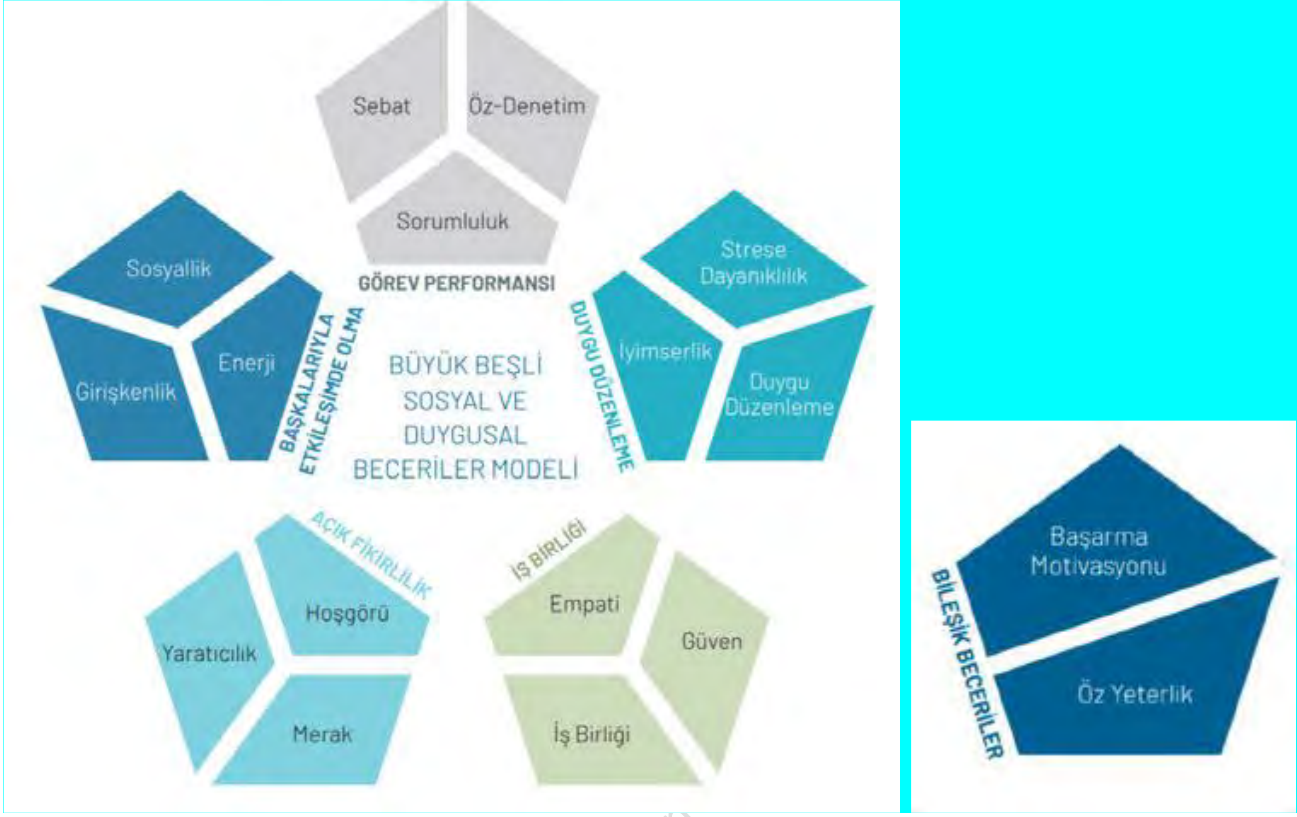
PIRLS de IEA tarafından uygulanmakta olup okuma boyutunu içermesi sebebiyle TIMSS'i tamamlar nitelikte bir çalışmadır. PIRLS, öğrencilerimizin okuma becerilerine ve okuma düzeylerine dair önemli çıktılar sağlamaktadır. Türkiye PIRLS'e ilk defa 2001 yılında katılmış, 20 yıllık bir aradan sonra alınan kararla 2021 yılında tekrar dâhil olmuştur. Bu uygulamaya ilişkin sonuçlar 2022 yılının sonunda açıklanacaktır. Türkiye'nin PIRLS'e katılmasının bir diğer avantajı her iki çalışmaya katılan ülkelerin sonuçlarının (PIRLS ve TIMSS'ten elde edilen sonuçlar) ilişkilendirilebilmesidir.

## OECD SOSYAL VE DUYGUSAL BECERİLER ARAŞTIRMASI

Tüm dünyada sosyal ve duygusal becerilerin önemi ile birlikte eğitim sistemlerindeki ağırlığının artmasıyla beraber 2017 yılında OECD tamamen sosyal ve duygusal beceriler odaklı yeni bir araştırma başlatmıştır. Türkiye de İstanbul'la bu çalışmaya katılmıştır.

sosyal ve duygusal beceriler dikkate alınıyor ki bu diğer tüm izleme araştırmalarından en önemli farkıdır. İkincisi 10 ve 15 yaş grupları ayrı ayrı ele alınmaktadır ki duyuşsal özellikler, sosyal ve duygusal özellikler yaşa bağlı olarak, özellikle de ergenlik dönemine bağlı olarak önemli değişimler gösterebilmektedir. Ergenlik sürecini de içerecek şekilde iki ayrı yaş düzeyinin dikkate alınması bu becerilerin nasıl değiştiği hakkında önemli çıkarımlar sağlamaktadır. Üçüncüsü ölçümler, dünyanın farklı bölgelerinde kabul gören beş faktörlü kurama dayalı olarak (Big five modeli) yapılmaktadır.

Bu beş faktörlü model aşağıda verilmiştir.



Beş faktörlü modeli oluşturan ana alanlar ve bu alanları oluşturan alanlar: açık fikirlilik (yaratıcılık, hoşgörü, merak), iş birliği (empati, iş birliği, güven), duygu düzenleme (iyimserlik, strese dayanıklılık, duygu düzenleme), görev performansı (sebat, öz denetim, sorumluluk) ve başkalarıyla etkileşimde olma (sosyallik, girişkenlik, enerji) olarak tanımlanmıştır. Her bir alanı oluşturan alanlar ilişkide ve temelde hepsi kendi içerisinde sosyal dayanağı olan sosyal beceri olarak tanımlanan özelliklerdir. Bir de birleşik beceriler diye tanımlanan ve bu beşliye dâhil olmayan ama onunla ilişkili olup sosyal beceriler açısından önemli görülen iki ayrı beceri ise başarıya motivasyonu ve öz yeterlik algısıdır.

