



## MODÜL 2

### ÖLÇME VE DEĞERLENDİRME (ÖZET)

**EĞİTİM SİSTEMİNDE ÖLÇME VE DEĞERLENDİRMENİN YERİ:** Bir bütüne sistemdiyebilmek için girdi, süreç ve çıktı öğelerinin olması gerekir. Eğitim de bir sistemdir çünkü birincisi gerçekleştirmek istediği hedefler bulunmaktadır. Hedefler şu biçimde tanımlanabilir:

1. **Uzak hedefler:** Devletlerin hedefleridir,
2. **Genel hedefler:** Kurumların (MEB, , okul yönetimleri vb.) hedefleri vb.
3. **Özel hedefler:** Dersin hedefleridir(kazanımlar)

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 vb.

**2. Süreç:** Dersin hedeflerinin öğrenciye kazandırıldığı eğitim durumlarını kapsar.

**3. Çıktı:** Öğrencilerin sahip olduğu niteliklerdir. 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.. Çıktılar hedeflerle tutarlı ise eğitim sistemi doğru bir şekilde işliyor demektir. Eğer çıktılar hedeflerle tutarlı değilse eğitim sisteminde bir şeyler aksıyor demektir.

### YİRMİ BİRİNCİ YÜZYIL DEĞİŞEN EĞİTİM PARADİGMASI

Sanayi Paradigması	Bilgi ve İletişim Teknolojisi Paradigması
Enerji yoğun	Bilgi yoğun
Çizim bürolarında yapılan tasarım ve mühendislik	Bilgisayar destekli tasarım
Ardışık tasarım ve üretim	Eş zamanlı mühendislik
Standart tasarım	Sipariş üzerine tasarım
Tahsis edilmiş tesis donanım	Esnek üretim sistemleri
Otomasyon	Sistemasyon
Tek firma	Ağ yapı
Hiyerarşik yapılar	Düz yapılar
Bölünmüş	Bütünleşik
Servisi olan üretim	Ürünleri olan servis
Merkezî yapı	Dağılmış yapı
Özgün beceri	Çoklu beceri
Planlama	Vizyon

Tablo 1. Değişen üretim paradigması

Sanayi Paradigması	Bilgi ve İletişim Teknolojisi Paradigması
Sınıflarda eğitim	Bireysel araştırma
Pasif öğrenme	Yaparak-yaşayarak öğrenme
Bireysel çalışma	Grupla çalışma
Her şeyi bilen öğretmen	Rehberlik eden öğretmen
Sabit içerik	Esnek içerik
Homojen	Heterojen

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

Eğitimde denetleme (kontrol) ölçme ve değerlendirme ile olanaklıdır ancak değişen hedeflerde bireylere kazandırılacak olan yeni davranışların ölçülmesi ve değerlendirilmesinde de doğal olarak bir paradigma değişimi söz konusudur.

## ÖLÇME, ÖLÇÜT VE DEĞERLENDİRMENİN KAVRAMSAL TEMELLERİ:

**Ölçme:** Öğrencinin belirli bir özelliğini<sup>3</sup> gözleyerek<sup>4</sup> o özelliğe sayı, sembol ya da sıfat/kategori adı<sup>5</sup> verme işlemidir.

Örneğin Can'ın genel yetenek testi sonucu, zekâ bölümünün 132 olduğu belirlendi. Zeynep yapılan test sonucunda 1. oldu.

- Ölçmede gözlenen özelliğin bir betimlemesi yapılır. Bu nedenle tek başına bir anlam ifade etmez. Ölçmenin anlam ifade etmesi değerlendirme ile olanaklıdır.
- Ölçmenin farktan doğduğu ifade edilebilir.
- Eğitimde akademik başarı dışında yetenek, ilgi, tutum, özel gereksinim gereken alanlar vb. özellikler de ölçmeye konu olur.
- Ölçme yapılmadan sisteme yönelik girdi, süreç ve çıktılar kontrol altına alınamaz.

**Ölçüt:** Ölçülen özellik hakkında karar alabilmek / yargıya varabilmek / değerlendirme yapabilmek için dayanak alınan referans noktası<sup>6</sup> ya da referans aralığıdır.

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

**Değerlendirmenin basamakları :** 1. Ölçme ⇒ Ölçme sonucu (ölçüm), 2. Ölçüt ve 3. Karardır.

- Zeynep dersten geçmiştir / kalmıştır.
- Mehmet öğretmen olarak atanmıştır / atanamamıştır.
- Hakan burs almaya hak kazanmıştır / burs alamamıştır.

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

## ÖLÇME TÜRLERİ

**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, öğrencilerin cinsiyetleri, boyları, kalem tutma becerisi, pas verme becerisi vb. özellikler doğrudan gözlenebilir.

**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, Eğitimde ölçmeye konu olan özelliklerin pek azı doğrudan ölçülebilir.

**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 vb. özellikler türetilmiş ölçmeye örnek olarak verilebilir.

**ÖLÇÜT TÜRLERİ:** Eğitimde ölçüt, mutlak ve bağıl olmak üzere ikiye ayrılır.

**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, doktora başvurusu için ALES'ten en az 60 almak gerekir,

**Bağıl Ölçüt (Norm Referanslı Ölçüt):** Eğer ölçüt ölçme işleminin yapıldığı grubun belirli bir özelliğine dayalı olarak belirleniyorsa "bağıl"dır. Bağıl ölçüt ancak ölçme işlemi sonrasında belirlenebilir. **Örneğin** aritmetik ortalamaya / ortancaya / moda denk ve üzerinde puan alan geçer.

- ➡ Eğer değerlendirme mutlak ölçüte göre yapılıyorsa "**mutlak değerlendirme**", bağıl ölçüte göre yapılıyorsa "**bağıl değerlendirme**" adını alır. **Örneğin**
  - Selim ..... fen lisesine yerleştirilmiştir. → Bağıl değerlendirme
  - Güzin lisans programları için tercih yapabilecektir. → Mutlak değerlendirme
- ➡ Eğitsel kararların verilmesinde **mutlak ve bağıl ölçütler bir arada da** kullanılabilir. **Örneğin**
  - Yükseköğretim Kurumları Sınavı
  - KPSS Ortaöğretim Memur Atamaları
  - Ortalama 40-60 ise mutlak ya da bağıl değerlendirme yapılır; ortalama 60'ın üzerindeyse mutlak değerlendirme, 40'ın altındaysa bağıl değerlendirme zorunludur.
- ➡ Öğrenci ile ilgili önemli / hayati kararlar verileceği zaman **mutlak değerlendirme** yapılması gerekir. **Örneğin** tek ders sınavları vb.
- ➡ Başvuran kişi sayısının çok, alınacak kişi sayısının az olduğu (arz-talep dengesizliği olan) sınavlarda yerleştirme / atama vb. **bağıl değerlendirme** ile yapılmak zorundadır. **Örneğin** LGS, YKS, KPSS, TUS vb.
- ➡ **Muafiyet sınavlarında mutlak değerlendirme** yapılması gerekir. **Örneğin** yabancı dil muafiyet sınavları vb.

## AMACA GÖRE DEĞERLENDİRME TÜRLERİ

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

**Sınıf içi ölçme ve değerlendirme:** Amaç öğrencinin ders bağlamındaki hazır bulunuşluluk düzeyini belirlemektir. Amaç not vermek değildir.

**Okul / kurum geneli ölçme ve değerlendirme:** Yine öğretim sürecinin başında, öğrenciyi tanımak ve onu uygun olan programa / kura / sınıfa yerleştirmek amacıyla yapılır. 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 belirlemek amacıyla yapılır.
- Biçimlendirme-yetiştirme amacıyla yapılacak değerlendirmede kullanılan testlere alanda "**izleme testi**" ya da "**tarama testi**" denir. (quiz), ünite tarama testleri buna örnektir.

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

- ✚ Belirli bir öğretim sürecinin sonunda, not vermek<sup>8</sup> amacıyla yapılan değerlendirmedir.
- ✚ Öğrencinin belirli bir öğretim süreci sonundaki durum tespittir.
- ✚ Bu amaçla uygulanan testlere "**alanda erişim testi**" denir. Bitirme sınavları, sertifika sınavları, üniversitedeki vize ve finaller, örnek olarak verilebilir.  
İlk üç değerlendirme türü öğrenci başarısına odaklıdır,



**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.

**Mesleki rehberlik:** öğrencilerin kendi yetenekleri, kişilik özellikleri, ilgileri vb. ile özelliklerini belirlemek amacıyla standart testlerden yararlanılmaktadır.

**5. Program Değerlendirme:** Eğitim programları bileşenlerinin yapısı ve niteliğine ilişkin olarak sistematik bilgi toplama ve değerlendirme sürecidir.

## BİR ÖLÇME ARACINDA BULUNMASI GEREKEN PSİKOMETRİK NİTELİKLER

► Psikolojik özelliklerin ölçülmesinde genellikle ölçme araçları kullanılır ve bu araçlardan elde edilen puanların hatasız ya da az hatalı olması ve bu araçların belirli psikometrik niteliklere sahip olması gerekir. Bunlar:

**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.

► Bir ölçme aracından elde edilen puanların hatasız ya da az hatalı olması **geçerli ve güvenilir ölçme yapmanın tek koşuludur.**

**HATA:** Ölçmede gözlenen bir özelliğin gerçek değeri (sayı, sembol ya da sıfat) bulunmak istenir. Ancak ölçmeye karışan çeşitli hatalar nedeniyle gerçek değer (puan) ölçme yoluyla doğrudan elde edilemez. Psikometride gerçek puan kuramı olarak ele alınan bu kavram basit bireşitlik ile gösterilir<sup>9</sup>. Bu çerçevede hatalar; kaynağı, yönü ve miktarı göz önünde tutularak sabit, sistematik ve tesadüfi olmak üzere üç türde incelenebilir:

**A. Sabit Hata:** Miktarı ölçmeden ölçmeye değişmeyen, diğer bir deyişle her ölçme işlemine aynı miktarda karışan hatalardır. Örneğin Bir öğretmen, sınavında herkese 10 puan fazla veriyorsa yine karışan hata sabit olacaktır.

**B. Sistematik 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.

Gerek sistematik gerekse sabit hatalarda ölçme sonuçlarına karışan hata miktarı, yönü ve kaynağı bellidir. Bu nedenle bu tür hataların ölçme sonuçlarına karışmasını engellemek ya da bu tür hataları düzeltmek görece daha kolaydır.

### C. 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. Tesadüfi hataların genellikle dört kaynağı olduğu kabul edilir. Bunlar:

**1. Ölçme işlemini yapan kişiden / öğretmenden kaynaklanan hata:** Öğretmenin test etme sürecine ilişkin davranışlarını iki aşamada ele alabiliriz:

**Ölçme işlemi sürecinde:** Test katılımcısının dikkatini dağıtacak davranışlar vb.

**Ölçme işlemi sonrasında:** Puanlamadaki dikkat ve titizliğin zamandan zamana değişmesi, yorgunluk, öncelik-sonralık yanılgısı, maddi hata vb.

## 2. Ölçme aracından kaynaklanan hata

- Ölçme araçları hazırlanırken maddelerin iyi ifade edilmemesi, test katılımcılarının yanlış anlamalarına ve dolayısıyla hataya neden olacaktır.
- Ölçtüğü özellik ve kapsam bakımından homojen (benzeşik) maddelerden oluşan bir araç, heterojen (ayrışık) maddelerden oluşan bir testten daha güvenilirdir.
- Ölçme aracının yapısıyla ilgili diğer bir konu ise **aracın uzunluğudur**.
- Ölçme aracıyla ilgili diğer bir faktör ise bilişsel özellikleri ölçen testlerde eğer **seçenek verilmişse** şans başarısı<sup>10</sup> karışma olasılığıdır.

**3. Bireyden / öğrenciden kaynaklanan hata:** Bireylerin ölçme işlemi sürecinde içinde bulundukları fiziksel, fizyolojik ve psikolojik durumlar test puanına etki eder. Uykusuzluk, açlık, hastalık, ağrı, motivasyon eksikliği, kaygı vb. faktörler

**4. Fiziksel ortamdan kaynaklanan hata:** Sıcaklık, ışık, ses, koku, görsel vb. etmenler hata miktarını artırabilir.

## GEÇERLİLİK, GÜVENİLİRLİK VE HATA İLİŞKİSİ

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



Şekil 1: Geçerlilik, güvenilirlik ve hata ilişkisi

Bir testin güvenilir olması onun geçerli olacağı anlamına gelmez ancak bir test geçerli ise büyük olasılıkla güvenilirdir.

Ölçme araçlarının psikometrik nitelikleri şu durumlara göre değişebilir:

- Ölçme amacının değişmesi,
- Uygulama grubunun değişmesi,
- Dilin eskimesi / değişmesi,
- Maddelerde ve/veya alt ölçeklerde yapılandırılmış değişiklikler,
- Farklı kültürler,
- Kuramsal bilgi birikiminde değişiklikler.

Bir testin şöhretli olması onun geçerli olduğu anlamına gelmez. Modern test kuramına göre geliştirilen ölçekler eleştiriye açıktır. Böylece testlerin niteliği ve test bilgisi gelişir.

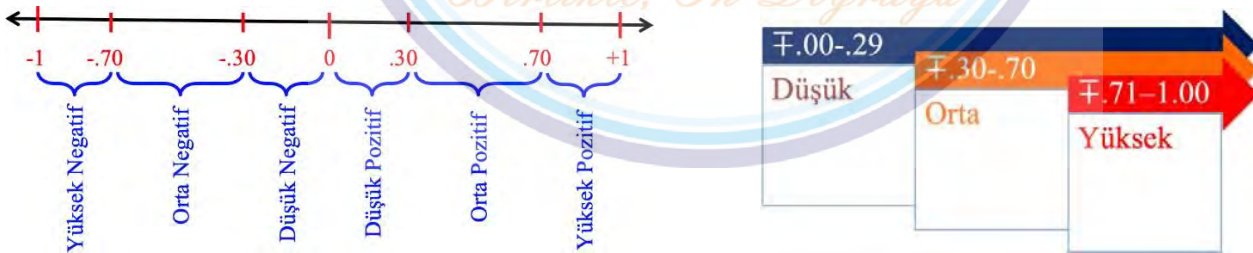
**KORELASYON:** Korelasyon (co-relation), en az iki değişken arasında karşılıklı bir ilişki bulunup bulunmadığı, eğer ilişki varsa bu ilişkinin yönü ve miktarı hakkında bilgi veren istatistik bir tekniktir. "r" ile sembolize edilir. Örneğin

- Öğrencilerin derse ilişkin tutumları ile ders başarıları arasında bir ilişki var mıdır?
- Saç uzunluğu ile zekâ arasında bir ilişki var mıdır?

- ❏ Korelasyon -1 ile 1 arasında değer alır..
- ❏ **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 edilmesibeklenir.
- ❏ **±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.
- ❏ Korelasyon katsayısı hakkında. Yön, negatif ya da pozitif olarak değerlendirilirken miktar, kabaca düşük, orta ya da yüksek olarak nitelendirilir. Miktar için kesin sınırlar olmamakla birlikte alanyazında genellikle kabul gören aralıklar Şekil 3'te gösterilmiştir
- ❏ Korelasyon **mutlak değer** olarak değerlendirilmelidir. Bir korelasyon katsayısının negatif ya da pozitif olması büyüklük-küçüklük belirtmez, yön bildirir.
- ❏ Korelasyon katsayısı ile neden-sonuç ilişkisi kurulamaz.



Şekil 2: Korelasyonun temel kavramları



Şekil 3: Korelasyonun miktar

## GEÇERLİLİK SORGULAMA YÖNTEMLERİ





## A. KAPSAM GEÇERLİLİĞİ:

- Kapsam geçerliliği özellikle başarı testlerinde aranan bir geçerlilik sorgulamasıdır.
- Kapsam geçerliliği bir testin ölçülmek istenen davranışları ne derece kapsadığıyla ilgilidir.

Bir testin kapsam geçerliliğinin yüksek olduğunun söylenebilmesi için:

1. Testteki soruların / maddelerin ölçülecek özellikler evrenini (konu kapsamını / içeriği) yeterli ve dengeli bir biçimde ölçüyor olması ve
2. Her bir sorunun / maddenin ölçmek istediği özelliği doğrudan ölçmesi, diğer bir deyişle kazanımla doğrudan ilgili olması gerekir.

## Kapsam geçerliliği sorgulama yöntemleri:

- **Mantıksal / rasyonel yöntemler:** Bu yöntemler belirtke tablosu hazırlanması ve uzman görüşüne başvurulmasıdır.
- **Belirtke tablosunun hazırlanması:** Öğretmen satırda davranışları, sütunda hedefleri yazar; belirlediği madde sayısı doğrultusunda davranış ve hedefleri yeterli ve dengeli bir biçimde temsil eden alanları seçer.
- **Uzman görüşüne başvurulması:** Uzmana belirtke tablosu ve maddeler sunulur, uzmandan soruların / maddelerin kazanımlarla doğrudan ilgili olma durumunu değerlendirmesi istenir
- **İstatistiksel yöntemler:** Uzmanlardan alınan dönütler betimsel / muhakemeye dayalı bir yolla çözümlenebilir ya da uzmanların "uygundur / uygun değildir" vb. biçimde değerlendirme yapması istenebilir

Tablo 3. Belirtke tablosu örneği

Konular	Hedefler						Toplam
	Bilgi	Kavrama	Uygulama	Analiz	Sentez	Değerlendirme	
1. Ölçme ve değerlendirmenin temel kavramları							%43.3
1.a. Ölçme ve değerlendirmenin eğitim sistemi içindeki yeri		1				1	2
1.b. Eğitimde paradigma değişimi ve bu değişimin ölçme ve değerlendirme uygulamalarına yansımaları	1			2			3
1.c. Ölçme, değerlendirme ve ölçütün kavramsal temelleri		2					2
1.d. Ölçme türleri		1					1
1.e. Ölçüt türleri	1					1	2
1.f. Değerlendirmenin amaca göre türleri		1			1	1	3
2. Bir ölçme aracıda bulunması gereken psikometrik nitelikler							%56.7
2.a. Geçerlilik, güvenirlik ve kullanılabilirliğin kavramsal temelleri		1					1
2.b. Ölçmede hata kavramı	1						1
2.c. Hata türleri		1		1			2
2.d. Geçerlilik ve güvenirlik ile hata ilişkisi						1	1
2.e. Geçerlilik ve güvenirlikle ilgili temel kavramlar		1	2				3
2.f. Geçerlilik sorgulama yöntemleri		2		1			3
2.g. Güvenirlik kestirim yöntemleri	1	1	1			1	4
2.h. Geçerliliği ve güvenirliği artırma yolları		1		1			2
TOPLAM	4	12	4	5	1	5	30

## B. 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.

Ölçüt dayanaklı geçerlik 2'ye ayrılır:

1. **Yordama Geçerliliği:**
2. **Zamandaş Geçerlilik:**

## 1. Yordama Geçerliliği:

- Ölçme araçlarının çoğunda bireylerin gelecekteki davranışlarının kestirilmesi söz konusudur. Yordama geçerliliği, ölçme aracının bu kestirim işini ne ölçüde doğru ve isabetli yaptığı ile ilgili bir sorgulamadır.
- Özellikle iki amaçla uygulanan ölçme araçlarının yordama geçerliliğinin yüksek olması istenir. Bunlar **seçme ya da yönlendirme amaçlı kullanılan testler**dir. Seçme amaçlı testlere **YKS, KPSS, LGS** vb. araçlar; diğer taraftan yönlendirme amaçlı testlere ise alan seçiminde ve/veya bir üst öğretim kurumuna yönlendirmek için uygulanan **yetenek testleri, ilgi envanterleri, mesleki kişilik envanterleri** vb. örnek olarak gösterilebilir.
- Yordama geçerliliğinde ölçme aracından elde edilen puanlar, ölçme aracının tahmin ettiği puanla (ölçüt puan) karşılaştırılır ve tahminin ne ölçüde doğru olduğu belirlenmeye çalışılır. Yordama geçerliliğinde ölçüt puan gelecekte belli olacağından beklemek gerekmektedir. **Yordama geçerliliğinde geçerlilik katsayısının 0 ile 1 arasında değişmesi beklenir.**

	YKS Puanı	Akademik Ortalama
Tarik	438.636	3.54
Esra	348.545	2.45
Begüm	456.344	3.39
Ozan	399.987	2.23
Defne	382.654	2.34
.....	480.334	3.45
.....	287.987	1.67
.....	381.954	3.88
.....	250.897	3.91
Umut	343.019	3.26

Ölçüt gelecekte belli olacağından beklemek gerekmektedir.

Korelasyon (r) 1'e yakın

Geçerlilik Katsayısı

Şekil 5: Bir yordama geçerliliği çalışma örneği

**Yordama geçerliliğinde en zor ve önemli nokta ölçütün doğru bir biçimde belirlenmesidir.**

Ölçütün belirlenmesinde şu noktalar dikkate alınmalıdır: Ölçüt(ün);

1. Ölçme aracının yordamaya çalıştığı değişkenle doğrudan ilgili olmalı,
2. Kararlı olmalı, günden güne değişmemelidir.
3. Bireylerin özelliğini gerçekten yansıtan nesnel ve güvenilir bir ölçüt olmalıdır.
4. Elde edilmesi kolay ve ekonomik olmalıdır.

## 2. Zamandaş Geçerlilik:

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.

	Benlik Algısı Ölçeği Puanları	Ölçüt Puan Öz-saygı Ölçeği Puanları
İrem	54	32
Metin	98	54
Erdal	128	98
Dilek	89	65
Çağdaş	101	54
.....	87	56
.....	98	58
.....	67	65
.....	122	80
Sedef	79	88

Daha önce geliştirilmiş ve geçerliliğinin yüksek olduğu bilinen araç

Korelasyon (r) 1'e yakın

Geçerlilik Katsayısı

Şekil 6: Bir zamandaş geçerlilik çalışma örneği





Şekil 7: Geçerlilik katsayısı

**Zamandaş geçerlilikte de en zor ve önemli nokta ölçütün doğru bir biçimde belirlenmesidir.**

Ölçütün belirlenmesinde şu noktalar dikkate alınmalıdır: Ölçüt(ün);

1. Ölçme aracının ölçmeye yöneldiği özelliklikle doğrudan ilişkili olmalıdır.
2. Geçerliliği yüksek olmalıdır.

### C. YAPI GEÇERLİLİĞİ:

Yapı geçerliliği, bir testin dayandığı kuramsal temelleri ne derece iyi örneklediğiyle ilgilidir. Yapı geçerliliği bir yandan testin ölçtüğü niteliklerin neler olduğunu araştırma, diğer yandan testi alan kişilerin elde ettikleri puanların ne anlama geldiğini açıklama çabalarıyla ilgilidir. Örneğin bir kişi, geleneksel aile biçimi ile çocuk yetiştirme biçimi arasındaki ilişkiyi incelemek amacıyla bir ölçek geliştirip bu ölçeğin yapı geçerliliğini ortaya koymak istediğinde, geleneksel aile yapısı ve çocuk yetiştirme kavramlarının ne anlamlara geldiğini, ölçme aracındaki maddelerin bu anlamlara uygunluğunu araştırarak yapı geçerliliği konusunda karar verebilir.

### GÜVENİLİRLİK

- Güvenilir bir ölçme aracı, aynı özelliklikle 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.
- Geleneksel olarak hataların varlığı gözlenen puanın gerçek puandan yüksek ya da düşük olmasına neden olur.

Yarışlara hazırlanan bir koşucu, bir mesafeyi farklı zamanlarda 23.7, 24.0, 24.2,... 25.1, 25.2 saniyelerde koşmuş olsun. Bu ölçümlerin ortalaması 24.7 ise gerçek puanı 24.7'dir. Koşucu bu puana daha önce 23.7 saniyede koştuğu ölçümü göstererek itiraz edebilir. Bu durumda koşucuya 23.7 ve 25.2 saniyede koştuğu durumlara birçok faktörün etkisinin olabileceği ve 23.7 saniyede tekrar koşmasının belki de hiç olanaklı olmayacağını, gerçek puanının 24.7 olduğunu söyleyebiliriz. Bu söylem **gerçek puan kuramına** dayanır.

Ölçme işleminde hata miktarı arttıkça gözlenen puanın gerçek puana olan farkı artmaya başlar. Ölçmenin standart hatası

$$S_h = s\sqrt{1 - r_x}$$

formülü ile bulunur.

$S_h$ = Ölçmenin standart hatası= Standart sapma

$r_x$ = Güvenilirlik katsayısı

Bir ölçme işleminde standart hatanın düşük olması, formülden de anlaşılacağı üzere, **güvenilirlik katsayısının yüksek, standart sapmanın görece düşük olmasına** bağlıdır. Örnek: Bir ölçme işleminde  $s=8$  ve  $r_x=.75$  ise  $S_h=8\sqrt{1-.75}=4$ . Bu sonuç bireylerin puanlarına  $\pm 4$  puan hata karıştığı anlamına mı gelir?

- Gerçek puanın tahmin edilmesinde genellikle üç olasılık değeri kullanılır. Bunlar yaklaşık **%68, %95 ve %99'dur**. %68 olasılık için bireyin puanından bir standart hata çıkarılır ve puanına eklenirken %95 olasılık için bireyin puanından iki standart hata çıkarılır ve puanına eklenir, son olarak %99 olasılık için bireyin puanından üç standart hata çıkarılır ve puanına eklenir.

$$\%68 \text{ olasılık için } X \pm 1 \times S_h$$

$$\%95 \text{ olasılık için } X \pm 2 \times S_h$$

$$\%99 \text{ olasılık için } X \pm 3 \times S_h$$

Dolayısıyla bir kişi 100 üstünden 50 puan almış ve  $S_h=4$  ise

- %68 olasılık:  $X \pm 1 \times S_h = 50 \pm 1 \times 4 =$  Bireyin gerçek puanı %68 olasılıkla 46-54 arasında değişir.
- %95 olasılık için  $X \pm 2 \times S_h = 50 \pm 2 \times 4 =$  Bireyin gerçek puanı %95 olasılıkla 42-58 arasında değişir.
- %99 olasılık için  $X \pm 3 \times S_h = 50 \pm 3 \times 4 =$  Bireyin gerçek puanı %99 olasılıkla 38-62 arasında değişir.

- Ölçmenin standart hatası yükseldikçe ölçme işlemi bireylere rastgele puan vermekten farklı olmayacaktır. Bu nedenle ölçme işlemlerinde güvenilirliğin kestirilmesi çok önemlidir.
- Güvenilirlik katsayısı **0 ile 1 arasında** değişir. Genel yetenek gibi bilişsel özellikleri ölçen testlerde güvenilirlik katsayısının 0.90 ve üzerinde olması arzu edilir.



Şekil 8: Güvenilirlik katsayısı



Şekil 9: Güvenilirlik kestirim yöntemleri

**A. 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.

	Anksiyete Ölçeği Puanları $X_1$	Anksiyete Ölçeği Puanları $X_2$
Ceren	54	50
Filiz	98	110
Ayдын	128	120
Ece	89	80
Deniz	101	101
.....	87	85
.....	98	106
.....	67	59
.....	122	119
Dursun	79	82

**Kararlılık Katsayısı**

**Korelasyon (r) 1'e yakın**

Şekil 10: Test-tekrar test yöntemi çalışma örneği

- Bir ölçme aracının güvenilirliğine ilişkin kanıt toplamak isteyen bir kişi **kararlılığın yanı sıra tutarlılığa ve duyarlılığa ilişkin sorgulamaları** yapmak durumundadır.
- Bu yöntem, **ölçülen özelliğin kararlı olduğu durumlarda** uygulanması gereken bir yöntemdir. Testin ölçtüğü özellik **sürekli değişkenlik gösteriyorsa** bu yöntemle testin güvenilirliği hesaplanmamalıdır. Örneğin genel zihin yetenekleri, kişilik testleri, ilgi envanterleri, vb. gibi testlerin kararlılık bağlamında güvenilirlikleri bu yöntemle hesaplanabilir.
- Bu yolla test güvenilirliğini kestirmede karşılaşılan önemli bir sorun, **testin iki uygulaması arasındaki zaman aralığının ne kadar olması gerektiğidir. Birinci uygulamada maddeleri hatırlamayacak kadar uzun, ölçülen özellikte köklü değişimler oluşmayacak kadar kısa olmalıdır**
- Aradaki zamanın belirlenmesinde dikkat edilmesi gereken bir başka nokta, **özelliğin değişim hızıdır.** (genel yetenek düzeylerinin kısa bir zamanda değişmesi gibi)
- Ayrıca özelliklerin değişimi yaşla da ilişkilidir.
- Test-tekrar test yöntemi, **başarı testleri için çok tercih edilmemektedir.** Çünkü kısa vadede başarı kolay değişebilen, kararsızlık gösterebilen bir özelliktir.
- Güvenilirliğin bir boyutu da testin tutarlılığıdır. (**iç tutarlılık katsayısı**)

## **B. 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 başvurulan yöntemler:  
a) ilk yarı ve son yarı      b) tek ve çift ve      c) rastlantısaldır.

Testteki

- maddeler basitten zora doğru sıralanmış ise
- maddeler konu içeriklerine ya da faktörlere göre kümelenerek yerleştirilmiş ise
- madde sayısı çok fazla ise İlk ve son yarı yöntemi için uygun değildir.

- ☑ Testin tümüne ilişkin bir güvenilirlik katsayısı **Spearman-Brown formülü aracılığıyla** hesaplanır. Elde edilen katsayı testin tamamına ilişkin **iç tutarlılık** bağlamındaki güvenilirlik katsayısı olarak kabul edilir.



	<u>Türkçe Testi</u>	<u>X<sub>1</sub> YARI</u>	<u>X<sub>2</sub> YARI</u>
Deniz	48	20	28
Şebnem	98	50	48
Ufuk	66	32	34
Işıl	88	44	44
Sinem	32	20	12
.....	70	30	40
.....	62	28	34
.....	40	20	20
.....	60	28	32
Sinan	66	36	30

Testin yarısına ilişkin tutarlılığı verir. Bu korelasyon testin tamamına nasıl genellenir?

Korelasyon (r)

Spearman-Brown Eşitliği

İç tutarlılık katsayısı

Şekil 11: Test yarılama yöntemi çalışma örneği

**C. KUDER-RICHARDSON 20 VE 21 YÖNTEMLERİ:** 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 seçeneğinden birinin yapıldığı durumlardır.
- ☑ Eğer testteki maddeler farklı ağırlıklarla puanlanmışsa ya da test puanları şans başarısı için düzeltilmişse bu formüller kullanılmaz.

**D. 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.

### GÜVENİLİRLİĞİ VE GEÇERLİLİĞİ ARTIRMA YOLLARI

1. Bir testteki madde sayısı arttıkça birimler küçüldüğünden duyarlılık artar, bu nedenle hata miktarı azalacağından güvenilirlik artar.
2. Bir ölçme işleminde genel bir ilke olarak **puanlayıcı sayısı arttıkça** güvenilirlik artar.
3. **Puanlama nesnelliği**, güvenilirliği büyük ölçüde etkiler. Bir testin farklı kişiler tarafından puanlanması ya da aynı kişinin farklı zamanlarda verdiği puanlar arasındaki tutarlılığa **puanlama güvenilirliği** adı verilir.
4. Testteki maddeler açısından **homojenlik** arttıkça güvenilirliğin artması, diğer taraftan **heterojenlik** arttıkça güvenilirliğin düşmesi beklenen bir durumdur.
5. Veri elde edilecek **grubun (örneklem)** büyüklüğü arttıkça güvenilirlik artar.
6. Maksimum performansı ölçen testlerde ortalama güçlüğü yaklaştıkça ( $\phi = .50$ ) grup heterojenleşir ve dolayısıyla güvenilirlik artar.
7. Maddelerin dil bilgisi kurallarına uygun, açık ve anlaşılır yazılması güvenilirliği artıran etmenlerden biridir.
8. Öğrencilere test uygulaması öncesinde ve gerekliyse süreçte **yönerge verilmesi** güvenilirliği artırır.
9. **Fiziksel ortamın ses, sıcaklık, ışık, koku, dikkat dağıtıcı uyarıcılar** vb. etmenler açısından uygun hâle getirilmesi güvenilirliği artırır.
10. **Testin süresinin yeterli olması** güvenilirliği artırır.
11. Uygulamada **bireylerin dikkatini dağıtacak davranışlardan kaçınılması** güvenilirliği artırır.
12. Maddelerin teste **düzgün yerleştirilmesi**, okumayı güçleştirecek bir unsurun bulunmaması; test katılımcısının yaşına, gelişim düzeyine **uygun bir punto büyüklüğü** kullanılması; baskı hataları bulunmaması vb. etmenler güvenilirliği artırır.

**NOT: Güvenilirliği artıran faktörler geçerliliği de artırır.** Ancak geçerlilik yalnızca tesadüfi hatalardan değil, aynı zamanda sabit ve sistematik hatalardan da etkilenir. O nedenle öğretmenin tesadüfi hataların yanı sıra sabit ve sistematik hatalardan arındırmaya yönelik alacağı tüm tedbirler geçerliliği de artıracaktır.

### 3. 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

#### TEST GELİŞTİRME:

Öğretmenler açısından sınıf içi ölçme ve değerlendirmelerde uygulanması daha olanaklı bir test planı şu şekilde yapılabilir:

1. **Sınavın amacının** belirlenmesi (puanların hangi amaçlarla kullanılacağı)
2. Sınavda **yoklanacak davranışların** belirlenmesi
3. Sınavın kapsayacağı **konuların listelenmesi**
4. Okulun takviminin ve sonuçların kullanılacağı zamanın dikkate alındığı **bir sınav gününün belirlenmesi**
5. Bir ders saatine uygulanabilecek uzunlukta bir sınav süresinin belirlenmesi,
6. **Belirtke tablosunun hazırlanması**
7. Soru/madde türlerinin belirlenmesi. Soru/madde türleri sınavda kullanılacak davranışlara göre belirlenir. Bir sınavda tek bir soru/madde türü kullanılabileceği gibi birden fazla da kullanılabilir
8. **Soru/madde sayısının** belirlenmesi
9. **Sınav süresinin** belirlenmesi
10. Sınavda kullanılacak soruların/maddelerin "ortalama güçlüğü" ve "güçlük dağılımının" belirlenmesi
11. Soruların/maddelerin yazımında, redaksiyonunda ve teste alınacak soruların/maddelerin seçiminde izlenecek yolun belirlenmesi
12. **Cevap anahtarının ve puanlama yönteminin** belirlenmesi
13. Ölçme aracını yazma ve çoğaltma yönteminin belirlenmesi
14. Sınavın uygulanma kurallarının belirlenmesi (yönerge yazılması)
15. Sınavın uygulanması
16. Test ve madde istatistiklerinin hesaplanması

#### BAŞARININ ÖLÇÜLMESİNDE YÖNTEMLER



Şekil 12: Ölçme araç ve yöntemleri



Şekil 13: Geleneksel yöntemler (Madde / soru türleri)

## GELENEKSEL ÖLÇME YÖNTEMLERİ

### 1. DOĞRU-YANLIŞ MADDELERİ

- Çoktan seçmeli madde türünden sonra **en çok kullanılan ikinci madde** tipidir. Genellikle öğrencinin bilimsel gerçekleri, tarihî olayları, kesin yargıları tanıması ve hatırlaması; olguları ve gerçekleri kişisel yargı ve görüşlerden ayırt etmesi ölçülmek istendiğinde kullanışlıdır.
- Avantajları:** a. Hazırlaması, uygulaması ve puanlaması kolay, diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistematiğe hata karışma olasılığı pek yoktur.
- Dezavantajları:** a. Şans başarısı olasılığı yüksektir. b. İyi yazılmazsa genellikle üst düzey bilişsel özellikleri ölçemez. c. Öğrenme eksiklerini belirleyemez. d. Yanlış üzerinden öğretim yapılmaz.

#### Doğru-Yanlış Maddeleri Yazım Kuralları:

- Her madde **tek ve belirli bir fikri belirtmelidir**. Özellikle aynı maddede biri doğru, öteki yanlış iki fikir ifade etmekten kaçınılmalıdır.
- Doğru yanlış maddesi, **kesinlikle doğru ya da kesinlikle yanlış olmalıdır**. Maddenin doğruluğu ya da yanlışlığı, başka bir açıklamaya gerek kalmadan belirlenebilmelidir.
- Bir maddenin yanlışlığı önemsiz bir ayrıntıda ya da aldatıcı bir noktada olmamalıdır. Bir yargı temelden yanlış olmalıdır.
- İyi kurgulandığı takdirde doğru-yanlış maddeleri ile üst düzey öğrenmeler de yoklanabilir.
- Doğru-yanlış maddeleri aynı kapsamdan geliyorsa **ortak bir soru kökü altında** birleştirilebilir.
- İpucu vermekten kaçınmak gerekir**. Özellikle "asla, daima, hiçbiri, tümü" vb. ifadelerin yanlış olma olasılığı yüksektir. Öte yandan "genellikle, çoğunlukla, bazen, zaman zaman, sıklıkla" vb. ifadelerin de doğru olma olasılığı yine yüksektir.
- Üst düzey becerileri yoklamak için tablo, grafik, harita ya da okuma parçası gibi bir öncülün kullanılması önerilir.



## 2. EŞLEŞTİRME MADDELERİ:

**Yönerge / Açıklama**

Yönerge: Aşağıdaki "A" sütununda Türkiye Cumhuriyeti tarihi ile ilgili bazı önemli olaylar, "B" sütununda ise bu olayların geçtiği yıllar verilmiştir. Bir olayın hangi yılda geçtiğini kararlaştırdıktan sonra, o yılın önündeki harfi, olayın solundaki boşluğa yazınız. "B" sütunundaki bir tarih, bir veya birden fazla kez kullanılabileceği gibi, hiç kullanılmayabilir.

**Öncüller / Maddeler**

**A Sütunu**

C	1. Osmanlı Meclis-i Mebusan'ın son toplanması
D	2. Ankara'nın başkent olması
H	3. Tek dereceli seçimlerin ilk kez yapılması
C	4. TBMM'nin açılması
D	5. Lozan Antlaşmasının imzalanması
D	6. Cumhuriyet'in ilanı
C	7. Sevr Antlaşmasının imzalanması

**B Sütunu**

A. 1918
B. 1919
C. 1923
D. 1925
E. 1928
F. 1930
G. 1934
H. 1936
I. 1946

**Muhtemel cevaplar / Seçenekler**

**Avantajları:** a. Hazırlaması, uygulaması ve puanlaması kolay, diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistematiik hata karışma olasılığı pek yoktur.

**Dezavantajları:** a. Şans başarısı olasılığı yüksektir. b. İyi yazılmazsa genellikle üst düzey bilişsel özellikleri ölçemez.

### Eşleştirme Maddeleri Yazım Kuralları:

- Her eşleştirme soru grubunda yönerge/açıklama yazılmalıdır.
- Seçeneklerin kullanılıp kullanılmayacağı, birden fazla kez kullanma durumu belirtilmelidir.
- Okul öncesinde, özel eğitimde eşleştirme şekillerle ya da resimlerle / fotoğraflarla yapılabilir.
- Bir eşleştirme maddesinde yer alan öncüller ile seçeneklerin her biri benzeşik (homojen) öğelerden oluşmalıdır. Diğer bir ifadeyle aynı kapsamdan gelmelidir.
- Uzun ifadeler öncül olarak kullanılmalı
- Bir eşleştirmeli maddeler grubundaki madde sayısı, en az 6, en çok 15 olmalıdır.
- Maddelerin tümü aynı sayfada bulunmalıdır. Aksi hâlde cevabın seçimini güçleştirir.
- Cevapların seçileceği sütun, bir kelime listesi ise alfabetik sıraya göre düzenlenmelidir.

## 3. ÇOKTAN SEÇMELİ MADDELER:

### Çoktan Seçmeli Maddelerin Özellikleri:

- ✓ Hazırlanması zor ancak puanlaması kolaydır.
- ✓ Puanlama objektiftir.
- ✓ Üst düzey bilişsel becerilerle ilgili madde yazmak zordur.
- ✓ Doğru cevabı içinde barındırması nedeniyle sentez düzeyinde soru sormak olanaklı değildir.
- ✓ Çok sayıda soru sorulabilir.
- ✓ Sistematiik hata karışma ihtimali çok düşüktür.
- ✓ Çeldirme mantığı temellidir.

### Çoktan Seçmeli Madde Yazım Kuralları:

- Gereksiz görsel kullanılmamasına da özen gösterilmelidir.
- Madde kökünde gereksiz yere sözcük kullanılmamalıdır.
- Seçeneklerde gereksiz sözcük tekrarı kaçınılmalıdır.
- Bir maddenin doğru yanıtı, başka bir maddenin kökünde ya da seçeneklerinde bulunmamalıdır.
- Birbirinin tam zıttı olan durumların seçeneklerde verilmesi çoğunlukla uygun değildir.
- Seçenekler yazılırken anlamca birbirini içeren / kapsayan ifadelerin kullanılmamasına dikkat edilmelidir.
- Seçeneklerin uzunlukları birbirine eşit olmalıdır.
- Doğru cevaplar teste dağıtılırken dikkat edilmeli, doğru cevapların bir örüntü göstermemesi sağlanmalıdır.
- "Hepsi" ve "hiçbiri" bir seçenek olarak kullanılmamalıdır.
- Çoktan seçmeli maddelerde seçenek sayısı genellikle 4 ya da 5 olsa da seçenek sayısını belirleyen temel faktör öğrencinin içinde olduğu gelişim dönemidir. Lise ve daha üstü öğretim düzeylerinde 5, ortaokul düzeyinde 4 ve ilkokul düzeyinde 3 seçenek kullanılabilir.

## AÇIK UÇLU SORULAR

### 1. ESSAY (YAZILI YOKLAMA):

**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.

- Sınırsız cevap özgürlüğü, anlatım becerisi iyi olan öğrencilere bir avantaj sağlayabilir.
  - Bu sınav türü, dünyada en eski ve günümüzde de öğretmenlerin hâlâ en çok tercih ettiği testlerden biridir. Bunun iki nedeni vardır: a. Hazırlaması kolaydır. b. Üst düzey bilişsel becerileri ölçmek için geleneksel yöntemler arasında en avantajlı yazılı test etme yoludur (Sentezde tek yol.).
  - Yazma, problem çözme, bilgileri organize etme, analiz vb. özgün bir ürün ortaya koyma (yaratıcı düşünme), eleştirel düşünme, yeni durumlara beceriyi transfer etme, analitik düşünme, bilimsel düşünme, hipotez oluşturma ve neden-sonuç ilişkilerini açıklama, veri düzenleme, güçlü ve zayıf yönleri belirleme gibi becerileri ölçmek için çok kullanışlıdır.
  - Çok fazla soru sorulamaması bir dezavantajdır.
  - Puanlar aynı zamanda öğrencinin anlatım biçimini, yazı güzelliğini, bilgisini örgütleme biçimini de yansıtır. Bu ise ölçme sonuçlarındaki sistematik hata miktarını artırabilir.
  - Puanlama özneliği belki de en önemli dezavantajdır.
- "Essay"de Nesnellik Artırma Yolları:**
- Ayrıntılı bir cevap anahtarı çıkarılmalıdır.
  - Bütüncül (holistik) ya da analitik rubrik hazırlanabilir.

### 2. KISA CEVAPLI SORULAR:

- ☑ Kısa cevaplı testler; öğrencinin bir sözcük, bir rakam, bir tarih ya da en çok bir cümle ile cevaplandırabileceği sorulardan oluşur.
- ☑ Eğer konu kapsamı bakımında homojen bir yapıda ise gruplandırılarak sorulması daha uygun olacaktır.
- ☑ Uzun cevap gerektiren sorularda olduğu gibi bu soru türünde de ayrıntılı bir cevap anahtarı hazırlanmalıdır.

**Avantajları:** a. Hazırlaması, uygulaması ve puanlaması kolay; diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistematik hata karışma olasılığı pek yoktur.

İyi yazılmazsa genellikle üst düzey bilişsel özelliklerin ölçülememesi dezavantajdır.

### 3. CÜMLE TAMAMLAMA SORULARI:

- Hazırlanması oldukça kolaydır.
- Cümleler ders kitabından aynen alınmamalı, öğretmen kendi sözcükleriyle soruları yeniden yazmalıdır.

**Avantajları:** a. Hazırlaması, uygulaması ve puanlaması kolay; diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistematik hata karışma olasılığı pek yoktur.

İyi yazılmazsa genellikle üst düzey bilişsel özelliklerin ölçülememesi dezavantajdır.

#### Cümle Tamamlama Soruları Yazım Kuralları:

- ✖ Her kısa cevap maddesi, yalnızca tek bir doğru cevabı olacak biçimde yapılandırılmalıdır.
- ✖ Bir maddenin ifadesinde, o maddenin cevabının bulunmasında işe yarayacak ipuçları vermekten kaçınılmalıdır.
- ✖ Cümlelerin gramer yapısı bağlamında ipucu genellikle eklerle verilmektedir.
- ✖ İlköğretimde cevapların ayrı bir cevap kâğıdına işaretletilmesi uygun görülmemektedir ancak ortaöğretimde bunun pek bir sakıncası bulunmamaktadır.
- ✖ Bir cümlede çok sayıda boşluk bırakılmamalı

#### 4. SÖZLÜ YOKLAMA:

Sözlü sınavlar çok iyi organize edilse dahi ciddi bir ölçme hatası barındırma potansiyeline sahiptir. Sözlü yoklamalar özellikle **dile dayalı becerilerin ölçülmesi** için geleneksel yöntemler içindeki tek türdür. Örneğinsözlü anlatım, diksiyon, yabancı dil dersinde konuşma, solfej ve şarkı söyleme vb.

- Bireysel bir test olması nedeniyle tüm öğrencileri test etmek için gereken süre oldukça fazladır.
- Cevaplama çoğu zaman üzerinde düşünmeye ve tasarlamaya olanak olmadan verilir.
- Puanlama hemen yapılır. Bu ise cevabın doğruluğunun genel izlenimle yapılmasına neden olur.
- Öğretmenin ölçme sonucuna sistematik hata karıştırma olasılığı yükselir.
- Sözlü anlatım becerisi iyi olan bir öğrenci, iyi olmayan bir başka öğrenciden daha az biliyor olsa da yüksek puan alabilir.
- Öğrencinin kılık-kıyafeti, diksiyonu, hâl ve hareketleri puanlamaya artı ya da eksi yönde etki edebilir.
- Çok fazla soru sorulamayacağı için **kapsam geçerliliği** problemi olabilir.

#### Sözlü Yoklama Uygulama Kuralları:

- Cevap anahtarı hazırlanmalıdır.
- Sınavın yapılması aşamasında öğrencilerin kaygısını artıracak davranışlardan kaçınılmalıdır. Hazır bulunuşluluk artırılmalıdır.
- Puana ilişkin dönüt hemen verilmelidir.
- Puanlamada sistematik hata kaynaklarına karşı dikkatli olunmalıdır.

#### GENEL DEĞERLENDİRME

- ☑ Her madde / soru türünün avantajları olduğu gibi dezavantajları da vardır. Dezavantajları asgariye indirmenin yolu madde / soru çeşitlemesi yapmaktır ancak özellikle ilkokulda, çocukların gelişim süreçleri de dikkate alındığında, madde / soru çeşidi sayısını çok artırmamak gerekir.
- ☑ Geleneksel test etme yaklaşımları ürün odaklıdır. Eğitim sürecinde ürün elbette önemlidir ancak ürün kadar öğrencinin süreçte bilgiyi nasıl yapılandığındının izlenmesi de beklenen ürünün ortaya konulabilmesi açısından çok önemlidir. Bu nedenle geleneksel ölçme yöntemleri, destekleyici yöntemlerle (alternative assessment) desteklenmelidir.

#### DESTEKLEYİCİ DEĞERLENDİRME YAKLAŞIMLARI

#### 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.
- ✚ Öğrenciyi başarılı / başarısız olarak sınıflandırmayı sağlayan bir araç değildir.
- ✚ Öğrencilerin özelliklerini birbirleri ile karşılaştırmak amacıyla oluşturulmuş araçlar değildir. Öğrenci portfolyosunun kabaca iki türü vardır. Bunlar:
  - a. **Süreci yansıtan portfolyolar:** Öğrencinin öğrenme ve gelişim sürecini yansıtır.
  - b. **Ürünü yansıtan portfolyolar:** Öğrenme sürecinden çok bitmiş görevleri içerir.

Portfolyonun içeriği:

- Öğretmen kayıtları (gözlemler, anekdot kayıtları)
- Öğrencinin çalışmaları
- Öğrencinin sözel ve psikomotor becerilerini gösteren teyp ve video kayıtları
- Öğrencinin kendi çalışmaları hakkındaki düşünceleri, günlükler
- Öğrencinin yazdığı mektuplar



## Portfolyo Oluşturma Süreci



### TOPLAMA:

1. Hangi çalışmaların toplanacağına ve hangi özelliklerin gözleneceğine karar verilmesi
2. Her öğrenci için sınıfta çalışmalarının toplanacağı ayrı bir kutu, dosya vb. oluşturulması
3. Her bir çalışmanın ve öğretmenin kaydının üzerine tarih yazılması

### SEÇME

1. Öğrenci seçimini kendi başına veya öğretmenin rehberliğinde yapabilir.
2. Portfolyonun türüne ve öğretmenin koyduğu koşullara bağlıdır.
  - a. **Süreci yansıtan portfolyolarda** öğretmenin belirlediği konuları / gelişim alanlarını yansıtan çalışmaların seçilen örnekleri.
  - b. **Ürünü yansıtan portfolyolarda** öğretmenin belirlediği konular ile ilgili ortaya çıkan ürünler arasından seçilenler yer alır.

### YANSITMA:

Portfolyoyu herhangi bir çalışma dosyasından ayıran en önemli aşamadır. Bu aşamada öğrenci:

1. Portfolyosuna seçtiği her bir çalışmayı niçin seçtiğini açıklar.
2. Çalışmalarını yaparken geçirdiği süreci ve bu süreçte öğrendiklerini anlatır.
3. Kendi başarısını görür, bunu ifade eder ve değerlendirme sürecine katılır.

#### Yansıtma Soruları:

- |  |  |
|--|--|
| <input type="checkbox"/> Bu çalışmayı nasıl yaptım?                        | Bu çalışmadan ne öğrendim?                       |
| <input type="checkbox"/> Bu çalışmayı daha da geliştirebilir miyim? Nasıl? | Bana zor gelen bir çalışmam var mı? Varsa neden? |
| <input type="checkbox"/> Bu çalışmayı portfolyoma neden koydum?            |  |

### SONUÇ

- ☐ Bu aşamada öğrenci "Bu çalışmayı niçin yaptık?" sorusunu yanıtlar.
- ☐ Okulda yaptığı çalışmalarla öğrendikleri arasında somut bağlar kurar.
- ☐ Tamamlanan portfolyo çalışmalarının öğrenci tarafından sınıf arkadaşları, öğretmeni ve ailesinden oluşan bir gruba sunumu yapılmalıdır.

## PERFORMANS DEĞERLENDİRME



Şekil 16: Performans değerlendirme araç ve yöntemleri

**1.ÖZ DEĞERLENDİRME (öğrenci değerlendirmesi):** Öğrencinin belirli bir konuda (örneğin bir ürünü ortaya koymada gösterdiği performans vb.) kendi kendisini değerlendirmesine denir.

- ✘ Öğrencilerin kendi özellikleriyle (yetenek, ilgi, beceri vb.) ilgili farkındalığının artmasını, zayıf ve güçlü yönlerini keşfetmesini sağlar.
- ✘ Öz düzenleme becerisi artar.
- ✘ Ölçütlü düşünme becerisi artar.
- ✘ Öğrenme motivasyonunu artırır.

**Öz Değerlendirmenin Olası Dezavantajları:** a. Öğrencilerin kendi performanslarını değerlendirirken yanlış davranışları söz konusu olabilir. b. Başlangıçta deneyimsizlik nedeniyle performansın değerlendirilmesinde yanlışlar olabilir.

**2.AKRAN DEĞERLENDİRME (öğrenci değerlendirmesi) :** Öğrencinin ortaya koyduğu performansa ilişkin arkadaşlarının değerlendirmesine denir.

- ☑ Daha katılımcı, aktif bir eğitim ortamı sağlanabilir (Sorumluluk duygusu artar.).
- ☑ Öğrencilerin eleştirel düşünme becerileri artar.
- ☑ Öğretmen dışındaki bir kaynaktan dönüt almak öğrencinin performansını artırabilir.
- ☑ Eleştiri kültürü (olumlu-olumsuz) gelişebilir.

### Öğretmen değerlendirmeleri

**1. RUBRİK (DERCELE PUANLAMA ANAHTARI):** Amaç, öğretmen tarafından ürünün genel izlenimle puanlamasındaki özneliğini azaltmaktır. Rubrikler ikiye ayrılır: **Bütüncül (Holistik) rubrik ve Analitik rubrik**

- Rubrikler performans görevleriyle birlikte öğrenciye verilmelidir.
- Rubriklerin geliştirilmesi uzmanlık gerektirir. Rubrikler için geçerlilik ve güvenilirlik çalışmaları yapılmalıdır.

Ölçüt	Puan
Maketi yaptığı materyal kalitelidir.	10
Ek yerlerinde pürüz bulunmamaktadır.	15
Özgün bir fikir ortaya koymuştur.	20
Maket çevre düzenlemesi gibi görsel öğelerle zenginleştirilmiştir.	15
Renklendirmeler uyumludur.	10
.....	....
.....	....
.....	....
<b>Toplam</b>	<b>100</b>

Ölçüt	1 Puan	2 Puan	3 Puan	Puanı
Maketi yaptığı materyal kalitelidir.	Karton	Kontrplak	Masif	
Ek yerlerinde pürüz bulunmamaktadır.	>0.5 cm	0.3-0.5 cm	<0.3 cm	
Özgün bir fikir ortaya koymuştur.	Ders örnekleriyle aynı	Ders örneklerine yeni öğeler katmış	Tamamen özgün bir fikir	
Maket çevre düzenlemesi gibi görsel öğelerle zenginleştirilmiştir.	Zenginleştirme yetersiz	Ağaç, park vb. yapmış ancak yetersiz	Çevreyi tüm öğeleriyle yansıtmış	
Renklendirmeler uyumludur.	Renklendirmede özensiz	Kısmen uyumlu	Renklendirmeler çok iyi	
.....	.....	.....	.....	
.....	.....	.....	.....	
.....	.....	.....	.....	
<b>Toplam</b>				

Bütüncül Rubrik Örneği (Maket Yapımı)

Analitik rubrik örneği (Maket Yapımı)

## 2. KONTROL LİSTELERİ:

- ➡ Gözlenen performans ürününün ölçütlere uygunluğunu "evet-hayır", "var-yok", "gösterdi-göstermedi" vb. bir biçimde kategorik (1-0) olarak puanlama amacıyla kullanılan araçlardır.
- ➡ Özellikle sergilenecek performans detaylı ve ardışık eylemler gerektirdiği zamanlarda kullanışlıdır. Örneğin bir deneyin eyleme dökülmesi vb.
- ➡ Pek çok işlem adımında oluşan performanstaki eksik adımları belirlemek için oldukça uygundur.

**3. 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.

**4. GÖZLEM FORMLARI:** Özellikle fen derslerinde, meslek liselerinin somut performans ürünlerinin geliştirildiği vb. alanlar için oldukça uygundur. Gözlemler öğrenciler hakkında doğru ve hızlı bilgi elde edilmesini sağlar.

## GELENEKSEL VE DESTEKLEYİCİ YÖNTEMLERİN KARŞILAŞTIRILMASI

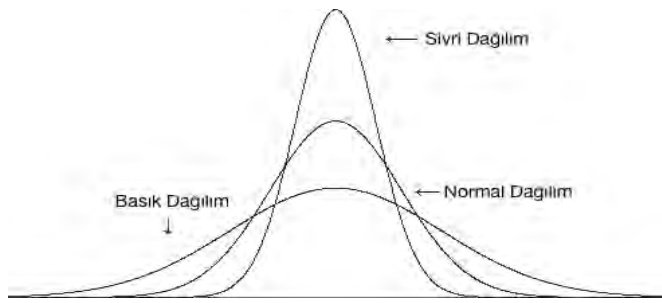
GELENEKSEL YÖNTEMLER	DESTEKLEYİCİ YÖNTEMLER
Ürün değerlendirilir.	Süreç ve ürün birlikte değerlendirilir.
Öğrencinin ulaştığı noktanın tespiti önemlidir.	Ne öğrendikleri yanında, öğrendiklerini nasıl kullandıklarıyla ilgilenilir.
Essay dışında genellikle üst düzey düşünme becerilerinin değerlendirilmesinde yetersizdir.	Üst düzey bilişsel düşünme becerilerine odaklanır.
Değerlendirme öğrenmeden ayırır.	Değerlendirme öğrenmeyle bütünleşmiştir.
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.

## TEST VE MADDE İSTATİSTİKLERİNE GENEL BİR BAKIŞ

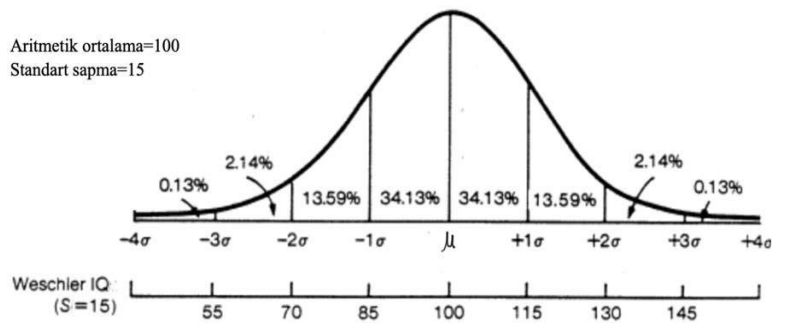
### A. TEST İSTATİSTİKLERİ

- Merkezî eğilim ölçüleri: **aritmetik ortalama, medyan (ortanca) ve moddur**. Bu üç istatistiğin de temel işlevi bir **puan dizisindeki merkezi bulmaktır**.
- Merkezî eğilim ölçüleri, örneğin ortalama, grubun başarısı (ya da hangi özellik gözleniyorsa) hakkında kabaca da olsa **fikir yürütmemizi** sağlar.
- Diğer taraftan değişkenlik ölçüleri **ranj (dizi genişliği), varyans, standart sapma ve çeyrek sapmadır**.
- Puanların değişkenliği, grup hakkında biraz daha ayrıntılı bir değerlendirme yapılmasını sağlar.
- Puanların dağılım özellikleri yine grup hakkında bilgi elde etmemizi sağlayan istatistiklerdir





Aritmetik ortalama=100  
Standart sapma=15



### Normal, sivri ve basık dağılımlar

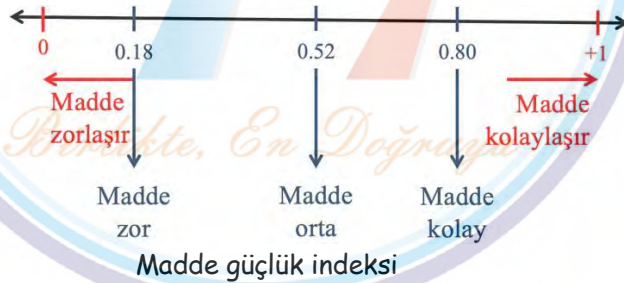
### Standart normal dağılım altında yer alan alanlar

- ☑ Normal dağılım ortalama ve standart sapma gibi basit iki istatistikle özetlenebilir.
- ☑ Normal dağılım ortalama etrafında üç standart sapma sağa ve sola, toplamda **altı alanda** özetlenebilir (%99,74'lük kısmı). Dolayısıyla bir kişinin ortalamadan kaç standart sapma uzaklaştığı, grup içerisindeki yerini gösterir.

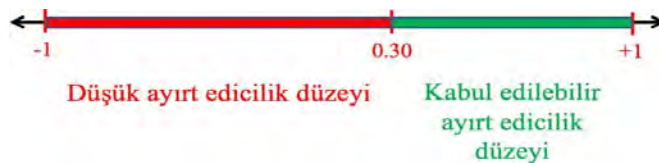


**B. MADDE İSTATİSTİKLERİ:** Ölçme işlemi sonucunda bireylerin madde puanları üzerinden hesaplanan istatistiklerdir. Eğitimde en sık kullanılan istatistikler madde güçlük indeksi, madde ayırt edicilik indeksi, madde güvenilirliği vb.dir.

Madde güçlük indeksi; maddenin kabaca zor mu, orta mı, kolay mı olduğunu gösteren bir istatistiktir.



- ➡ Madde güçlük indeksi 1'e yaklaştıkça maddenin kolaylaştığı, 0'a yaklaştıkça zorlaştığı anlaşılar. 0.50 civarı ise ortalama güçlükte olduğu biçiminde yorumlanır.
- ➡ Bir test zorlaştıkça dağılımın sağa çarpık, kolaylaştıkça sola çarpık olma olasılığı artar. Diğer bir ifadeyle grup başarı düzeyi açısından homojenleşmeye başlar ve homojen bir dağılımda testin bilenle bilmeyeni birbirinden ayırt etme gücü azalabilir ve güvenilirlik katsayısı düşebilir.
- ➡ Madde ayırt edicilik indeksi ise maddenin bilenle bilmeyeni ayırt edip edemediğine ilişkin bir istatistiktir. -1 ile 1 arasında değer alır ve genellikle alanyazında 0.30 ile 1 arası ayırt edici olarak nitelendirilir.



### Madde ayırt edicilik indeksi

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

## Kavramlar

**Erişi ya da başarı kavramı**, sunulan bilgilere, sunulan içeriğe o eğitimi alan bireylerin ne kadar ulaşabildiğine dair bir gösterge, erişim düzeyi olarak ifade edilmektedir.

**Beceri**; bilgilerin bir araya toplanıp, doğru bilgilerin seçilip ardından seçilen bilgilerin doğru şekilde kullanılması, gerçek yada tasarlanan bir durumda kullanılabilme yetisini ifade etmektedir.

Günümüzde dünyada önem atfedilen eğitim izleme araştırmalarından birisi olan Uluslararası Öğrenci Değerlendirme Programında (PISA) ölçülen özellikler "okuryazarlık" olarak tanımlanmaktadır.

## Uluslararası İzleme Çalışmaları

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. **Uluslararası Matematik ve Fen Eğilimleri Araştırmasında (TIMSS)** matematik ve fen, **Uluslararası Okuma Becerilerinde Gelişim Araştırmasında (PIRLS)** ise okuma becerileri değerlendirilmektedir.

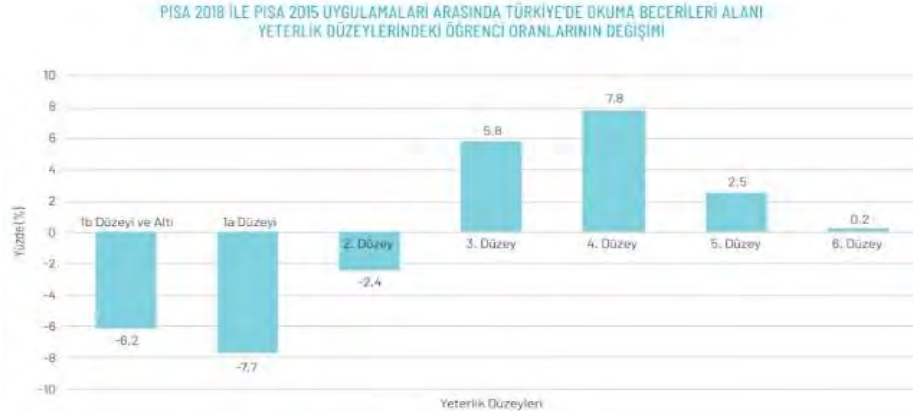
## Günümüzde En Yoğun Katılım Gösterilen Çalışmalar

- **PISA**, Ekonomik Kalkınma ve İşbirliği Örgütü (OECD) tarafından geliştirilen ve üç yıllık periyotlarla uygulanan izleme çalışmasıdır.
- **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

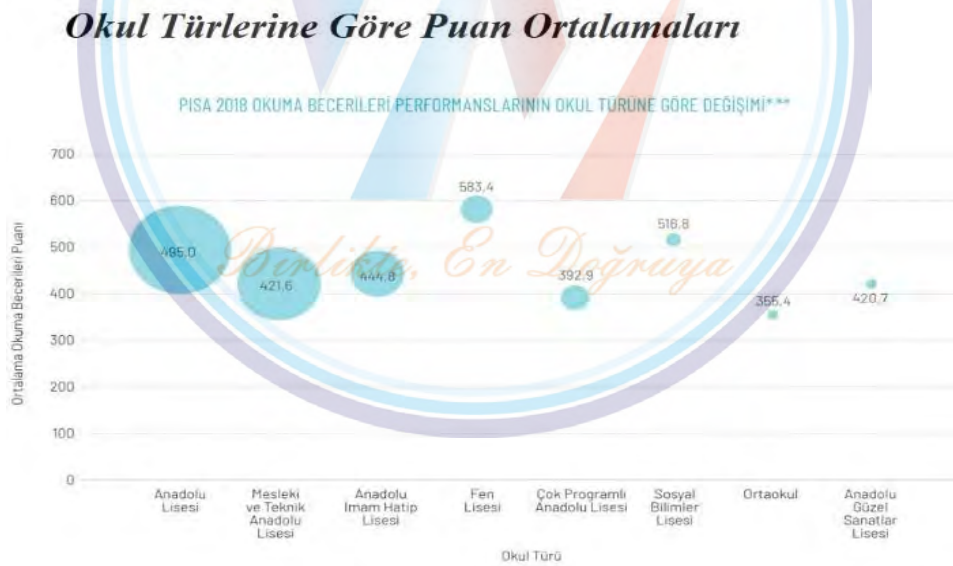
- En güncel PISA uygulaması olan 2018 döngüsüne 79 ülke/ekonomi katılım göstermiştir.
- PISA, **15 yaş grubunu** hedef grup olarak almaktadır. Bunun nedeni birçok Avrupa ve OECD ülkesinde 15 yaşa tekabül eden dönemin yaklaşık olarak ortaokulun sonuna denk gelmesi ve bu dönemde zorunlu eğitimin tamamlanmasıdır..
- PISA, 2000 yılından bu yana uygulanmakta ve **her üç yıllık döngüde** bir alana ağırlık verilmektedir. 2000 yılında ağırlıklı alan okuma becerileri iken 2003 yılında matematik okuryazarlığı ve 2006 yılında fen okuryazarlığı, 2018 yılında tekrar okuma becerileri olacak şekilde çapraz bir modelleme kullanılmaktadır.

## Yeterlik Düzeylerindeki Değişim



Şekil 1. Okuma becerisi ortalama puanlarındaki ve yeterlik düzeylerindeki değişim

Yeterlik düzeyleri son derece önemlidir çünkü her bir yeterlik üzerinde bulunan öğrencinin o alanda ne yapıp ne yapamadığını, davranış bazında hangi davranışları gösterip hangi davranışları gösteremediğine dair somut açıklamalar yapmaktadır. Örneğin en üst düzeye yeterlik düzeyi olan 6. düzeydeki bir öğrencinin çok karmaşık metinleri rahatlıkla okuyabilmekte, anlamlandırabilmekte, metin içerisindeki dinamiklerin farkına varıp bunları ilişkilendirebilmekte, farklı metinlerdeki örtülü anlamları farkedebilmektedir.



Şekil 2. Okul türlerine göre okuma becerisi puan ortalamaları ve performans farkları

**Grafiğin yorumu:** Şekil 2'nin sol tarafında bulunan dairelerin büyüklüğü öğrenci sayısının büyüklüğünü temsil etmektedir. Burada bir diğer önemli gösterge de geçen yıllara kıyasla okullar arasındaki başarı farkının kısmen azalmış olmasıdır. Elbette bu farkın daha da azalması hedeflenmektedir.



## TIMSS KAPSAMI VE SONUÇLARI

- 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.
- IMSS, ilk defa 1995 yılında uygulanmaya başlanmış olup 2019'da araştırmanın yedinci döngüsü tamamlanmıştır.
- TIMSS'te tabakalı örnekleme yöntemi kullanılmaktadır: Türkiye'deki tüm okulların listesi uluslararası merkeze gönderilmekte bu merkez iki ayrı sınıf düzeyi için (4 ve 8. sınıf) önce okulları tesadüfen seçmektedir. Ardından seçilen okullar içerisinden tesadüfi olarak şubeler seçilmekte ve örneklem seçim süreci tamamen bağımsız ve yansız bir şekilde uluslararası merkez tarafından gerçekleştirilmektedir.
- Uygulama sürecine ise öğrenciler matematik veya fen testi ile başlamaktadır. Sınav yaklaşık 36-45 dakika sürmekte, ardından 15-30 dakika arası mola verilmektedir. Diğer testin uygulaması tamamlandıktan sonra 5-15 dakikalık bir anket uygulaması gerçekleştirilmektedir.



Şekil 6. Matematik yeterlik düzeyleri oranları

**Grafiğin yorumu :** Şekil 6'da görüldüğü gibi PISA'ya benzer şekilde TIMSS sonuçlarımızda da hem 4 hem de 8. sınıfta matematik ve fen alanlarında önemli iyileşmeler söz konusudur. Fen bilimleri alanında yeterlik dağılımları incelendiğinde ileri düzey seviyesindeki öğrenci oranları 2011'de %3'e, 2015'te %4'e ve son uygulamada %12'ye çıkmıştır.

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.

## OECD SOSYAL VE DUYGUSAL BECERİLER ARAŞTIRMASI

2017 yılında OECD tamamen sosyal ve duygusal beceriler odaklı yeni bir araştırma başlatmıştır. Türkiye de İstanbul'la bu çalışmaya katılmıştır.

Öncelikle bu çalışma neden önemlidir ve nasıl bir fark yaratmaktadır?

- ➡ **Birinci olarak** tümüyle 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.
- ➡ **Üçüncüsü** ölçümler, dünyanın farklı bölgelerinde kabul gören beş faktörlü kurama dayalı olarak (Big five modeli) yapılmaktadır. Bu beş faktörlü model, Şekil 7'de verilmiştir.



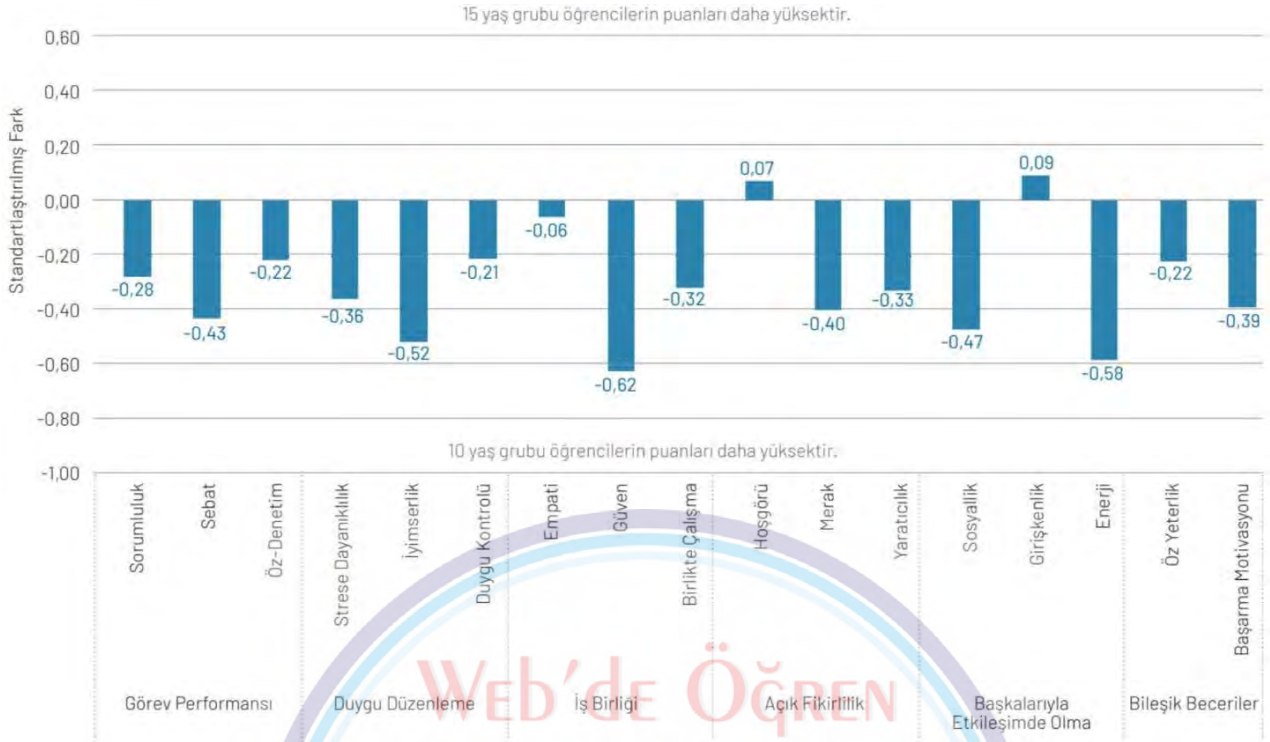
Şekil 7. Büyük Beşli Sosyal ve Duygusal Beceriler Modeli

➤ Beş faktörlü modeli oluşturan ana alanlar ve bu alanları oluşturan alanlar:

1. **açık fikirlilik**(yaratıcılık, hoşgörü, merak),
2. **iş birliği** (empati, iş birliği, güven),
3. **duygu düzenleme** (iyimserlik, strese dayanıklılık, duygu düzenleme),
4. **görev performansı** (sebat, öz denetim, sorumluluk)
5. **başkalarıyla etkileşimde olma** (sosyallik, girişkenlik, enerji)

➤ Bir de birleşik beceriler diye tanımlanan ve bu beşliye dâhil olmayan ama onunla ilişkili olup sosyal beceriler açısından önemli görülen iki ayrı beceri ise **başarma motivasyonu ve öz yeterlik algısıdır**.

➤ Çalışmaya Türkiye, 10 yaş grubundan yaklaşık 2.700 ve 15 yaş grubundan ise yaklaşık 3.200 öğrenci ile katılmıştır.



Şekil 8. Yaş düzeylerine göre sosyal ve duygusal beceriler

**Grafiğin yorumu** Şekil 8'de görüldüğü gibi çalışmanın enteresan sonuçlarından biri 10 yaş grubunun sosyal ve duygusal beceri puanlarının 15 yaş grubundan daha yüksek olmasıdır.

Ulusal izleme çalışmaları kapsamında Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE) araştırması, üst düzey bilişsel becerilere öğrencilerimizin ne kadar sahip olduklarını incelenmekte; Türkçe-Matematik-Fen Öğrenci Başarı İzleme Araştırması (TMF-ÖBA) ise öğrencilerimizin eğitim programına bağlı kazanımlara ulaşma düzeylerini belirlemek üzere yapılmaktadır.