

ÖLÇME VE DEĞERLENDİRMENİN TEMEL KAVRAMLARI

Bir bütüne sistem diyebilmek 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, bazı ülkelerde çerçevesi anayasa ile güvenceye alınmıştır.

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

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.

2. **Süreç:** Girdilerin hedefler doğrultusunda bir araya getirildiği, dersin hedeflerinin öğrenciye kazandırıldığı eğitim durumlarını kapsar.

3. **Çıktı:** Öğrencilerin sahip olduğu niteliklerdir.

Eğitimde kontrol (denetleme) ölçme ve değerlendirme ile mümkündür. Gözlenmeyen süreçler ve ürün değerlendirilemez.

DİKKAT: 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: Öğrencinin belirli bir özelliğini gözleyerek o özelliğe sayı, sembol ya da sıfat/kategori adı verme işlemidir. Örneğin

• *Kerim Temel Yeterlilik Testinden (TYT) 450 puan aldı.*

• *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.

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.

• *Kerim gıda mühendisliği programına yerleştirilmiştir / yerleştirilememiştir.*

• *Can üstün yetenekliler sınıfına kabul edilmiştir / edilmemiştir.*

• *Zeynep dersten geçmiştir / kalmıştır.*

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

Eğitimde ölçüt, mutlak ve bağıl olmak üzere ikiye ayrılır.

- 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.
- 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.

DİKKAT :

- ❖ 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..*
- ❖ Muafiyet sınavlarında mutlak değerlendirme yapılması gerekir.

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 üniteadaki öğrenme eksiklerini belirlemek amacıyla yapılır. Amaç yine not vermek değildir. 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 çünkü artık hedeflenen öğretim sürecinin sonuna gelinmiştir. Bu amaçla uygulanan testlere “alanda eriş 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:

A) Özel eğitim: Belirli alanlarda özel gereksinimi olan çocuklara uygun eğitim ortamları ve/veya programı uygulamak amacıyla ölçme ve değerlendirme süreçleri işe koşulmaktadır. 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.

B) 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.

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

1. **Geçerlilik** ölçmek istenilen özelliğin, başka özelliklerle karıştırılmadan, doğru ve tam olarak ölçülebilesidir. Diğer bir ifadeyle ölçme aracından elde edilen puanın amaca hizmet etme derecesidir.
2. **Güvenilirlik** ölçme işleminden elde edilen puanların tesadüfi hatalardan arınık olma derecesidir. Diğer bir ifadeyle puanların kararlı, tutarlı ve duyarlı olmasıdır.
3. **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. Ölçme işlemlerinde hataları en aza indirebilmek için öncelikle hataların tanımlanması gerekir ki ölçme işlemini yapan kişi neye müdahale edeceğini, hangi durumlara karşı önlem alması gerektiğini bilsin.

HATA

- a) **Sabit Hata:** Miktarı ölçmeden ölçmeye değişmeyen, diğer bir deyişle her ölçme işlemine aynı miktarda karışan hatalardır.
- b) **Sistemik Hata:** Ölçülen büyüklüğe, öğretmene ya da ölçme koşullarına göre miktarı değişen hatalardır.
- c) **Tesadüfi (Rastlantısal) Hata:** Şansla ortaya çıkan ne yönde ve ne ölçüde karıştığı genellikle bilinmeyen hatalardır.

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 ve/veya kaygısını artıracak davranışlar.
Ölçme işlemi sonrasında: Puanlamadaki dikkat ve titizliğin zamandan zamana değişmesi, yorgunluk, öncelik-sonralık yanılgısı, maddi hata
2. **Ölçme aracından kaynaklanan hata:** Ölçme araçları hazırlanırken maddelerin iyi ifade edilmemesi, test katılımcılarının yanlış anlamalarına ve dolayısıyla hataya neden olacaktır. O nedenle hangi özellik ölçülürse ölçülsün, araçta

kullanılan dilin dil bilgisi, imla kurallarına uygun olması; maddelerde anlatım bozukluğu bulunmaması, muğlak ya da belirsiz (müphem) ifadelerin bulunmaması gerekir. Ölçülecek özellikler evrenini yeterince temsil etmeyen maddelerden oluşan bir araçtan elde edilen puanlara da hata karışmaktadır. Ölçme aracındaki madde sayısı arttıkça bireyde yorgunluk, dikkat azalması vb. etkenler nedeniyle hata miktarı artacaktır. Ölçme aracıyla ilgili diğer bir faktör ise bilişsel özellikleri ölçen testlerde eğer seçenek verilmişse şans başarısı 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 nedeniyle bireyler gerçek puanlarından uzaklaşabilirler. Ayrıca grubun homojen ya da heterojen olması da hata kaynağını arttıran ya da azaltan bir etkiye sahiptir. Daha heterojen gruplarda güvenilirlik katsayısı artarken daha homojen gruplarda güvenilirlik katsayısı azalır.
4. **Fiziksel ortamdan kaynaklanan hata:** Ölçme işleminin gerçekleştiği fiziksel ortama ilişkin bazı etmenler de ölçme sonuçlarına hata karıştırabilir.

GEÇERLİLİK, GÜVENİLİRLİK VE HATA İLİŞKİSİ

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

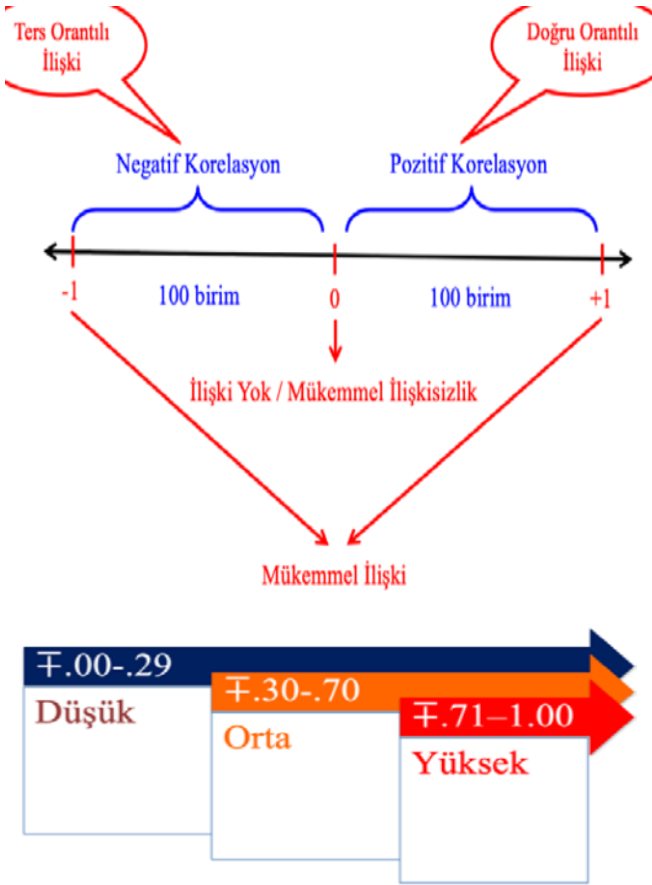


Bu nedenle güvenilirlik, geçerlilik için bir ön şart ancak yeterli şart değildir. Diğer bir ifade ile bir testin güvenilir olması onun geçerli olacağı anlamına gelmez ancak bir test geçerli ise büyük olasılıkla güvenilir. Geçerlilik ve güvenilirlik bir varlık yokluk sorunu değil, derece sorunudur. Bu nedenle bir testten elde edilen puanlar yüksek düzeyde geçerli, orta düzeyde geçerli, düşük düzeyde geçerli; yüksek düzeyde güvenilir, orta düzeyde güvenilir, düşük düzeyde güvenilir biçiminde nitelendirilir.

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.

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.



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. 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:

a) **Mantıksal / rasyonel yöntemler:** Bu yöntemler belirtke tablosu hazırlanması ve uzman görüşüne başvurulmasıdır.

- **Belirtke tablosunun hazırlanması:** Bir kapsam geçerliliği sorgulamasında öncelikle ölçmeye konu olan kapsam dâhilinde davranışların belirlenmesi gerekir. Bu noktada en çok kullanılan yöntemlerden biri belirtke tablosu hazırlamaktır.
- **Uzman görüşüne başvurulması:** Uygulamada uzman ile kastedilen öncelikle ölçme ve değerlendirme tekniklerini de bilen bir alan uzmanıdır.

b) **İstatistiksel yöntemler:** Uzmanlardan alınan dönütler betimsel / muhakemeye dayalı bir yolla çözümlenebilir ya da uzmanların “uygundur / uygun değildir” ya da “uygundur / düzeltme gerekir / soru kullanılmamalıdır” vb. biçimde değerlendirme yapması istenebilir. Eğer uzmandan ikinci yolla dönüt istenmişse uzmanlar arasında uyum olup olmadığı, çeşitli istatistiksel yöntemlerle test edilir. Alanda uzmanlar arasında uyumu test eden pek çok indeks bulunmaktadır

2. YORDAMA GEÇERLİLİĞİ

Yordama, eldeki bu nitelikteki veriden yola çıkarak geleceğe, henüz gerçekleşmemiş bir olguya ilişkin yapılan tahmindir.

Yordama geçerliliğinde en zor ve önemli nokta ölçütün doğru bir biçimde belirlenmesidir. Ölçütün belirlenmesinde şu noktalar dikkate alınmalıdır:

Ölçüt(ün);

- ❖ Ölçme aracının yordamaya çalıştığı değişkenle doğrudan ilgili olmalı, ölçme aracı hangi özelliği kestirmeyi amaçlıyorsa onun doğrudan bir temsili olmalıdır.
- ❖ Kararlı olmalı, günden güne değişmemelidir. Açıktır ki kendisi kararsız olan bir özellik hiçbir araçla yordanamaz.
- ❖ Bireylerin özelliğini gerçekten yansıtan nesnel ve güvenilir bir ölçüt olmalıdır. Söz gelimi okulda alınan notlar bir ölçüt olarak alınmışsa öğrencilere verilen notlara başarının dışındaki etmenler etki etmemelidir.
- ❖ Elde edilmesi kolay ve ekonomik olmalıdır.

3. ZAMANDAŞ GEÇERLİLİK

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.

Zamandaş geçerliliğinde geçerlilik katsayısının -1 ile 1 arasında değişmesi beklenir. Ölçüte bağlı olarak geçerlilik katsayısı ∓ 1 'e yaklaştıkça artar, 0'a yaklaştıkça düşer.



Zamandaş geçerlilikte de en zor ve önemli nokta ölçütün doğru bir biçimde belirlenmesidir. Ölçütün belirlenmesinde şu noktalar dikkate alınmalıdır: Ölçüt(ün);

- ❖ Ölçme aracının ölçmeye yöneldiği özelliklerle doğrudan ilişkili olmalıdır. Bu ilişki doğru orantılı ya da ters orantılı olabilir.
- ❖ Geçerliliği yüksek olmalıdır. Geçerliliği yüksek olmayan bir ölçüt puanla bakılacak korelasyonun düşük olması kaçınılmazdır.

4. 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 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 almaktadır. 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?”

Örneğin 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 olamayacağını, gerçek puanının 24.7 olduğunu söyleyebiliriz. Bu söylem gerçek puan kuramına dayanır.

$$X = T + E$$

X = Bireylerin ölçme aracından elde ettiği gözlenen puanı

T = Bireylerin gözlenemeyen gerçek puanı

E = Ölçmeye karışan hata miktarı

Ölçmenin standart hatası

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

formülü ile bulunur.

$Sh = \text{Ölçmenin standart hatası}$

$s = \text{Standart sapma}$

$rx = \text{Güvenilirlik katsayısı}$

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 $rx=.75$ ise $Sh=8\sqrt{1-.75}=4$. Bu sonuç bireylerin puanlarına ± 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.

Güvenilirlik katsayısı 0 ile 1 arasında değişir. Güvenilirlik katsayısı için alan yazında genellikle 0.70 ve üstü ölçütü kabul edilir. Ancak bu değer de çok yüksek bir değer olmadığı, 0.70-0.80 arası güvenilirliğin ancak ön bilgi elde etmek amacıyla kullanılabileceği ifade edilir. Genel yetenek gibi bilişsel özellikleri ölçen testlerde güvenilirlik katsayısının 0.90 ve üzerinde olması arzu edilir.



1. 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.

Bu yöntem, ölçülen özelliğin kararlı olduğu durumlarda uygulanması gereken bir yöntemdir. Testin ölçtüğü özellik sürekli değişkenlik gösteriyorsa bu

yöntemle testin güvenilirliği hesaplanmamalıdır. Diğer bir deyişle bu yöntem daha çok iki uygulama arasında kolaylıkla değişmeyen özellikleri ölçen testler için uygundur. Örneğin genel zihin yetenekleri, kişilik testleri, ilgi envanterleri, tutum ölçekleri vb. gibi testlerin kararlılık bağlamında güvenilirlikleri bu yöntemle hesaplanabilir.

Güvenilirliğin bir boyutu da testin tutarlılığıdır. Tek uygulamaya dayalı güvenilirlik sorgulama yöntemleri ile "Test kendi içinde tutarlı bir bütün oluşturur mu?" sorusuna yanıt aranır. Bu nedenle bu yöntemlerin tümünden elde edilen katsayı **iç tutarlılık katsayısı** olarak adlandırılır.

2. 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 başvuru yöntemler:

- ❖ ilk yarı ve son yarı
- ❖ tek ve çift
- ❖ rastlantısalıdır.

İlk ve son yarı yöntemi her test için uygun değildir. Uygun olmadığı durumlar: Testteki

- maddeler basitten zora doğru sıralanmış ise
- maddeler konu içeriklerine ya da faktörlere göre kümelenerek yerleştirilmiş ise
- madde sayısı çok fazla ise elde edilen katsayı korelasyonun yarı puanlar üzerinden elde edilmesi nedeniyle testin tamamına ilişkin bir tutarlılık bilgisi vermez, yarısına ilişkin verir. Diğer bir ifadeyle elde edilen katsayı eşdeğer yarılardan birinin güvenilirliği olarak kabul edilir. 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.

3. KUDER-RICHARDSON 20 VE 21 YÖNTEMLERİ

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.

4. 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 **ic 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. KR-20 ve KR-21 yöntemlerinde bahsedildiği gibi testteki maddeler açısından benzeşiklik (homojenlik) arttıkça güvenilirliğin artması, diğer taraftan ayrışıklık (heterojenlik) arttıkça güvenilirliğin düşmesi beklenen bir durumdur.
5. Testten elde edilen puanların güvenilirlik kestirimi için veri elde edilecek grubun (örneklem) büyüklüğü arttıkça grubun heterojenleşmesinden dolayı güvenilirlik artar.
6. Aslında bir üst maddeyle de paralel bir biçimde maksimum performansı ölçen testlerde ortalama güçlüğü yaklaştıkça grup heterojenleşir ve dolayısıyla güvenilirlik artar. Heterojen gruplarda testin ölçmeye konu olan özelliğe sahip olanla olmayanı ayırt etme gücünün artması beklenir.
7. Maddelerin dil bilgisi kurallarına uygun, açık ve anlaşılır yazılması belki de güvenilirliği en çok artıran etmenlerden biridir.

8. Öğrencilere test uygulaması öncesinde ve gerekliyse süreçte yönerge vermek, test almaya güdülenme ve hazırbulunuşluluğun artması açısından önemlidir.
9. Fiziksel ortamın ses, sıcaklık, ışık, koku, dikkat dağıtıcı uyarıcılar, ergonomi vb. etmenler açısından uygun hâle getirilmesi yine güvenilirliği artıran bir başka etmendir.
10. Eğer süreli bir test uygulaması ise sürenin yeterli verilmesi gerekir.
11. Uygulamada bireylerin dikkatini dağıtacak ve/veya kaygısını artıracak davranışlardan kaçınılması, puanlama ya da veri girişi yapılırken dikkatli ve titiz davranılması ölçme işlemini yapan kişiden kaynaklanabilecek hataları azaltacağından dolayı 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.

TEST GELİŞTİRME VE MADDE / SORU TÜRLERİ

Sınıf içi ölçme ve değerlendirmelerde uygulanması daha olanaklı bir test planı önerisi;

1. Sınavın amacının belirlenmesi (puanların hangi amaçlarla kullanılacağı)
2. Sınavda yoklanacak davranışların belirlenmesi
3. Sınavın kapsayacağı konuların listelenmesi
4. Okulun takviminin ve sınav sonuçlarının 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
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



1. DOĞRU-YANLIŞ MADDELERİ

Çoktan seçmeli madde türünden sonra en çok kullanılan ikinci madde tipidir. Genellikle öğrencinin bilimsel gerçekleri, tarihî olayları, kesin yargıları tanınması ve hatırlanması; 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. Sistematik hata karışma olasılığı pek yoktur.

Dezavantajları: a. Şans başarısı olasılığı yüksektir. b. İyi yazılmazsa genellikle üst düzey bilişsel özellikleri ölçmez. c. Öğrenme eksiklerini belirleyemez. d.

Yanlış üzerinden öğretim yapılmaz. e. Her derste/konuda kesin yanlış denebilecek durumlar belirlemek zordur.

Doğru-Yanlış Maddeleri Yazım Kuralları:

- Her madde tek ve belirli bir fikri belirtmelidir.
- Doğru yanlış maddesi, kesinlikle doğru ya da kesinlikle yanlış olmalıdır.
- Bir maddenin yanlışlığı önemsiz bir ayrıntıda ya da aldatici bir noktada olmamalıdır.
- Mümkün olduğunca olumsuz ifade kullanılmamalıdır. Özellikle iki olumsuz ifadeden kesinlikle kaçınılmalıdır.
- Kanı ifadeleri kendiliklerinden doğru ya da yanlış olmadığından bir kaynağa dayandırılarak verilmelidir.
- Bir maddenin ifadesi kısa, açık ve yalın olmalıdır.
- Okuduğunu anlama süreçlerinin ilk basamağı, metin içerisinde açıkça verilmiş bilgiden doğrudan çıkarım yapmadır.
- Öğrencinin bir maddede ifadenin doğruluğuna ya da yanlışlığına karar vermesi gerektiğinde öğretmen özellikle vurgulamak istediği bir sözcük / terim vs. varsa onun altını çizebilir.
- Doğru-yanlış maddesinin özellikle yanlış olduğu durumlarda, öğrencinin şansa puan almasını engellemek amacıyla ifadeyi düzeltmesi ya da yanlışlığın ne olduğunu belirtmesi istenebilir.
- İyi kurgulandığı takdirde doğru-yanlış maddeleri ile üst düzey öğrenmeler de yoklanabilir.
- Doğru-yanlış maddeleri aynı kapsamdan geliyorsa ortak bir soru kökü altında birleştirilebilir.
- Doğru ve yanlış maddelerin testteki sıralanışı rastgele olmalıdır, belirli bir örüntü içinde olmamalıdır.
- Doğru ve yanlış maddelerin ifadesi yaklaşık olarak aynı uzunlukta olmalıdır.
- Bir doğru-yanlış testindeki doğru ve yanlış maddelerin sayısı, yaklaşık olarak birbirine eşit olmalıdır.
- Maddeleri işaretleme yöntemi, öğrenciye açık ve anlaşılır bir biçimde açıklanmalıdır
- Mümkün olduğunca kısa yazılmalı, basit cümle yapısı tercih edilmeli, özellikle “ve” gibi bağlaçlara dikkat edilmelidir.
- Olumsuz anlama sahip bir sözcük ya da ifade varsa altı çizilmelidir.
- İpucu vermekten kaçınmak gerekir.
- Üst düzey becerileri yoklamak için tablo, grafik, harita ya da okuma parçası gibi bir öncülün kullanılması önerilir.

2. EŞLEŞTİRME MADDELERİ

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

Avantajları: a. Hazırlaması, uygulaması ve puanlaması kolay, diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistemik 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. Bir eşleşme takımındaki öncüller ile cevaplar eşit sayıda olmamalıdır.
- Eşleştirmede ortaokul ve üstü öğretim kademelerinde öncülün başında bir boşluk bırakılarak öğrenciden seçeneğin harfini yazmaları istenebilir.
- 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.
- Uzun ifadeler öncül olarak kullanılmalı ve öncüller sütunu sayfanın sol sütununa, cevaplar sütunu ise sayfanın sağ tarafına yerleştirilmelidir.
- Bir eşleştirmeli maddeler grubundaki madde sayısı, en az 6, en çok 15 olmalıdır.
- Maddelerin tümü aynı sayfada bulunmalıdır.
- Cevapların seçileceği sütun, bir kelime listesi ise alfabetik sıraya göre (küçükten büyüğe ya da büyükten küçüğe) düzenlenmelidir.
- Bu eşleştirme maddeleri öğrencilerin yalnızca hatırlama düzeyi dışında, kavrama düzeyindeki öğrenmelerinin de yoklanabileceğine ilişkin örnektir.

3. ÇOKTAN SEÇMELİ MADDELER

Hazırlanması zor ancak puanlaması kolaydır. Puanlama objektiftir. Üst düzey bilişsel becerilerle ilgili madde yazmak zordur. O yüzden yazılan maddeler genellikle bilgi, kavrama ve uygulama basamaklarındadır. Doğru cevabı içinde barındırması nedeniyle sentez düzeyinde soru sormak olanaklı değildir. Bu nedenle bir test yalnızca çoktan seçmeli maddelerden oluşuyorsa yaratıcılığın denetlenmediği yönünde haklı bir eleştiri vardır. Doğru cevabı içinde barındırması nedeniyle puanlara şans başarısı karışma ihtimali vardır. Çok sayıda soru sorulabilir. Sistemik hata karışma ihtimali çok düşüktür. Çeldirme mantığı temellidir.

Çoktan Seçmeli Madde Yazım Kuralları:

- İdeal bir çoktan seçmeli maddede, bilen bir öğrenci, seçenekleri görmesine gerek kalmaksızın doğru cevabı verebilmelidir.
- Öğrencinin sözcük repertuarında bulunmayan, henüz öğrenmediği sözcük, kavram, terim vb. kullanılmamalıdır.
- Bir durumun resim ya da fotoğrafla anlatılabildiği durumda, çokça cümle kurmak yerine görsellerden yararlanılabilir.
- Gereksiz görsel kullanılmamasına da özen gösterilmelidir.
- Bir maddeyi okuyan bir öğrenci öğretmenin zihninden geçeni keşfetmek zorunda değildir.
- Bazı öğretmenler karışık bir dille madde yazdığında sorunun zorlaşacağını düşünür. Oysa bu oldukça yanlış bir yöntemdir. Karışık bir dil yerine ortalama bir öğrencinin anlayabileceği bir dil kullanılmalıdır. Bir çoktan seçmeli maddeyi zorlaştırmak isteyen öğretmen temelde iki yöntem kullanabilir: a. Çeldiricileri doğru yanıtı yaklaştırmak (güçlendirmek) ve b. Davranışı üst düzey becerileri ölçen bir formda hazırlamak.
- Madde kökünde gereksiz yere sözcük kullanılmamalıdır.
- Seçeneklerde gereksiz sözcük tekrarı kaçınılmalıdır.
- Bir maddenin doğru yanıtı, başka bir maddenin kökünde ya da seçeneklerinde bulunmamalıdır.
- Seçeneklerin hepsi birbirleriyle tutarlı ve ayrı gramatik yapıya sahip olmalıdır.
- İpucu vermekten kaçınmak gerekir.
- 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. Özellikle doğru cevabın daha uzun ya da daha kısa olmamasına özen gösterilmelidir.
- “Hepsi” ve “hiçbiri” bir seçenek olarak kullanılmamalıdır.
- Seçenekler bir sıra ile verilmelidir. Seçenekler baş harflerine göre alfabetik olarak, eğer sayı iseler büyüklük olarak sıralanmalıdır.
- Dil bilgisi ve imla kuralları konusunda yüksek bir titizlik ve özen göstermek çok önemlidir. **İyi bir madde yazarı:** a. madde yazılan alanı çok iyi bilmeli, b. madde yazma teknik ve yöntemlerinden haberdar olmalı, c. Maddelerin yazılacağı dili kullanmada becerikli olmalı ve d. testin uygulanacağı öğrencilerin gelişim düzeyini çok iyi bilmelidir.
- Çoktan seçmeli maddelerde seçeneklerin baş harfleri büyük olmalıdır.
- Bir madde kökü ile seçenekleri aynı sayfada olmalıdır.
- Eğer ortak köke dayalı madde grubu yazılacaksa hangi soruların ortak köke göre yanıtlanacağı belirtilmelidir. İlkokulda madde sayısını çok artırmamak gerekir.

4. ESSAY (YAZILI YOKLAMA)

Cevaplayıcıların sorulara cevap verme konusunda sınırsız bir özgürlüğü vardır. **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. **Pratik nedenler:** Hazırlaması kolaydır. b. **Eğitsel nedenler:** 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 sadece öğrencinin sahip olduğu bilgiyi değil; 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 Nesnelliği Artırma Yolları:

- ❖ Ayrıntılı bir cevap anahtarı çıkarılmalıdır. Öğrenci ne yaparsa kaç puan verileceğinin belirlenmesi gerekir.
- ❖ Bütüncül (holistik) ya da analitik rubrik hazırlanabilir.

5. KISA CEVAPLI SORULAR

Kısa cevaplı testler; öğrencinin bir sözcük, bir rakam, bir tarih ya da en çok bir cümle ile cevaplandırabileceği sorulardan oluşur. İstenen cevapların kısa ve belirgin olması, kısa cevaplı testlerin puanlanmasını genellikle nesnel kılar.

Uzun cevap gerektiren sorularda olduğu gibi bu soru türünde de ayrıntılı bir cevap anahtarı hazırlanmalıdır. Öğrencilerin verecekleri cevap çeşitliliğinin öğretmenin cevap repertuarından her zaman daha fazla olacağı göz önüne alındığında, anahtarın zümrece hazırlanması önerilebilir. **Avantajları:** a. Hazırlaması, uygulaması ve puanlaması kolay; diğer bir ifadeyle kullanışlıdır. b. Soru sayısı artırılabilir. c. Sistematik hata karışma olasılığı pek yoktur. **Dezavantajı:** İyi yazılmazsa genellikle üst düzey bilişsel özelliklerin ölçülememesi yine bir dezavantajdır.

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 fazla yeri çıkarır ve yerine bir boşluk koyar. Öğrenciden o cümleyi anlamlı bir biçimde tamamlaması istenir. Cümle tamamlama soruları boşluk doldurma olarak da geçmektedir.

Hazırlanması oldukça kolaydır. Bu yargı, özellikle kim, ne, nerede, ne zaman? sorularına cevap olabilecek olgusal bilgileri ölçmeye yönelik soruların yazılmasında geçerlidir. 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 yine bir 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.
- ✓ Maddenin cevabında birden çok ayrıntı varsa o ayrıntıların her birine ayrı puan vermek gerekir.
- ✓ Bir cümlede çok sayıda boşluk bırakılmamalı ve cümlelerden sadece anahtar niteliğindeki anlamlı ve önemli sözcükler çıkarılmalıdır.
- ✓ Herkesçe aynı biçimde algılanacak, öğrencinin “Acaba öğretmenim bununla neyi kastediyor?” sorusunu sormayacağı biçimde soru yazmak gerekir.

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.**

Sorular sözlü olarak sorulur ve cevap sözlü olarak verilir. Öğretmen ile öğrenci arasında karşılıklı ve devamlı bir etkileşim vardır. Bireysel bir test olması nedeniyle her öğrenciye ayrı soru sorma zorunluluğu vardır. Bireysel bir test olması nedeniyle tüm öğrencileri test etmek için gereken süre oldukça fazladır. Öğrencinin cevaplarını gözden geçirme şansı yoktur. Cevaplama çoğu zaman üzerinde düşünmeye ve tasarlamaya olanak olmadan verilir. Puanlama hemen yapılır. Bu ise cevabın doğruluğunun genel izlenimle yapılmasına neden olur. Öğrencinin yerinde ayağa kaldırılması ya da tahtaya kaldırılması kaygı vb. bireyden kaynaklanan hata miktarını artırabilir. 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:

- Sözlü sınavların yapılacağı gün ve saat, yazılı sınavlarda olduğu gibi ilan edilmelidir.
- Sorular önceden hazırlanmalıdır. Soruların seçiminde amaca uygunluğa ve her öğrenciye benzer güçlükte soru belirlenmesine dikkat edilmelidir.
- Cevap anahtarı hazırlanmalıdır.
- Sınavın yapılması aşamasında öğrencilerin kaygısını artıracak davranışlardan kaçınılmalıdır. Hazırbulunuşluluk artırılmalıdır.
- Basit olan sorudan başlanmalıdır.
- Kayıt altına alın(a)mıyorsa puanlama hemen yapılmalıdır.
- Puana ilişkin dönüt hemen verilmelidir.
- Puanlamada sistematik hata kaynaklarına karşı dikkatli olunmalıdır.

NOT: Yoklanacak olan davranışlar, yazılı sınav türleriyle de yoklanabiliyorsa sözlü yoklama yapmaktan kaçınılmalıdır.

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.

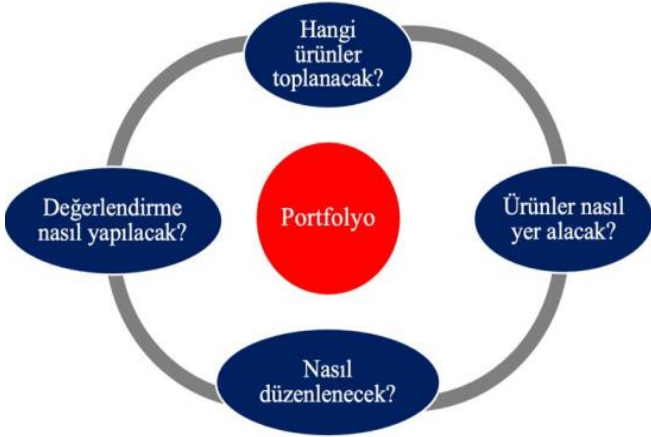
Portfolyolar ile öğrencinin öz geçmişi, ilgileri, tutumları, becerileri, yetenekleri, öğrenme biçimi, sosyal özellikleri vb. özelliklerinin tanınması amaçlanır. Sürecin paydaşları; öğrenci, öğretmen, arkadaş, aile ve iletişimde olduğu diğer kişilerdir. Öğrencinin yalnızca okulda değil; aile ortamında, oyun ortamında ve dış çevrede tanınması amaçlanır.

Portfolyolar öğrencinin yaptığı çalışmaların bir araya getirildiği herhangi bir çalışma dosyası **değildir**. 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. Başlangıç çalışmalarını, süreçteki çalışmalarını, karşılaşılan güçlükleri ve öğrenme ürünlerini içerir.

b. Ürünü yansıtan portfolyolar: Öğrenme sürecinden çok bitmiş görevleri içerir. Öğrencinin en iyi olduğunu düşündüğü çalışmaları içerir.



Şekil 15: Portfolyo değerlendirme süreci

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
- ❖ Öğrenciye yazılan mektuplar
- ❖ Öğrencinin yazdığı mektuplar
- ❖ Öğretmenin aileye ve diğer öğretmenlere yazdığı mektuplar

Portfolyo Oluşturma Süreci



A. TOPLAMA:

- Hangi çalışmaların toplanacağına ve hangi özelliklerin gözleneceğine karar verilmesi
- Öğrencilere çalışmalarının bir dosyada toplanacağını açıklanması ve çalışmalarını saklama konusunda öğrencilerin teşvik edilmesi
- Her öğrenci için sınıfta çalışmalarının toplanacağı ayrı bir kutu, dosya vb. oluşturulması

- Her bir çalışmanın ve öğretmen kaydının üzerine tarih yazılması.

B. SEÇME

Öğrenci seçimini kendi başına veya öğretmenin rehberliğinde yapabilir. Portfolyonun türüne ve öğretmenin koyduğu koşullara bağlıdır.

a. Süreci yansıtan portfolyolarda öğretmenin belirlediği konuları / gelişim alanlarını yansıtan çalışmaların seçilen örnekleri, öğrenme / gelişim sürecini yansıtacak şekilde yer alır.

b. Ürünü yansıtan portfolyolarda öğretmenin belirlediği konular ile ilgili ortaya çıkan ürünler arasından seçilenler yer alır.

C. 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:

- Bu çalışmayı nasıl yaptım?
- Bu çalışmadan ne öğrendim?
- Bu çalışmayı daha da geliştirebilir miyim? Nasıl?
- Çalışmalarım içinde en çok sevdiğim hangisi? Neden?
- Bana zor gelen bir çalışmam var mı? Varsa neden?
- Bu çalışmayı portfolyoma neden koydum?

D. SONUÇ

- a) Bu aşamada öğrenci “Bu çalışmayı niçin yaptık?” sorusunu yanıtlar.
- b) Okulda yaptığı çalışmalarla öğrendikleri arasında somut bağlar kurar.
- c) Tamamlanan portfolyo çalışmalarının öğrenci tarafından sınıf arkadaşları, öğretmeni ve ailesinden oluşan bir gruba sunumu yapılmalıdır. Portfolyonun sunumu, öğrencinin çalışmalarına önem vermesini sağlar ve kendine olan güvenini artırır.

2. PERFORMANS DEĞERLENDİRME



I. ÖZ DEĞERLENDİRME

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:

Öğrencilerin kendi performanslarını değerlendirirken yanlış davranışları söz konusu olabilir. Başlangıçta deneyimsizlik nedeniyle performansın değerlendirilmesinde yanlışlar olabilir.

II. AKRAN DEĞERLENDİRME

Öğrencinin ortaya koyduğu performansa ilişkin arkadaşlarının değerlendirmesine denir.

- ❖ Akranların değerlendirme sürecine katılması nedeniyle 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.

Olası Dezavantajları: Öğrencilerin yanlış davranışları söz konusu olabilir. Kişisel ilişkiler olumlu ya da olumsuz yönde değerlendirmeye etki edebilir. Genel izlenimle puan verme söz konusu olabilir.

III. RUBRİK (DERECELİ PUANLAMA ANAHTARI)

Performans ürünlerinin değerlendirilmesinde en sık kullanılan araçlardan biridir. Amaç, öğretmen tarafından ürünün genel izlenimle puanlamasındaki özneliliğini azaltmaktır.

Rubrikler ikiye ayrılır:

a. Bütüncül (Holistik) rubrik

b. Analitik rubrik: Üsttekine göre daha detaylı bir puanlama anahtarıdır.

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

IV. 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.

V. 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.

VI. GÖZLEM FORMLARI:

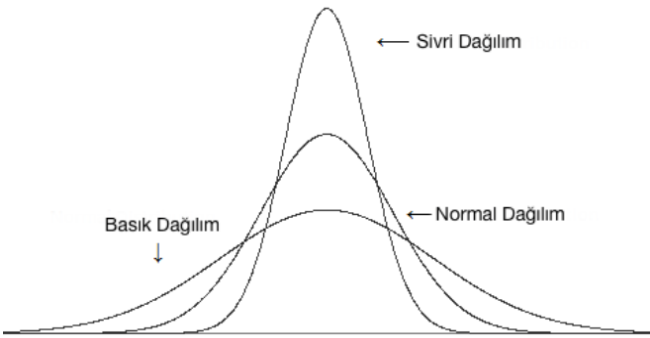
Öğrenme çıktılarının somut olarak gözlenebildiği bazı alanlarda bu yöntem oldukça kullanışlıdır. Özellikle fen derslerinde, meslek liselerinin somut performans ürünlerinin geliştirildiği vb. alanlar için oldukça uygundur. Gözlemler öğrenciler hakkında doğru ve hızlı bilgi elde edilmesini sağlar. Gözlem formları yarı yapılandırılmış biçimde olabileceği gibi tam yapılandırılmış bir formatta da olabilir.

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 değinir.
Değerlendirme öğrenmeden ayırdır.	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.

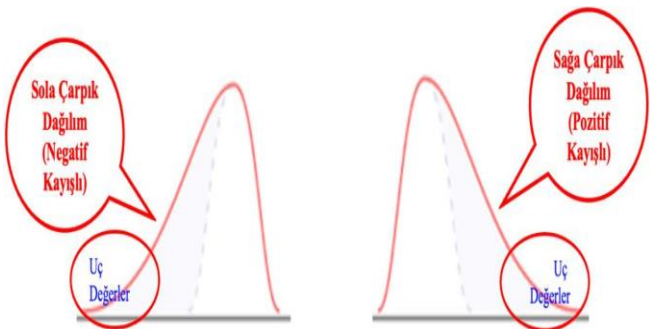
TEST İSTATİSTİKLERİ

Eğitimde en sık kullanılan istatistikler merkezi eğilim ölçüleri, değişkenlik ölçüleri, dağılım özellikleri, standart puanlar vb.dir. Merkezi eğilim ölçüleri: aritmetik ortalama, medyan (ortanca) ve moddur.

Diğer taraftan değişkenlik ölçüleri ranj (dizi genişliği), varyans, standart sapma ve çeyrek sapmadır. Değişkenlik ölçüleri grubun kabaca homojen ya da heterojen bir özellik gösterip göstermediği, standart sapma örnek olarak verilirse değişkenliğin ortalama etrafında nasıl dağıldığını gösterir. Puanların değişkenliği, grup hakkında biraz daha ayrıntılı bir değerlendirme yapılmasını sağlar.

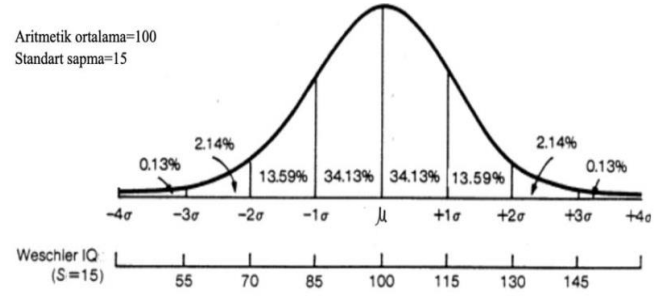


Şekil 17: Normal, sivri ve basık dağılımlar



Şekil 18: Sola ve sağa çarpık dağılımlar

DİKKAT: Bir test zorlaştıkça dağılımın sağa çarpık, kolaylaştıkça sola çarpık olma olasılığı artar.

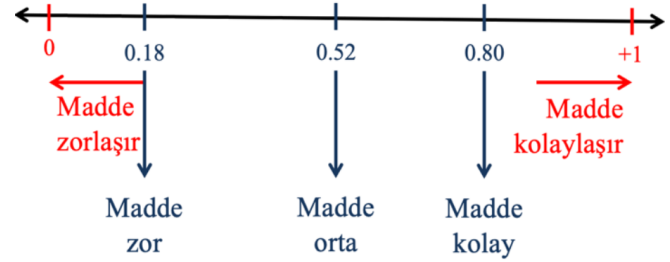


Şekil 19: Standart normal dağılım altında yer alan alanlar

MADDE İSTATİSTİKLERİ

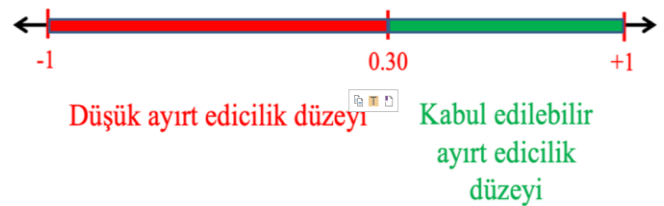
Ölçme işlemi sonucunda bireylerin madde puanları üzerinden hesaplanan istatistiklerdir. Eğitimde en sık kullanılan istatistikler madde güçlük indeksi, madde ayırt edicilik indeksi, madde güvenilirliği vb.dir.

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



Şekil 20: Madde güçlük indeksi

Madde ayırt edicilik indeksi ise maddenin bilenle bilmeyeni ayırt edip edemediğine ilişkin bir istatistiktir. -1 ile 1 arasında değer alır ve genellikle alan yazında 0.30 ile 1 arası ayırt edici olarak nitelendirilir.



Şekil 21: Madde ayırt edicilik indeksi

EĞİTİM İZLEME ARAŞTIRMALARI

Başarı, konu alanında verilen eğitim sonucunda öğrencilerin beklenen kazanımlara, becerilere ya da davranışlara sahip olma düzeyini ifade eden bir kavram.

Beceri ise başarıdan farklı olarak tekil kazanımlardaki erişim düzeylerinden ziyade öğrencilerin bilgi ve becerilere sahip olduktan sonra bunları ne kadar kullanabildiğine dair bir gösterge olarak değerlendirilmektedir.

Okuryazarlık, özellikle belirli bir alanda ilgili doğru bilgileri bulma, bu bilgileri yanlış bilgilerden ayıklama, bunun için temel okuryazarlık becerilerini kullanma, ardından doğru bilgileri seçerek ve uygun analiz yöntemlerini kullanarak gerçek ya da tasarlanmış durumlarda kullanma becerisi olarak tanımlanmaktadır.

Uluslararası İzleme Çalışmaları

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.

Aslında bu tarz izleme çalışmalarına ihtiyaç duyulmasının ve bu çalışmaların popüler olmasının sebeplerinden ilki karşılaştırılabilir veri ihtiyacıdır. İzleme araştırmalarının sağladığı ikinci katkı, ihtiyaç duyulan bağlamsal ve bütüncül ilişkilere dair bulgular sunmasıdır. Son olarak izleme çalışmaları, eğitim politikalarının bir bütün olarak oluşturduğu etki konusunda da önemli bulgular sağlamaktadır.

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.

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.

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. 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. 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. PISA, 2000 yılından bu yana uygulanmakta ve her üç yıllık döngüde bir alana ağırlık verilmektedir. Türkiye 2003 yılından itibaren dahil olmuştur.

Her döngüde bir alana ağırlık verilmekle beraber matematik okuryazarlığı, fen okuryazarlığı ve okuma becerileri alanlarının hepsine dair ölçme yapılmaktadır. Örneklem seçimi yedi coğrafi bölgeye göre değil, Türkiye'nin sosyoekonomik farklılıklarını da dikkate alan 12 bölgeden oluşan İBBS Düzey-1 istatistiksel örnekleme sistemine göre yapılmaktadır. Bu örnekleme sürecinde bölgesel öğrenci yoğunlukları dikkate alınmakta ve daha büyük illerimizin olduğu bölgeler daha fazla öğrenciyle, daha az olduğu yerler ise daha az öğrenciyle temsil edilmektedir.

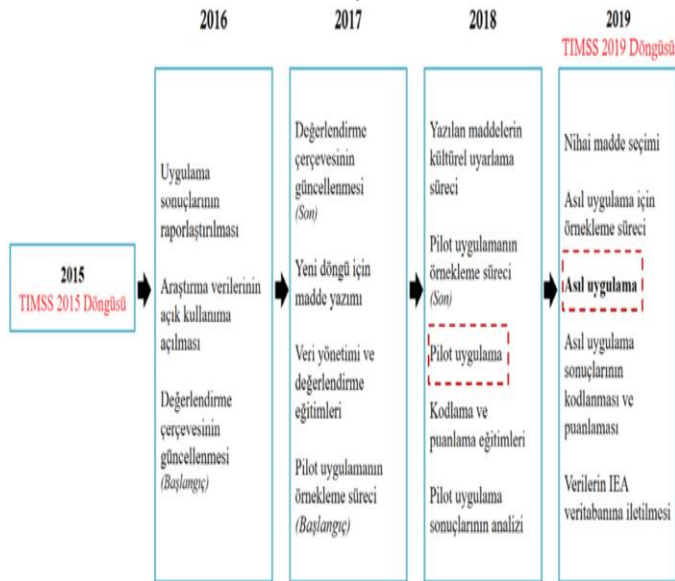
PISA'nın en önemli çıktılarından bir tanesi öğrencilerin hangi yeterlik düzeyinde olduğuna dair veriler sunmasıdır. Yeterlik düzeyleri son derece önemlidir çünkü her bir yeterlik üzerinde bulunan öğrencinin o alanda ne yapıp ne yapmadığını, davranış bazında hangi davranışları gösterip hangi davranışları gösteremediğine dair somut açıklamalar yapmaktadır.

Özetle 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.

TIMMS KAPSAMI VE SONUÇLARI

TIMSS, IEA tarafından dörder yıllık döngüler şeklinde gerçekleştirilmektedir. **PISA ile TIMSS arasındaki en 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.



Şekil 5'te de görüldüğü gibi her uygulama döngüsünde uygulama öncesi ve sonrasında yapılan işlemler kendi içerisinde birbirini besleyen süreçler içermektedir. 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. Burada PISA'da da kullanılan İBB-düzy-1 bölgeleri kullanılmaktadır. Ardından seçilen okullar içerisinde 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. Sonuç olarak TIMSS asıl uygulaması yaklaşık yarım gün içerisinde tamamlanmaktadır.

TIMSS'te 500 puan önemli bir ölçütü temsil etmekte olup bir uygulamadan diğerine aynı beceriyi ifade etmektedir. Türkiye ilk defa 4. sınıf seviyesinde 500 puanı aşmıştır. PISA'da önemli ölçüde puanlarımız artmış olmakla birlikte OECD ortalamasına henüz ulaşamamak da TIMSS'in eğitim programıyla daha ilişkili olması sebebiyle müfredata dayalı değerlendirmelerde öğrencilerimiz daha yüksek performans göstermektedir.

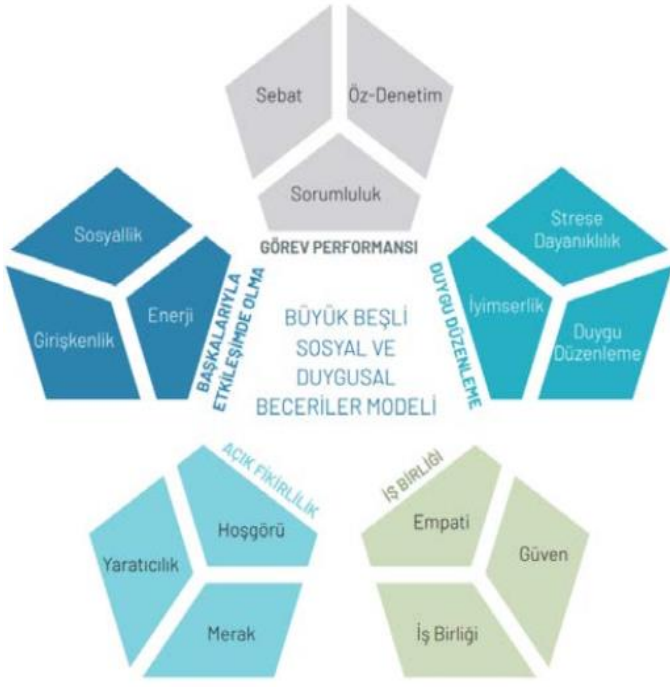
PIRLS KAPSAMI VE SONUÇLARI

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 katılımının 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.

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. 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, Şekil 7'de 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 başarma motivasyonu ve öz yeterlik algısıdır.

Bu çalışmayı özel kılan diğer bir özelliği de veri çeşitlenmesinin (data triangulation) kullanılmasıdır. Veri çeşitlenmesiyle kastedilen aynı sorular üzerinden hem öğrencilerden hem öğretmenlerinden hem de velilerinden bilgi toplanmasıdır ki bu tüm verilerin birleştirilerek sonuçların daha güvenilir hâle getirilmesine katkı sağlamaktadır. Veri çeşitlenmesi kullanılmasının sebebi utangaçlık ya da kendine güven gibi duyuşsal becerilere ilişkin soruları, öğrencilerin oldukları gibi değil olmak istedikleri veya görünmek istedikleri şekilde cevaplandırma eğilimlerinin önüne geçebilmek; daha gerçekçi sonuçlara ulaşabilmektir.

Bu çalışmaya katılan ülkeler ve şehirler şu şekildedir: Kanada (Ottawa), ABD (Houston), Kolombiya (Manizales ve Bogota), Portekiz (Sintra), Finlandiya (Helsinki), Türkiye (İstanbul), Rusya Federasyonu (Moskova), Güney Kore (Daegu) ve Çin Halk Cumhuriyeti (Suzhou).

Özetle çıkarmamız gereken en önemli sonuç, sosyal ve duygusal beceriler anlamında kız ve erkek öğrencilerin önemli farklılıklara sahip olduğu ve bu becerileri geliştirmeye dönük her türlü müdahalede cinsiyetler arası farkların dikkate alınması gerekliliğidir.

Sosyal ve duygusal beceriler bağlamında bireyin öz yeterliğini geliştirmek, her türlü konuda onu daha yeterli hissettirmek veya kendine yönelik yeterlik inancını artırmak önemlidir ama bunlar akademik başarı ile ne kadar ilişkilidir? Burada iki özellik ön plana çıkıyor: *Birinci özellik olan merak*, hem okuma hem de matematik becerileriyle anlamlı bir ilişkiye sahiptir. *İkincisi de sebat*, yani aynı konu üzerinde öğrenene kadar sabırla çalışmaktır.

Okul iklimi ile ilgili çıktılar ise rekabetçi okul iklimi, veli beklentileri ve öğretmen beklentilerini kapsamaktadır. Akademik çalışmalardan, öğrencilerden beklenti arttıkça (beklentinin gerçekçi olması koşuluyla) öğrencilerin akademik performansları ve kendi öz yeterlik algılarının arttığı bilinmektedir. Veli beklentileri ve öğretmen beklentileri arttıkça öğrencinin psikolojik iyilik hâli ve sınav kaygısı da artmakta ancak bir taraftan da psikolojik kaygıları ve yaşam doyumları kısmen artmaktadır. Diğer taraftan rekabetçi okul iklimi ve okul içi rekabeti destekleyen politikalar arttıkça öğrencilerin özellikle psikolojik iyilik hâlinde önemli eksikler olduğu görülmektedir. öğrencilere yaşam doyumunu sağlayacak, onları kabul edecek olumlu okul iklimi yaratmak son derece önemlidir. Okullarımızda zorbalığa maruz kalma, büyük oranda öğretmen ve öğrenci ilişkisi bağlamında çözülebilmektedir. Öğretmen ve öğrenci ilişkisi arttıkça, daha etkileşimli hâle geldikçe zorbalığa maruz kalma durumu azalmaktadır. Öğretmen ve öğrenci ilişkisi arttıkça okula aidiyet de artmaktadır. İstanbul, katılımcı şehirler arasında öğrencileri okula aidiyet oranı yüksek şehirlerden biridir.

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. Son olarak 2019 yılında ilk defa tamamen bilgisayar ortamında gerçekleştirilen okuma, dinleme, yazma ve konuşma olmak üzere dört temel becerinin hepsini kapsayan Türkçe dil becerilerinin izlenmesi araştırması yapılmış ve yaygınlaştırılmaktadır.