



Recep Tayyip Erdoğan Üniversitesi
Eğitim Fakültesi

Pedagojik Formasyon Eğitimi SERTİFİKA PROGRAMI

ÜNİTE 2

ÖLÇME ARACINDA BULUNMASI GEREKEN NİTELİKLER

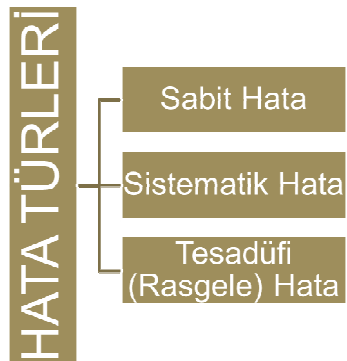
Doç. Dr. Ahmet TEKBİYİK

ÖLÇMEDE HATA

Somut ya da soyut niceliklerin ölçülmesinde daima hata yapma ihtimali vardır,

Hata kaynakları:

- Ölçülen özellik
- Ölçülen bireylerin özellikleri
- Ölçmenin yapıldığı ortam
- Ölçme aracı
- Ölçmeyi yapan kişi



1. Sabit Hata: Bir ölçmeden diğerine değişmeyen hata türüdür.

Örnek: Tüm öğrencilerin sınav notlarına 10 puan eklemek,
Baskül'ün her defasında 2 kg eksik tartması
Bir soruyu tüm sınıfın cevapsız bırakması

Sabit hataların miktarı ve kaynağı bellidir.

Bu nedenle güvenilirliği etkilemez, geçerliği doğrudan etkiler

Sabit hata ortalamayı değiştirir

Standart sapmayı etkilemez,

2. Sistemik hata: Miktarı değişken olan ve tüm ölçümlere aynı oranda etki etmeyen hatadır.

Ölçmeyi yapan kişiye, ölçülen niceliğe ve ölçme koşullarına bağlı olarak değişebilir.

Ölçmede amaçlanan değişkenin dışında başka bir değişkenin de ölçülmesi durumunda görülebilir

Örnek:

Sıcaklıktan etkilenen bir şerit metreyle farklı sıcaklıklarda ölçüm yapılması,

Bir terazinin her 1 kg da 25 gr fazla tartması

Öğretmenin kendisine yakın gördüğü öğrencilerin sınav kağıtlarını okurken fazla puan vermesi

Her sınav kağıdına %10 puan eklemesi

*Sistemik hatalar güvenilirliği etkilemezken, geçerliği düşürür.
Ortalamayı da etkiler*

3. Rasgele Hata: Kaynağı ve miktarı belli olmayan hatalardır. Ölçmeyi yapan kişiden, ortamdaki veya şans başarısından kaynaklanabilir.

Hatanın kaynağı, yönü ve miktarı bilinemediğinden arındırılmaz.

Örnek;

Dikkatsiz puanlama

Sınıf ortamının elverişsizliği

Öğrencilerin duyuşsal özellikleri (Sınav kaygısı vb.)

Rasgele hatalar öncelikle güvenirliliği ve dolaylı olarak geçerliliği düşürür

Kopya çekme davranışı ölçmede hata kapsamına girer mi?

BİR ÖLÇME ARACINDA BULUNMASI GEREKEN ÖZELLİKLER

Bir ölçme ve değerlendirme işleminde verilen kararların doğruluğu ve yerindeliği, **kararın dayandığı ölçümlerin hatasız ve ölçütün uygun olmasına bağlıdır.**

Bu nedenle ölçme araçlarında bazı özelliklerin aranması gerekmektedir.



GÜVENİRLİK

Bir ölçme aracının ölçtüğü özelliğin hatalardan arınık olarak ölçebilme derecesidir

Güvenilir bir ölçme aracı **Tutarlı, Kararlı, Hassas ve Objektif** ölçümler yapar

1. Kararlılık

Ölçülen özelliğin aynı ölçme aracıyla farklı zamanlarda birkaç kez ölçüldüğünde birbirine yakın değerler vermesi

2. Tutarlılık

Bir ölçme aracındaki maddelerin birbirine ve testin bütününe uyumluluğu (iç tutarlılık)

3. Duyarlılık (Hassasiyet)

Birim aralığının mümkün olduğunca küçük olması

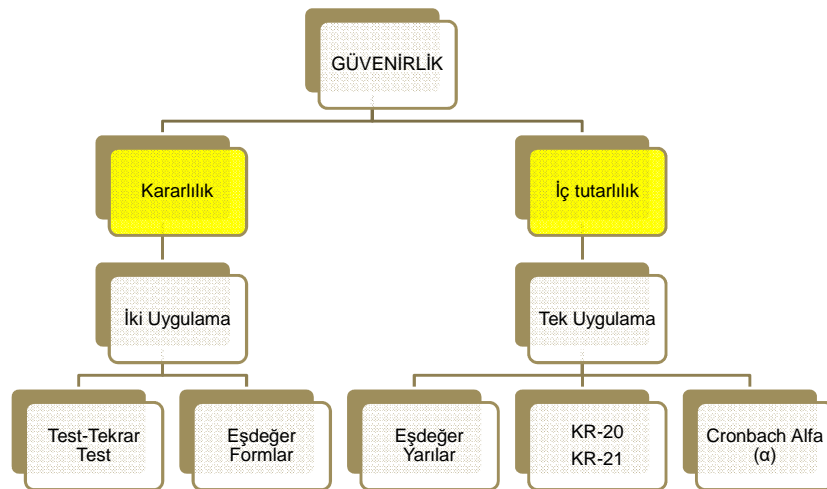
4. Objektiflik

Farklı Puanlayıcıların ölçme aracından aynı sonuçları elde etmesi

Güvenirliđi Etkileyen Özellikler

- 1.Soru sayısının fazla olması güvenilirliđi artırır
- 2.Puanlama objektifliđi: Birden fazla puanlayıcının aynı sonuçlar elde etmesi güvenilirliđi artırır
- 3.Testin uygulama koşullarının elverişli olması güvenilirliđi artırır
- 4.Soruların orta güçlükte ve yüksek ayırt edicilikte olması güvenilirliđi artırır
- 5.Maddelerin homojen olması, benzer konulardan oluşması güvenilirliđi artırır
- 6.Şans başarısının en aza indirilmesi güvenilirliđi artırır

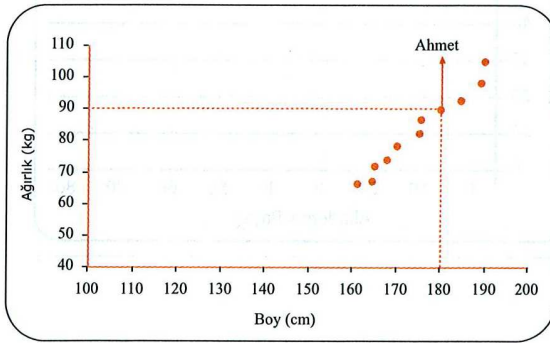
GÜVENİRLİK TAHMİN YÖNTEMLERİ



KORELASYON KAVRAMI

Korelasyon: İki değişken arasındaki ilişkinin miktarını ve yönünü belirler

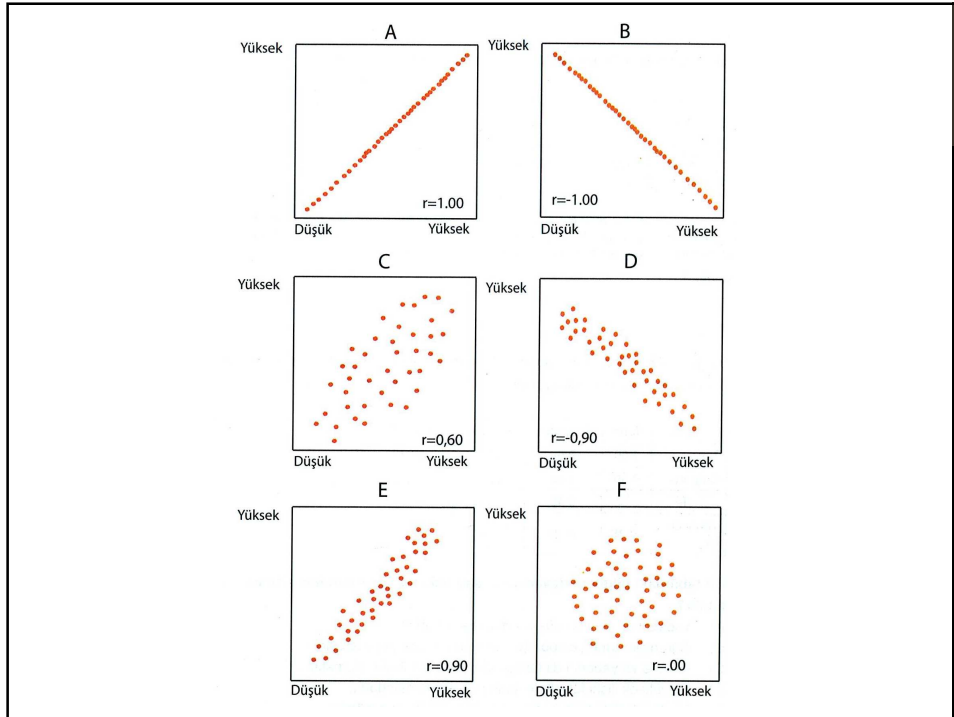
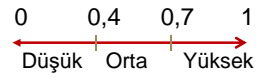
Korelasyon katsayısı -1 ↔ +1 aralığında değer alır



+1: Güçlü olumlu ilişki

0: İlişki Yok

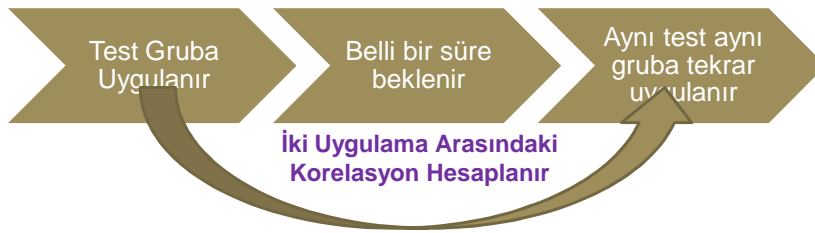
-1: Güçlü olumsuz ilişki



Test-tekrar test yöntemi (Kararlılık)

- 1 test-2 uygulama

0,70 ve üzeri yüksek güvenilirlik



Eşdeğer/Paralel Formlar Yöntemi (Kararlılık)

2 test-2 uygulama

Aynı özelliği ölçen iki test

0,70 ve üzeri yüksek güvenilirlik



Testi Yarılama/Eşdeğer yarılar yöntemi (İç tutarlılık)

1 test-1 uygulama

Test bir gruba bir kez uygulanır, iki yarıya ayrılır

- Rasgele iki bölüme
- Tam ortadan ikiye bölme
- Tek numaralı sorular-çift numaralı sorular



Güvenirlik için Spearman Brown Formülü

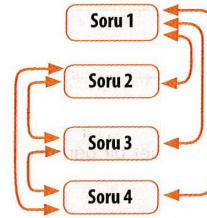
$$\text{Tüm testin güvenilirliği} = \frac{2r}{1+r}$$

r: Eş yarılar arasındaki korelasyon

Örneğin iki yarı arasından elde edilen korelasyon katsayısı 0.70 ise bu testin güvenilirliği:

$$\text{Güvenirlik} = \frac{2 \times 0.70}{1+0.7} = 0.82 \text{ olarak bulunur.}$$

- **Kuder-Richardson (KR-20 ve KR-21)**
- **ve Cronbach Alfa (İçtutarlılık) Yöntemi**
 - 1 test, 1 Uygulama
 - Maddeler arası uyuma yöneliktir



- **KR-20: Maddeler 1 veya 0 şeklinde kategorik kodlanmışsa**
- **KR-21: Maddeler farklı ağırlıkta puanlanmış veya düzeltme formülü kullanılmışsa**
- **Cronbach α: Doğru cevabı olmayan maddelerde (1-Katılmıyorum, 2- Kararsızım, 3-Katılıyorum ...)**

KR-20, KR-21 ve Cronbach Alfa'nın Yüksek Olması Durumunda Yapılabilecek Yorumlar:

- Test tesadüfi hatalardan arınıktır
- Ölçülen özellik tek boyutludur
- Testi oluşturan maddeler homojendir, maddeler arası uyum yüksektir
- İç tutarlılık yüksektir
- Testin yapı geçerliğinin yüksek olması muhtemeldir

Ölçmenin Standart Hatası

- Bir ölçme aracında tesadüfi hata miktarı tahmin edilemez. Ancak her bir öğrenci başına düşen tesadüfi hata miktarı hesaplanabilir.
- Standart hata teste karışan hatanın test puanı cinsinden miktarını verir
- Ölçmenin standart hatası arttıkça güvenilirlik azalır

S_e = Standart Hata

r = Güvenirlik katsayısı

S_x = Standart sapma

$$S_e = S_x \sqrt{1 - r}$$

Puanlayıcılar Arası Güvenirlik (Objektiflik)

Subjektif testlerde iki farklı puanlayıcının verdiği puanların uyumlu olması puanlama güvenilirliğini artırır.



GEÇERLİK

Bir ölçme aracının ölçmek istediği özelliği başka özelliklerle karıştırmadan ölçebilme derecesidir. Bir anlamda testin uygunluğudur.

Bir ölçme aracının veya yönteminin onunla ölçülmek istenen değişkeni ölçüp ölçmediği; ölçebiliyorsa, onu başka değişkenlerden ne derece arınık olarak ölçtüğüdür (Turgut, 1993)

!!! ölçme araç ve yöntemlerinin nitelikleri arasında en önemlisi geçerliktir.

✓Geçerlik, ölçme aracının amaca hizmet etme derecesidir. Bir ölçme aracının geçerli olabilmesi için onun sadece ölçmek için düzenlendiği özelliği ölçmesi, yani onu başka özelliklerle karıştırmadan ölçmesi gerekir.

✓Eğitimde kullanılan araç ve yöntemlerle ölçülmek istenen değişkenler bireylerin yeteneği, başarısı gibi deyimlerle tanımlanmaya çalışılan değişkenlerdir. Aracın veya yöntemin ölçülmek istenen değişkenin ölçüsü olabilecek bir puan vermesi, bu ölçüyü başka değişkenlerle karıştırmaması beklenir. Bir araç veya yöntem, bu niteliği sağlayabileceği derecede geçerli puanlar verir.

✓Bir testin geçerliği bir derece sorundur. Bir test geçerli veya geçersiz olmaktan öte daha çok veya daha az geçerli olabilir.

21

Testin geçerli olması için güvenilir olması gerekir, ancak güvenilir olmak geçerliği garanti etmez.

Bir test geçerli ise;

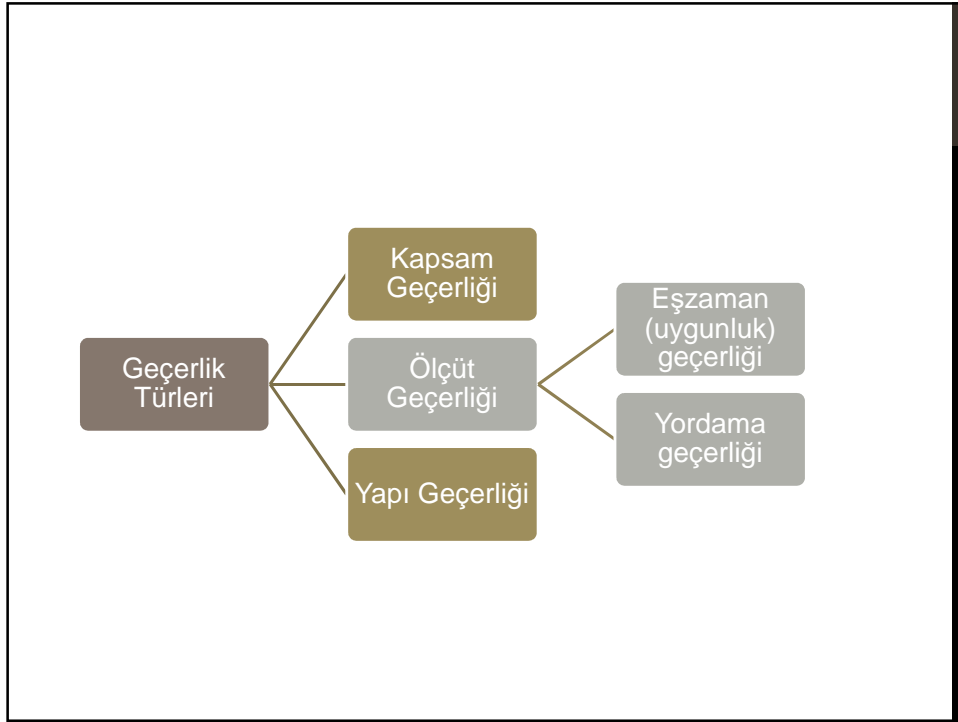
Test amaca hizmet etmektedir,

Sorular öğretimin hedeflerini ve konuları temsil etmektedir

Test ölçmek istediği özellikleri başka değişkenlere karıştırmadan ölçmektedir



22



1. KAPSAM GEÇERLİĞİ:

Ölçme araç ve yöntemlerinin kapsamlarına ve içeriklerine bakılarak varılan geçerlik yargılarına kapsam geçerliği denir. Bir bütün olarak testin ve testteki her bir maddenin amaca ne derece hizmet ettiği. Bir testin kapsam geçerliği;

- Testteki toplam maddelerin ölçülecek davranışları ve konu içeriğini örnekleme derecesine,
- Testteki sorulan soruların ilgili davranışları ölçmede uygun olup olmadığına bağlıdır.

Kapsam Geçerliğinin belirlenmesinde iki yol izlenebilir

1.1. Mantıksal yol

-Belirtke tablosu hazırlama: İlk sütunda konu ana başlıklarının, ilk satırda ise davranışların bilişsel boyutunun yer aldığı tablodur.

	Davranış Düzeyleri				
	Bilgi	Kavrama	Uygulama	Toplam	Yüzde
Üniteler					
1.Giriş					
2.Ölçmenin Temelleri					
3. Güvenirlik ve Geçerlik					
4. Yazılı Yoklama					
5.Kısa cevaplı testler					
6. Doğru-yanlış testler					
7. Çoktan seçmeli testler					
8. Sözlü sınavlar					
Yüzde					

- Uzman görüşü alma: Konu alanı uzmanlarının görüşlerinin alınması yöntemidir. Hazırlanan ölçme aracı uzmanlara gösterilerek ölçülecek konu alanını yeterince temsil edip etmediğinin sorulmasıdır.

25

2.1. İstatistiksel yol Aynı alanda daha önce geliştirilmiş, aynı davranışları ölçen ve geçerliği yüksek bir test ile yeni hazırladığımız iki testin aynı gruba uygulayıp öğrencilerin bu testlerden elde ettikleri puanları arasındaki korelasyona (olumlu ilişkiye) bakma yöntemidir.

2. Ölçüt geçerliği:

İstatistiksel teknikler kullanılarak ve bilinenlerden yararlanılarak bilinmeyen durumlar hakkında yapılan geleceğe yönelik tahminlerde bulunma işlemidir. Bir tahmin işidir.

- ❑ Testten elde edilen puanlarla testin yordamak için düzenlendiği değişkenin doğrudan ölçüsü olan ve daha sonra elde edilen ölçüt arasındaki korelasyon da yordama geçerliğidir. Korelasyon katsayısının büyüklüğü yüksek yordama geçerliğine işaret eder.
- ❑ Eğitimde bir öğrencinin ya da grubun gelecekteki başarı durumunun nasıl olacağını tahmin etmek için kullanılır. Örneğin; YGS-LYS sınavları, uygulanan testlerden elde edilen puanlar, adayların üniversitedeki ve gelecekteki başarılarının yordayıcısı olarak kullanılmaktadır. Bu amaçla kullanılan testlerin amaca hizmet etme derecesi yani yordama geçerliğinin olması gerekmektedir.

27

Ölçüt geçerliği, iki grupta incelenebilir;

2.1.Yordama geçerliği

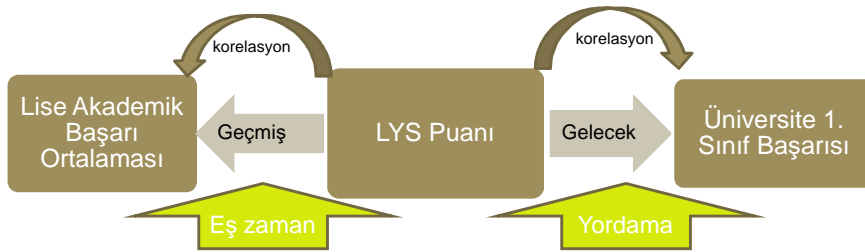
- ✓ Yordama geçerliği aranan test uygulanır, o testin yordayacağı özellik ya da puan elde edinceye kadar beklenir. Bir okulun giriş sınavında kullanılan testin o okuldaki başarıyı yordama derecesine bakmak istediğimizde testi uyguladıktan sonra o öğrencilerin okulda ders notları ortalamalarını beklememiz gerekmektedir.
- ✓ Söz konusu puanlar geldikten sonra iki puan grubu arasındaki korelasyona bakılır. Elde edilen korelasyon değeri giriş sınavında kullanılan testin girilen okuldaki başarıyı yordama derecesi olarak kabul edilir. Öğrencilerin giriş testi puanları yordayıcı, öğrencilerin başarıları ise ölçüttür.
- ✓ Yordama geçerliği, özellikle seçme amacıyla kullanılan testlerde aranması gereken bir özelliktir.
- ✓ Örneğin personel seçmek amacıyla yapılan bir ölçmenin yordayıcılığı yüksek ise personelin iş başarısının da yüksek olabileceği yorumu yapılır

28

2.2. Uygunluk geçerliği (mevcut hal, eş zaman geçerliği):

Yeni bir ölme aracı geliştirildiğinde, bu aracın sonuçlarının geçerliği olup olmadığını bulmak için, yeni geliştirilen test ile ilişkili olan ve daha önce uygulanmış bir testin sonuçları ölçüt olarak kullanılabilir. Bu durumda iki test sonucu arasındaki korelasyona ilişkin kanıtlara bakılır.

Örneğin LYS puanlarının yordama ve uygunluk geçerliği belirlenmek isteniyorsa,



29

3. Yapı geçerliği: ölçme işleminin teorik bir yapıyı (özelliği) ölçüp ölçmediğinin belirlenmesidir.

- ✓ Birbiri ile ilgili olduğu düşünülen belli öğelerin yada öğeler arasındaki ilişkilerin oluşturduğu bir örüntüye **yapı** denir.
- ✓ *Yapı geçerliği ile maddelerin dersin yapısını oluşturan özelliklere ve birbiriyle olan uyumuna bakılır*
- ✓ Bu, bir ölçüde bilimsel kuram geliştirme süreciyle aynıdır.
- ✓ Eğitimde ölçmeye çalışılan bazı özellik yada değişkenler birer kuramsal yapıdır.
- ✓ Örneğin zeka türü ölçülecekse ölçme aracının 8 ayrı alt boyutlu bir yapıya sahip olması beklenir.

30

TEST TÜRLERİNDE ARANAN GEÇERLİK-GÜVENİRLİK ve GÜÇLÜK

	BAŞARI TESTİ	SEÇME SINAVI	İZLEME TESTİ
Kapsam Geçerliği	Yüksek	Orta veya Yüksek	Yüksek
Yordama Geçerliği	-----	Yüksek	-----
Güvenirlik	0,70 ve üzeri	0,70 ve üzeri	0,70 ve üzeri
Güçlük	Orta düzeyde 0,40-0,60 arası	Test zor olmalı 0,40 ve altında	Kolay olmalı 0,60 ve üzeri

Görünüş geçerliği: Bir testin gerçekten ne ölçtüğüyle değil, onun ne ölçüyor görüldüğü ile ilgilidir. Bir testin görünüş geçerliği, o testin ölçmek istediği şeyi ölçüyor görünmesidir. Görünüş geçerliği olan bir test kapağında ne yazılıysa onu ölçüyor görünür.

Örneğin; kapağında "Fizik testi" yazılı olan bir testin içinde fizikle ilgili sorular varsa söz konusu testin geçerliği vardır denilir. Bunun için konu alanı uzmanlarının görüşü yeterlidir.

Kullanışlılık

Araç ve yöntemin uygulanabileceği eğitim seviyesi, uygulama süresi, ölçmeden istediği beceriler, maliyet ve ekonomi gibi etkenlerin o araç ve yöntemin kullanılmasına etkileri *kullanışlılık* olarak ifade edilmektedir. Kullanışlılığı artıran önlemler güvenilirliği ve geçerliği de dolaylı olarak artırır. Ölçme araç ve yöntemlerinin kullanışlılığı hakkında bir yargıya varırken göz önünde bulundurulması gereken noktalar şöyledir;

Ekonomi; Araç ve yöntem maliyet yönünden ekonomik olmalı, öğretmene, öğrenciye, yöneticiye ağır mali yük getirmemelidir.

Hazırlama süresi; Her araç ve yöntem bir hazırlık gerektirir. Testlerde bu süre daha uzun, yazılı yoklamada daha kısadır. Hazırlama süresi kısa olan bir sınav tercih edilmelidir. Yalnız kalabalık gruplarda yazılı yoklamaların cevaplarının puanlanması zorluğu da göz önünde bulundurulmalıdır.

33

Uygulama süresi; Bir yöntemin kullanışlılığı, verdiği puanın önemine oranla, uygulamada gerektirdiği zaman azaldıkça artar.

Hazırlayıcı ve uygulayıcıların nitelikleri; Bir aracın uygulanması ne kadar özel beceriler gerektirirse, kullanışlılığı da o oranda sınırlanır.

Cevapların Nitelikleri; Öğrenciler cevaplama yönergesini kolayca anlayabilir ve cevaplarını kolayca kaydedebilirlerse, o araç kullanışlıdır.

34

Uygulama kolaylıkları; Testlerde sayfa düzeni, baskı mükemmelliği, cevap kağıdının düzeni, yazılı yoklamada soruların yazılı verilmesi gibi konular uygulama kolaylığı o araçların kullanılışlığını artırır.

Puanlama kolaylıkları; Testlerde cevap kağıdının ayrı olması, iyi düzenlenmesi, cevapların optik okuyucu ile puanlanabilmesi, puanlama anahtarının olması kullanılışlığı artırır.

Puanları yorumlama kolaylıkları; Bir araç yorumlamaya yardımcı, yorumlamayı kolaylaştırıcı ve yorumda hatayı azaltıcı araçlarla beslenirse daha kullanışlı olur.

35

Geçerliği Etkileyen faktörler

Ölçme yöntemi ve soru sayısı: Bir testte kullanılan yöntem (yazlı, yoklama, sözlü yoklama, testler) öğrencinin alacağı puanı (hata miktarını) etkileyebilir. Testteki soru sayısının fazla olması, kapsam geçerliğini artırır.

İstenmeyen özelliklerin ölçülmesi: Her soru, o araçla ölçülmek istenen davranışlardan en az bir tanesini yoklamalı, başka özellikleri ölçmemelidir.

Örneğin; matematik bilgisini ölçerken yazı güzelliği, tertip düzen gibi özelliklerin ölçülmesi geçerliği düşürür.

Ölçülecek Özelliği Tam Olarak Ölçen Sorular:

Her soru, o soruyla ölçülmek istenen bilgi ve becerilere sahip öğrencilerin tereddütsüz doğru cevaplandırılabilceği ve bu bilgi, beceri ve yeteneklere sahip olmayanların doğru cevaplandırılmayacağı nitelikte olmalıdır.

Puanlayıcı yanlılığı:

36

Kapsamlı ve dengeli dağılım: Sınav belli bir sürede okutulan konular ve o konulara dayanılarak geliştirilmesi beklenen bilgi, beceri ve yetenekler bakımından hem kapsayıcı hem de dengeli olmalıdır.

Güvenirlilik: Herhangi bir puanın geçerliği için ön şart, onun güvenirliliğidir. Güvenirlilikten yoksun bir puan sadece tesadüfi hataları gösterir. Hatada geçerli ölçü olamaz. Her ölçme sonucu güvenirliliği oranında geçerlik kazanır. Fakat güvenirliliğin sağlanmış olması geçerliğinde mutlaka sağlanacağı anlamına gelmez.

Sınavın güçlük derecesi: Sınavın ve ölçme araçlarının güçlüğü de geçerliğe etki eder. Güçlük derecesi o sınavdan elde edilecek puanların kullanılacağı amaca göre ayarlanmalıdır.

37

- ❑ Aynı sınavların her yıl kullanılması, kopya, sınav süresinin yetersizliğinden boş bırakma, soruları sınavdan önce açıklama, derslerde sınav sorularının çözülmesi, öğrencilerin çeşitli nedenlerle sınavı yanda bırakması, gürültü, hastalık gibi nedenlerle sağlıklı bir cevaplandırmanın yapılmayışı gibi etkenler güvenirliliği ve geçerliği düşürür.
- ❑ Hedeflere uygun soru hazırlamak, etkili bir sınav planı geçerliği artırır.

38