



Ders 5: ÖLÇME VE DEĞERLENDİRME

Prof. Dr. Tevhide Kargın

Ölçme ve Değerlendirme

- Ölçme (measurement), bireylerin ya da nesnelerin belirli özelliklere sahip olup olmadığının,
- sahip ise, sahip oluş derecesinin belirlenerek sonuçların sembollerle ve özellikle de sayı sembolleriyle ifade edilmesidir.
- Ölçme, bir betimleme (tanımlama) işlemidir.



- Deęerlendirme (evaluation) ise, ölçme sonuçlarını bir ölçütle kıyaslayarak ölçülen nitelik hakkında bir karara varma sürecidir.
- Deęerlendirme, bir yargılama işlemidir ve ölçme sonucunun bir ölçütle karşılaştırılmasına dayanır (Tekin, 1996; Turgut, 1997).
- Ölçme işinde kullanılacak araçların belli niteliklere sahip olmaları beklenir.

Ölçme (Test etme)/Değerlendirme Sonuçlarını Etkileyen Faktörler

- Çocukla ilgili faktörler
- Ölçme/değerlendirme yapanla ilgili faktörler
- Testin kensisyle ilgili faktörler



Çocukla ilgili faktörler

- Kaygı ve motivasyon
- Önceki yaşantı
- Sağlık ve duygusal durum



Ölçme/değerlendirme yapanla ilgili faktörler

- Uygulama ve yorumlama farklılıkları
- Değerlendirme yapanın yanlılığı, etnik kökeni, dil ve kültür



Testin kendisiyle ilgili faktörler

- Geçerlik
 - Kapsam geçerliği
 - Yordama geçerliği
 - Yapı geçerliği
 - Görünüş geçerliği
- Güvenirlik
 - Test tekrar test
 - Paralel Formlar
 - İki yarı
 - KR 20



GEÇERLİK (VALIDITY)

- Geçerlik kavramı çeşitli yazarlar tarafından farklı biçimlerde tanımlanmıştır.
- Özgüven(1994. s,97), "bir ölçme aracının ölçmek üzere hazırlandığı amacı, ölçme derecesi",
- Tekin(1993. s,42), bir ölçme aracının ölçmeyi amaçladığı özelliği başka herhangi bir özellikle karıştırmadan doğru olarak ölçebilme derecesi",
- Hakan ve diğerleri(1991. s,267) "bir ölçme aracının ölçtüğü niteliklerde bir değişme olmadığı sürece birbirine yakın ölçüler verme özelliği",
- Karasar(1986. s,158), "ölçmek istenen şeyin ölçülebilmiş olma derecesi, ölçülmek istenenin başka şeylerle karıştırılmadan ölçülebilmesi"

- Demirel ve Ün(1987. s,68), "bir ölçme aracının ölçmek istenen niteliği ölçme derecesi", ve Yıldırım(1983. s,134), "ölçeğin kullanım amacına uygunluk niteliği", şeklinde tanımlamaktadır.
- Bu tanımlar incelendiğinde; ölçme aracının ölçmek istediği özelliği/davranışı tam ve doğru olarak, başka özelliklerle/davranışlarla karıştırmadan ölçmesine o aracın geçerliği olarak değerlendirildiği anlaşılmaktadır.

- Bir testin geerliđi ile gvenirliđi arasında yakın bir iliřki vardır. Geerlikte gvenirlik n kořuldur. Yani bir test geerli olmak iin gvenilir olmak zorundadır.



- Karasar(1984. s,159), geçerliđi; içerik, uygulama (deneysel) ve yapı geçerliđi olmak üzere üç türde incelerken
- Tekin(1993. s,43) ise kapsam, yordama, yapı ve görünüş geçerliđi olmak üzere dört türde incelemektedir. Burada Tekin'in yaklaşımı esas alınmıştır.



Kapsam Geerliđi

- Kapsam geerliđi, bir bütn olarak testin ve testteki her bir maddenin amaca ne derece hizmet ettiđidir.
- Bir testin kapsam geerliđi; o testteki toplam maddelerin llecek davranıřı ve konu ieriđini rnekleme derecesine ve testteki her bir maddenin lmek istediđi davranıřı ne derece iyi ltđne bađlıdır.
- Bir testin kapsam geerliđine sahip olabilmesi iin, lme konusu evreni yeteri derecede rneklemesi ve kapsaması ayrıca testteki her bir madde lmek istediđi davranıřı gerekten lmesi gerekir.

Kapsam Geerliđi

- Bir testin kapsam geerliđini belirlemek iin testteki btn maddelerin lmek istenen konuyu kapsayıp kapsamadıđına bakılır. Bunun iin konuların nem derecesini belirleyen belirtke tablosundan yararlanılır.
- rneđin: Trke dersinde 12 konuda 44 hedef davranıřı đrencilerin ne derece kazandıđını belirlemek iin yapılan sınavda eđer,
- 8 konuda ya da 30 edef davranıřı kapsayan soru hazırlanırsa kapsam geerliđi DřK olacaktır.



Kapsam Geerliđi

- Kapsam geerliđinin yksek olması iin izlenmesi gereken sıra:
- clecek zelliklerin veya hedef davranışların belirlenmesi,
- lme aracını oluřturacak zelliklerin veya davranışların belirlenmesi,
- lme aracını oluřturan zelliklerin yada davranışların, evreni (llmek istenen zellikleri veya davranışları) temsil etme dzeyinin belirlenmesi.
- lme aracındaki zelliklerin veya soruların, lmeye uygunluđunun kontrol edilmesi.



Yordama Geerliđi

- Bir testin yordama geerliđi, o testten elde edilen puanlarla testin yordamak iin dzenlendiđi deđiřkenin dođrudan ls olan ve daha sonra elde edilen lt arasındaki korelasyondur.
- Bu biimde hesaplanan korelasyon katsayısı yordama geerliđi katsayısı olarak adlandırılır.



Yordama Geerliđi

- Korelasyon katsayısının byklđ yksek yordama geerliđine iřarettir.
- rneđin; niversiteler arası seme sınavında alınmiř olan puanlarla seme sınavı sonularına gre niversitelere girmiř olan đrencilerin niversitelerde almıř oldukları notlar arasındaki korelasyon katsayısı, niversiteler arası seme sınavının yordama geerliđi katsayısıdır.



Yapı Geçerliđi

- Yapı geçerliđi genelde zeka, ilgi, kiřilik vb. niteliklerin ölçölmesiyle ilgilidir. Bu tür nitelikler ölçölürken önce ilgili niteliđin yapısı tanımlanır.
- Sonra bu niteliđin göstergesi olabilecek davranışları ölçecek araçlar geliştirilir.
- Bu tür araçların geçerliđi; davranışların saptanmasından, araçlardaki maddelere verilen cevapların yorumlanmasına kadar pek çok işlemin geçerli olmasına bađlıdır.



Görünüş Geçerliđi

- Görünüş geçerliđi, bir testin gerçekten ne ölçtüđü ile deđil onun ne ölçüyor göründüđü ile ilgilidir.
- Bir testin görünüş geçerliđi o testin ölçmek istediđi şeyi ölçüyor görünmesidir. Görünüş geçerliđi olan bir test kapađında ne yazıyorsa onu ölçüyor görünür. Örneđin kapađında "Fizik Testi" yazılı olan bir testin içinde fizikle ilgili sorular varsa söz konusu testin görünüş geçerliđi vardır denilir.
- Yani üzerinde "matematik testi" yazılı olan bir testin içinde kimya soruları bulunmamalıdır. Test görünüş olarak ölçmek istenilen konuda olmalıdır.

GÜVENİRLİK (RELIABILITY)

- Karasar'a (1984. s,155) göre güvenilirlik; "aynı şeyin bağımsız ölçümleri arasındaki kararlılık, ölçmek istenen şeyin sürekli olarak aynı sembolleri alması, aynı süreçlerin izlenmesi ve aynı ölçütlerin kullanılması ile aynı sonuçların alınması" şeklinde tanımlanmaktadır.



Güvenirlik

- Tekin(1993. s,41) ve Özgüven(1994. s,83) ise güvenilirliđi "bir ölçme aracının ölçtüđü şeyi tutarlı, kararlı bir şekilde ölçmesi" biçiminde tanımlamaktadırlar.
- Tanımlamalardan anlaşıldığına göre bir ölçme aracının güvenilir olması için o ölçme aracını aynı koşullar altında tekrar tekrar uyguladığımızda aynı ya da en azından benzer sonuçları vermesi gerekmektedir.



Ölçme hataları

- Gerçek değer ile ölçme işlemi sonucunda elde edilen değer arasındaki fark ölçmede hata olarak tanımlanır.
- Ölçme hatalarının kaynakları
 - Ölçülen özelliğın niteliğinden kaynaklanan hatalar
 - Ölçme aracından kaynaklanan hatalar
 - Ölçme işlemi yapan kişiden kaynaklanan hatalar
 - Ortamdan kaynaklanan hatalar



Ölçmede Hata Türleri

- Sabit hatalar
- Sistematik hatalar
- Raslantısal/tesadüfi hatalar



Ölçme hataları- Sabit hata

- Bir ölçmeden diğerine miktarı değişmeyen, her ölçmede aynı miktarda hatanın meydana geldiği hata türüdür.
- Örneğin toplam puanı 5 olan bir yazılı yoklamada 2 puanlık bir soru, verilen zamanın yeterli olmaması nedeniyle, tüm sınıf tarafından bos bırakılmış olan bir sınavda öğretmenin, tüm öğrencilere bu soruyu doğru cevaplamış gibi 2 puan fazla not vermesi.
- Terazinin ölçtüğü her şeyi 50 gram eksik tartması

Ölçme hataları-Sistemik hata

- Hata miktarının ölçümden ölçüme belli bir oranda az ya da fazla olduğu, daha çok ölçme işlemini yapanın subjektifliğinden kaynaklanan hata türüdür.
- Örneğin öğretmenin yazısı güzel öğrencilere daha yüksek puan vermesi,
- kız öğrencilere erkek öğrencilerden daha yüksek puan vermesi
- Sınava 50 puan alan öğrenciye 5 puan, 60 puan alan öğrenciye 10 puan eklemesi gibi hatalar bu türden bir hatadır.

Ölçme hataları- Tesadüfî hata

- Hata miktarının ölçmeden ölçmeye kuralsız ve tamamen rastlantısal bir şekilde değiştiği kaynağının büyüklüğünün ve yönünün bilinmediği hata türüdür.
- Örneğin şans başarısı (atıp tutturma),
- sınav koşulları,
- öğretmenin cevapları puanlarken dikkatsizlik yapması gibi etkenler bu türde hatalara yol açar.
- Bir ölçme sonucu, içindeki tesadüfî hataların azlığı oranında güvenilir sayılır.

Güvenirlilik Katsayısının Hesaplanması

- Güvenirlilik katsayısının hesaplanmasını Anastasi(1988. s,116-125) ve Tekin(1993. s,57-67) dört şekilde açıklamışlardır.
- Bunlar; test tekrar test yöntemi,
- birbirini takip eden ya da paralel testler yöntemi,
- bir testin iki yarıya bölünmesi yöntemi ve
- kuder-richardson 20 ve 21 formülleridir. Burada bu yöntemlerin ne oldukları kısaca açıklanmıştır



Test Tekrar Test Yöntemi

- Bu yöntemle test güvenliğini belirlemek için bir test aynı gruba belli bir zaman aralığı ile iki kez uygulanır.
- İlk uygulamadan elde edilen puanlar ile ikinci uygulamadan elde edilen puanlar arasındaki korelasyona bakılır.
- Korelasyon katsayısı yüksek ise testin güvenilirliği yüksek demektir.
- Bu yolla güvenilirlik katsayısı hesaplamada iki test uygulaması arasında geçen süre testin güvenilirliğini olumsuz olarak etkilemektedir.



Paralel Testler Yöntemi

- Bu yöntemle güvenlik tahmini için bir testin en az iki eşdeğer formunun geliştirilmesi gerekir.
- Paralel iki testin eşdeğer olabilmesi için her iki test içindeki maddelerin sayısı, niteliği ve ölçtükleri davranışlar bakımından birbirine denk olmalıdır.
- Bu yöntemle güvenilirlik tahmininde, testin her iki formu aynı anda aynı gruba uygulanır ve her iki paralel formdan alınan puanlar arasındaki korelasyon hesaplanır.



Bir Testi İki Yarıya Bölme Yöntemi

- Bu yöntemde test bir gruba uygulandıktan sonra iki yarıya bölünür ve testin iki yarısından aldıkları puanlar arasındaki korelasyon hesaplanır.
- Hesaplanan korelasyon katsayısı testin yarısının güvenilirliğini gösterir.
- Daha sonra istatistiksel bir formül kullanılarak (Sperman-Brown) testin tümünün güvenilirliği hesaplanır.



Kuder-Richardson 20 ve 21 Formülleri

- Bu yöntem ancak bir testin içindeki herbir maddenin güçlük derecesi biliniyorsa uygulanabilir.
- Kuder-Richardson formülleri, testteki herbir maddenin aynı değişkeni ölçtüğü, yani testin ölçtüğü şeyin homojen olduğu varsayımına dayanır.



Güvenirliđi etkileyen etkenler

- 1. Testin kendisiyle ilgili etkenler
 - a) Testte yer alan soru sayısı: Soru sayısının belli bir noktaya kadar artırılması güvenilirliđi artırır.
 - b) Test yönergesinin ve testte yer alan soruların ifadesi: Bir testteki soruların ifadesinin yalın ve anlaşılır olması testten elde edilen puanların güvenilirliđini olumlu yönde etkiler.

Güvenirliđi etkileyen etkenler

- c) Testin homojenliđi: Ölçtükleri davranış ve konu bakımından homojen (benzer) bir testten elde edilen puanlar, heterojen (farklı) bir testten elde edilen puanlardan daha güvenilirlerdir.
- d) Puanlamanın nesnelliđi: Bir testin deđişik kişilerce veya aynı kişi tarafından deđişik zamanlarda puanlanmasından elde edilen puanların tutarlılıđı testin güvenilirliđini olumlu yönde etkiler.

Güvenirliđi etkileyen etkenler

- 2. Testin uygulama kořullarıyla ilgili etkenler:
- Uygulama kořullarının her öđrenci için aynı olması testin güvenilirliğini olumlu yönde etkiler.



- **3. Testi alan öğrenci veya testin uygulandığı grupla ilgili etkenler:**
- **Bu açıdan bakıldığında bir testin güvenilirliği, testin uygulandığı kişinin uygulama sırasındaki fiziksel ve ruhsal durumuna ve o testin uygulandığı grubun ölçülen özellikler bakımından homojen veya heterojen olmasına bağlıdır.**



- Bir ölçme aracında güvenilirlik ve geçerliğin yanı sıra bulunması gereken bir başka nitelik ise kullanılabilirliktir.
- Kullanılabilirlik, bir ölçme aracının veya yönteminin, geliştirilme, çoğaltılma, uygulanma ve puanlama yönünden kolaylık, nesnellik ve ekonomiklik özelliklerine sahip olması anlamına gelir



- Testin uzunluđu
- Uygulanış kolaylıđı
- Fiyatı
- Başka formlarının olması
- Ulaşma kolaylıđı

