

BÖLÜM-11

EĐİTİMDE ÖLÇME VE DEĐERLENDİRME

Eđitimi; girdisi, bir işlem süreci ve bu işlem sürecinin sonunda da çıktısı, yani ürünleri olan bir sistem olarak ele aldığımızda şunları görürüz:

Eđitimin girdisi; “öđrenciler”, “öđretimi gerçekleştiren eđitciler”, “öđretim programları” ve “okul, kitap, defter, sıra, araç, gereç” gibi eđitimde kullanılan kaynaklardır.

Süreçte ise; “öđrenci katılımı ve performansı”, “eđitcinin performansı”, “programların uygulanması” ve “mevcut kaynakların rasyonel şekilde kullanılması” yer alır.

Çıktısı da “öđrenci kazanımları”, yani “program aracılığı ile kazandırılması veya deđiştirilmesi öngörülen eđitsel davranışlardır.”

Eđitim sisteminde ölçme deđerlendirmenin yeri

Sistemlerde, ürün elde etmenin bir maliyeti vardır. En azından bu maliyetin karşılığının alınıp alınamadığını öđrenebilmek için, süreç sonunda elde edilen ürünlerin, istenilen kalitede olup olmadığının bilinmesi istenir. Bu, hem ürünün kalitesinin sürekli iyileştirilmesi, hem de

varsa üretim sürecindeki aksaklıkların belirlenerek sistemin daha sağlıklı şekilde işletilebilmesi için gereklidir. Sistemlerde bu işlevi yerine getiren ögeye **geri bildirim** ya da **dönüt** ögesi denir.

Ölçme değerlendirme etkinlikleri, eğitim sisteminde, sistemin **geri bildirim ögesini** oluşturur. Ancak, uygulamadaki ele alınışı itibarıyla, ölçme değerlendirme, kendisinden beklenen kapsamlı geribildirim işlevini çoğu kez yerine getiremez. Çünkü sınıflardaki ölçme değerlendirme, çoklukla, öğrencilerin bir sonraki sınıfa ya da eğitim kademesine devam edip edemeyeceklerine karar vermek için yapılır.

Oysa ölçme değerlendirme, öncelikli olarak, öğrencileri geçti-kaldı diye sınıflandırmak için değil; mevcut gir-dilerle başlatılan süreç sonunda, öngörülen çıktının elde edilip edilemediğini, edilememiş ise neden elde edilemediğini ve elde edilebilmesi için bir dahaki sefere, öğretim süreçlerinde nelerin yapılması gerektiğini ortaya koymak için, yani sistemi sağlıklı bir şekilde işletmek ve geliştirmek için yapılmalıdır. Öğrencilerin başardı ya da başaramadı diye sınıflandırılması, ölçme sonuçlarıyla yapılan basit bir kıyaslama işleminin zaten olması gereken doğal bir sonucudur.

Ölçme değerlendirme etkinliklerinin amacına yönelik olarak yapılması kadar, ölçme sonuçlarının **var olan gerçek durumu** yansıtır olması da önemlidir.

Ölçme değerlendirme sonuçlarının, hatadan arındırılmış ve amacına uygun bir şekilde eğitim sisteminin gerçek durumunu yansıtır olması **iki açıdan** önemlidir.

Birincisi, ölçme sonuçları sistemin veya sınıfın gerçek durumunu ortaya koymadığı zaman, gerçek olmayan durumlara dayalı olarak yapılan işlerin sonunda, en hafifi kaynak ve zaman israfı olmak üzere, **bedeli farkı** şekil-

lerde ödenen olumsuzluklar⁶⁵ yaşanır. Öğretmenin yaptığı sınavlarda çok başarılı görünen öğrenciler, üniversite sınavı gibi standart sınavlarda başarısız olurlar. Ya da sistemin standart sınavlarında başarılıymış gibi görünen öğrenciler, uluslararası değerlendirmelerde düşük kırıklıkları yaşar ve yaşatırlar. En acısı da okul başarısı hayat başarısını getirmez. Daha da kötüsü, kendini eğitimi, saygın, üretken hatta zengin ve gelişmiş sanan toplum, çevresindeki ülkelerin pazarı ve ucuz işgücü olmaktan öteye gidemez.

İkincisi, insanın doğasında adaleti tecelli ettirme eğilimi vardır. İnsanlar hak ettiklerini alamadıkları zaman önce kırılırlar, daha sonra bu kırıklık güvensizliğe ve en sonunda da öfkeye dönüşür. Ardından bu olumsuz duygular başka alanlara da genellenmeye başlar. Öğretim kurumlarını; okuluna, öğretmenine, müdürüne, dekanına güvenmeyen öğrenciler doldurur. Daha sonraki aşamada, beklenen adalet gelmeyince, adalet bu kez meşru olmayan yollarla farklı şekillerde sağlanmaya çalışılır.

65 Kaynağına ulaşmak için yaptığım araştırmalarda, aşağıdaki kısa yol aracılığıyla İngilizcesine ulaştığım bir sosyal medya paylaşımının konuyla çok ilgisi var. **Doğru yapılan ölçme değerlendirmeyi, eğitim sisteminin sağlıklı işleminin anahtarı olarak gördüğümüzde**, eğitim sisteminin niteliksiz olmasının yol açacağı olumsuzlukları örneklemesi açısından söz konusu paylaşımı aşağıya aktarıyorum:

Bir ülkeyi yok etmek için atom bombası veya uzun menzilli füzelere ihtiyaç yoktur. Bunun için eğitim seviyesini düşürmek ve kopya çekilmesine müsaade etmek yeterlidir. Böylesi bir sistemin ürünü olan,

*Doktorların elinde hastaları can verir,
Mühendislerin mimarların yaptığı binalar, köprüler çöker,
Ekonomistler yüzünden para ve kaynaklar israf edilir,
Hâkimlerin elinde adalet katledilir,*

Doğmatik düşünen, bağınaz bilim insanlarının elinde insani değerler ve insanlık yok olur.

Bir ulusun eğitim sisteminin çökmesi, o ulusun çöküşüdür...

[<https://medium.com/@presidentdzitse/how-a-nation-is-collapsed-4e97625c4e07>]

Eđitim đretim ortamlarında adalet duygusunun **en somut** olarak yařatılacađı yer, đrenci bařarisının llmesidir. Bu nedenle lme deđerlendirme **nemli bir eđitici yeterliđidir**⁶⁶.

Neden hem “lme” hem de “deđerlendirme”?

“lme” ve “deđerlendirme”, birbirinden farklı iki iřlemi ifade eden kavramlardır.

lme; bir varlıđın, rneđin bir kiřinin, “cinsiyeti”, “boyu”, “kilosu”, “sa rengi”, “mesleđi”, “matematik puanı”, “yz metreyi ne kadar srede kořabildiđi” gibi eřitli

66 Meslek hayatım boyunca yaptığım her sınavda, đrenciler daha sınavdan çıkmadan, sorduđum soruların beklediğim cevaplarını ve bunların ne řekilde puanlanacağını ieren cevaplanmış bir sınav kâđidını alıřma odamın kapısına astım. Bu kâđıdın zerine de **CEVAPLAR** veya **CEVAP ANAHTARI** gibi kliře ifadeler yerine **NERİLEN CEVAPLAR** yazdım. (Nedenini soranlara da, “Olur da bir đrencim benim dřunemediđim, daha gzel bir cevabı formle edip yazarsa o da kabulmdr mesajı vermek iin” dedim.) Sınav kâđıtlarını, ok zel bir durum olmadıka, izleyen derse girerken yanımda getirdim ve đrencilere dađıtarak puanlamayı kontrol etmelerini sađladım. Bu uygulamada, sınav kâđıdını alan đrencilerin, ncelikle hangi sorudan ka puan aldıklarına deđil, aynı sorudan yanındaki arkadařının ka puan aldıđına, kendisinin ka puan aldıđına baktığını grdm ve bu yzden sınav kâđıtlarını hibir zaman zaman bir btn olarak puanlamadım. Sınavları soru soru puanladıđımda, hem kimin kâđıdını puanladıđımı bilemedim hem de her sorunun puanlanma ltleri herkes iin aynı olabildi. Son ařamada puanları topladıđımda, geme puanının bir iki puan altında kalan kâđıtlar iin de gereksiz itirazlar gelmedi, fazladan puan talebinde bulunlar olmadı. İzleyen yıllarda, sınıf mevcutlarının yzlerle ifade edildiđi gruplara ders vermek durumunda kaldıđımda oktan semeli sınavlar yapmak zorunda kaldım. Bu durumda da, cevap formunun en altına, zerine hangi soruda hangi seeneđi iřaretlediklerini yazdıktan sonra yırtarak sınav salonundan ıkarabilecekleri bir blm ekledim. Sınavdan hemen sonra da dođru cevap olan seenekleri paylařtım. Sınavı izleyen derste de tm soruları, dođru olmayan seeneklerin neden dođru olmadıklarını da gstererek tekrar yanıtladım. ok iddialı olduđum dersler verdim, đrencilerime kitaplarda bulamayacakları ayrıntıları đrettim ama rahle-i tediten geen đretmenler uzun yıllar sonra beni đrettiklerimle deđil, sınavlarda bu yaptıklarımınla hatırladı. Aktif meslek yařantımın sonlarına dođru niversiteler deđiřti, sahip olduđum deđerleri benimsemeyen, bunu derse ve bana genelleyen đrenciler oldu; anonim olarak arkamdan yazıldı izildi ama “adil bir hocaydı” lafı esirgenmedi.

özelliklerinden, ilgilendiđimiz herhangi bir özelliđini ele alarak, varlıđın bu özelliđe sahip olup olmadıđını, sahipse sahip olma derecesini, **sayı** ya da **sembolle** ifade etmektedir. Eđer özellik “boy”, “kilo”, “matematik puanı” gibi birimi olan bir özellikse 175 cm., 65 kg., 75 puan gibi **sayılarla**; “cinsiyet”, “meslek” gibi bir özellikse, bu kez sayı yerine “erkek”, “kadın”, “öđretmen”, “marangoz” gibi sayı dıřında **sembollerle** ifade edilir.

Deđerlendirme ise, ölçme sonucunun **bir ölçüt** (veya kriter) ile kıyaslanarak, bu ölçme sonucu hakkında “yeterli-yetersiz”, “geçer-kalır” gibi bir yargıda bulunmaktır. Özel güvenlik görevlisi yetiřtiren ve verdiđi iş ilanında “Adayların boylarının en az 175 cm olması gerekir.” kořulu bulunan bir eğitim kurumuna bařvuran adayların boy uzunluklarının belirlenmesi ölçmedir. Ancak adayın ölçme ile belirlenen boy uzunluđunun, ölçüt olan 175 cm ile kıyaslanarak, aday hakkında “kabul edilebilir” ya da “kabul edilmez” řeklinde bir yargıda bulunulması deđerlendirmedir. Ya da bir **sınavın puanlanması ölçme**, puanın “yeterli sayılmanın” ölçütü olan puan deđerı ile kıyaslanarak, öđrencinin **geçmesine** ya da **kalmasına karar verilmesi de deđerlendirmedir.**

Ölçmede birim sorunu

Varlıkların, ölçme için ilgilenilen özelliklerine yakından baktıđımızda řunu görürüz. “Cinsiyet”, “medeni hal”, “meslek”, “nüfusa kayıtlı olduđu il” gibi özellikler ile “boy”, “kilo”, “yılda okunan kitap sayısı”, “sınav puanı” gibi özellikler, özelliđin dođası ve işlevi açısından farklıdır.

Sembollerle ifade edilen “cinsiyet”, “medeni hâl” gibi özellikler, varlıđı, (“erkek-kadın”, “evli-bekar-bořanmıř” gibi) **bir gruba** veya **bir kategoriye ait kılar**, yani sınıflandırır.

Sayılarla ifade edilen (170 cm, 60 kg 85 puan řeklindeki) “boy”, “kilo”, “sınav puanı” gibi özellikler; **varlıđın bu özelliđe sahip olma “miktarını” ya da “derecesini”**

belirtir. Ancak, miktar ile ifade edilen bu tür özelliklerden söz ederken, varlığın o özelliğe sahip oluş miktarını, **herkes için anlaşılır kılmak amacıyla**; önceden tanımlanmış, herkesçe ne kadar olduğu bilinen bir özellik miktarına ihtiyaç duyulur. Bu **tanımlanmış özellik miktarına** birim⁶⁷ denir. Örneğin; belli bir uzunluk miktarına “metre”, onun yüzde birine “santimetre” denmiştir. Belli bir süreye “gün”, onun yirmi dördte birine “saat” denmiştir. Belli bir kütle miktarına “kilogram”, onun binde birine de “gram” denmiştir ve biz ölçme ile varlığın bu özelliğe kaç birim miktarı sahip olduğunu belirlemeye çalışırız.

Bu bakış açısıyla ölçme aslında bir tür “doğrudan” gözlemdir. Bir kalemin boyunu ölçerken, kalemin boyu içinde, yanına koyduğumuz cetveldeki birim özelliklerden (yani, santimetrelerden) kaç tane olduğunu gözleriz. Ya da, karpuzu eşit kodlu terazide tartarken, onu dengelemesi için diğer kefeye ne kadar gram konduğunu gözleriz.

Ölçme çeşitleri ve davranış ölçme

Yapılış biçimi itibarıyla farklı ölçme şekilleri vardır. Bir öğrencinin süreç sonunda ne kadar öğrendiğini belirlemek için yapılan ölçme ile bir cetvelle kalemin boyunu belirlemek için yapılan ölçme birbirinden oldukça farklıdır.

Doğrudan ölçme

Varlığın sahip olduğu özellik miktarının kaç birim olduğunun doğrudan gözlenmesi suretiyle yapılan ölçmelere **doğrudan ölçme** denir. Örneğin, şerit metreyle bir masa-

⁶⁷ Varlığın doğal bütünlüğünün aynı zamanda birimi olduğu durumlarda kullanılan (sınıftaki öğrenci sayısı gibi, ağaçtaki elma sayısı gibi) birimlere **doğal birimler** denir. Ancak varlığın doğal bütünlüğünün birimi olmadığı, ilgilenilen özelliğin kararlaştırılmış bir miktarının birim olduğu durumlardaki birimlere de **tanımlanmış birimler** denir. Tanımlanmış birimlerin **eşitlik** (farklı yer ve zamanlarda veya ardışık olarak kullanılan miktarların hep eşit olması), **yaygınlık** (herkesçe bilinmesi ve kabul görmesi) ve **amaca uygunluk** (kömür miktarını ölçerken ton, kuyumcudan altın alınırken gram kullanılması gibi) gibi özelliklerinin olması istenir.

nın enini veya boyunu ölçme doğrudan ölçmedir. Çünkü masanın enini veya boyunu ölçerken, masanın kenarına koyduğumuz şerit metre veya cetvel aracılığı ile, masanın en ya da boy özelliđi içinde kaç tane birim özellik (santimetre) olduğunu doğrudan gözleriz.

Dolaylı ölçme

Ancak, bir de varlığını bildiđimiz ama aldıđı deđeri ve deđerindeki deđişimi doğrudan gözleyemediđimiz özellikler vardır. Örneđin, havanın sıcaklığı veya bir cisme uygulanan kuvvet gibi. Miktarları belirlenecek “sıcaklık”, “kuvvet” gibi özelliklerin doğrudan gözlenerek ölçmesi olanaksızdır.

Böyle durumlarda ölçme, kendisini doğrudan gözleyemediđimiz özelliđin **üzerinde gözlenebilir deđişime yol açtığı** başka bir varlık aracılığı ile yapılır. Örneđin doğrudan gözleyemediđimiz sıcaklık, cisimlerin hacminde gözlenebilir bir genleşmeye yol açar. Eđer sıcaklığın kolayca genleştireceđi alkol veya cıva gibi bir sıvı ince bir cam borunun içine konursa, sıcaklıktaki deđişimin sonucu olarak, cam borudaki sıvının hacmindeki genleşme doğrudan gözlenebilir. Tıpkı kuvvet uygulanan cisimlere bir lastik veya yay bađlandığında, kuvvetteki deđişimin lastikte veya yayda neden olduđu uzamanın gözlendiđi gibi. Doğrudan gözlenemeyen özelliklerin, üzerinde yol açtıkları etkilerinin doğrudan gözlenebildiđi başka varlıklar aracılığı ile ölçülmesine **dolaylı ölçme** adı verilir⁶⁸.

Sınavlar aracılığı ile ölçme

Davranışlar da **dolaylı olarak** ölçülür. Daha açık bir ifadeyle, eđitimde davranış ölçme, bir tür dolaylı ölçmedir. Bireye eđitimle kazandırılmış, varlığını bildiđimiz

68 Bir de, birden fazla ölçümün sonuçlarının kullanılarak yapıldığı bir ölçme türü vardır. Bir cismin özkütlesini (yoğunluđunu) ölçmek için, hem hacmini hem kütlesini ölçmeniz; ardından da kütlesinin hacmine oranını bulmanız gerekir. Buna da **türetilmiş ölçme** denir.

ama gözleyemediğimiz potansiyel davranışlar, ölçmeyi yapanın gözleyebileceği bir şekilde dönüştürülerek ölçülmeye çalışılır.

Dolaylı ölçme olan sıcaklık ölçümünde, doğrudan gözlenemeyen sıcaklık değişimini gözlenebilir kılmak için kullanılan araç, cam boru içinde genişmesi gözlenebilen alkol ya da sıvıdır. Davranış ölçümünde de, bireyin doğrudan gözleyemediğimiz bilgisini, potansiyel davranışlarını gözlenebilir kılan araç ise en genel haliyle sınavlardır.

Ama bir farkla...

Öğrencinin sahip olduğu potansiyel davranış ile sınav sonucu arasındaki ilişki, sıcaklıktaki değişim ile sıvının genişmesi arasındaki, “şu kadar sıcaklık artışı hacimde şu kadar artışa yol açar” şeklinde fonksiyonunu yazabileceğimiz türden, düzgün bir ilişki değildir.

Eğiticiler; sınav hazırlarken, önce öğrenenin potansiyel davranışını açığa çıkarmak için, öğrenenin istenen davranışı içinde sergileyebileceği hayali bir durum tasarlar. Ardından eğitici, kendi zihninde tasarladığı bu hayali durumu, öğrenenin kendisinden istenilenleri anlayabileceği bir şekilde ifade etmeye çalışır. Bu sınav sorusudur ve bu hâliyle eğiticinin zihnindeki anlamı taşır.

Öğrenen, sınavda (yani, ölçme esnasında) bu soruyu okur ve soruya kendince bir anlam yükler. Öğrenenin okuduğu bu soruya yüklediği bireysel anlam, çoğu kez eğiticinin soruyu ifade ederken ona yüklediği anlamla bire bir örtüşmez çünkü öğrenen soruyu, kendi bakış açısıyla, kendi fiziksel ve davranışsal çevresiyle anlamlandırır.

Öğrenen, sorulan soruya kendince yüklediği bireysel anlama göre, sergilemesi istenen davranışın karşılığı olan sorunun cevabını zihninde oluşturur. Aslında bu cevap, eğiticinin ortaya çıkarmak ve gözlemek istediği davranıştır ama bu davranış eğiticinin soru ile ortaya çıkarmaya çalıştığı davranış ile ne kadar örtüşür, bu da pek bilinemez.

Ardından öğrenen zihninde oluşturduğu sorunun cevabını, dil ile kodlamak suretiyle dışa vurarak, gözlenebilir hale getirir. Ancak bu gözlenebilir cevabın, öğrenenin zihninde oluşturduğu davranışı, **ne kadar öğrenenin zihninde canlandırdığı haliyle yansıtabildiği** de öğrenenin dili kullanma becerisi ile sınırlıdır.

Son aşamada da eğitici, okuduğu veya işittiği bu cevaba kendince bir anlam yükler. Bu noktada da, öğrenenin ve eğiticinin bu cevaba yüklediği anlamların **ne derece örtüş-tüğü** ayrı bir tartışma konusudur⁶⁹.

Her ikisi de dolaylı ölçme olmasına rağmen; **sınav sorusuyla davranış ölçme** ile **termometreyle sıcaklık ölçme** arasındaki farka tekrar dönecek olursak, sıcaklıktaki artış ile cam borudaki sıvının hacmindeki genişleme arasındaki ilişki fonksiyonunu yazabileceğimiz, belirli ve düzgün bir ilişkidir. Ama, yukarıda özetlendiği hâliyle konu; potansiyel bir davranışı, sınav sorusu ile açığa çıkarma olunca, cevaplanması istenen bir sınav sorusunun, **ölçülmek istenen davranışı** ne derecede **doğru bir şekilde** ve hangi miktarda açığa çıkarabildiği ve **gözlenir hâle getirdiği** tartışmalıdır. Bu nedenle, davranışların dolaylı olarak ölçülmesi, **sanıldığından çok daha fazla hata içerir**.

Ölçmede hata ve hataların kaynakları

Dolaylı bir ölçme olan sıcaklık ölçümü, farklı termometrelerle farklı sınıflarda ve farklı zamanlarda yapıldığında bile (bu ölçmelere de hata karışma ihtimali nedeniyle) **ol-**

69 Öğrencilerin sınav sorularına verdikleri şu cevaplar, internette anonim olarak dolaşan paylaşımlardan derlenmiştir. Bir matematik sınavında sorulan ve sonu "*Hangisi daima tektir?*" ifadesiyle biten soruya verilen "*Allah*" cevabı, "*Divan üyelerinden kadının görevleri nelerdir?*" sorusuna verilen "*Yemek yapmak, bulaşık yıkamak, çocuklara bakmak*" cevabı, "*Devletin temel organlarına örnek veriniz.*" sorusuna verilen "*Mide, akciğer, kalp*" cevabı, "*Aşağıdaki şiirin ölçüsü nedir?*" sorusuna verilen "*Yaklaşık 9 cm.*" cevabı, "*Çin uygarlığına ait teknik buluşları yazınız.*" sorusuna verilen "*Çin malı olanlar*" cevabı ve "*Türkiye'nin başlıca geçitlerini yazınız.*" şeklindeki coğrafya sorusuna verilen, "*Alt geçit, üst geçit ve yaya geçidi*" cevabı.

dukça doğru sonuçlar elde edilir ve o ortamlardaki hava sıcaklıkları az bir hata ile belirlenmiş olur.

Ancak aynı eğiticinin, aynı sorularla aynı sınıfta, aynı sırada yan yana oturan, benzer özelliklere sahip iki öğrenci için yaptığı davranış ölçümü; bu iki öğrenciyi, ölçülen davranış açısından **birbirinden sağlıklı bir şekilde ayırt edemeyeceğimiz** kadar çok hata içerebilir.

Gerçekten de, sorulardan ne anladıkları, buna göre zihinlerinde neler oluşturdukları, zihinlerinde oluşturduklarını ne şekilde anlatabildikleri ve eğiticinin bunlara ne anlamlar yüklediği düşünüldüğünde **bunun bir abartma olmadığı** görülecektir.

Şimdi de, pek çok faktörün hataya neden olduğu dolaylı bir ölçme türü olan davranış ölçümünde, hataların kaynaklarına biraz daha sistematik olarak bakalım:

1-Ölçmeyi yapan kişinin yanlı olması, özensiz ve dikkatsiz davranması, davranış ölçme konusunda yeterli derecede bilgisinin olmaması ölçmede hataya neden olabilir. Bu tür hatalara **ölçmeyi yapan kişiden kaynaklanan hatalar** denir.

2-Ölçme aracının, yani sınav kağıdının, açık ve anlaşılır olmaması, ölçülmek istenen davranışı başka davranışlarla karıştırıyor olması, harf karakterlerinin boyu, satır aralıkları, karışık sayfa düzeni, sayfa sonunun böldüğü sorular gibi nedenlerle soruların okunmasında güçlükler olması da hataya neden olabilir. Bu hatalara da **ölçme aracından kaynaklanan hatalar** denir.

3-Sınav yerinin ısı, ışık ve havalandırma koşullarının kötü olması, sıra ya da masaların üzerinde yazı yazarken oynaması veya yüzeylerindeki pürüzler nedeniyle rahatça yazmaya uygun olmaması, gözetmen ya da sınav sorumlusunun giydiği sert topuklu ayakkabıların yürürken ses çıkarması, birden fazla gözetmenin olduğu durumlarda gözetmenlerin kendi aralarında konuşmaları da sınava

girenlerin performanslarını olumsuz yönde etkileyerek hataya neden olabilir. Bu tür hatalar da **ölçmenin yapıldığı ortamdan kaynaklanan hatalar** olarak adlandırılır.

4-Sınava giren kişinin derse veya sınava karşı olumsuz tutumları ya da ön yargıları, sınav anında yorgun, hasta veya canının sıkın olması, dikkatini toplayamaması, özellikle **performansa dayalı sınavlarda** kişiliğinin ölçmede kullanılan yöntemle uygun olmaması da ölçmede hataya neden olabilir. Bunlara da **davranışları ölçülen kişiden** (veya ölçülen özellikten) **kaynaklanan hatalar** denir.

5-Eđitimde ölçülecek davranışlar farklı farklı olduğu gibi bunları ölçmek için kullanılacak ölçme araç ve teknikleri de farklı farklıdır. Bir konunun olabildiğince değişik bakış açılarından yorumlanmasını gerektiren eğitsel bir davranış için çoktan seçmeli test uygun olmayabilir. Ölçmede kullanılan uygun olmayan ölçme aracı, uygun olmayan ölçme şekli de hataya neden olabilir. Bu tür hatalar da, **ölçmede tercih edilen yöntemden kaynaklanan hatalar** olarak adlandırılır.

Eđitimde ölçme değerlendirmeye, bir soruya cevap arayarak, bir de başka bir açıdan bakalım:

Aynı programı izleyen farklı cođrafi bölgelerde, aynı tipteki iki lisenin ikinci sınıfına devam eden iki öğrenci, aynı konulardan girdikleri üçüncü matematik yazılı sınavından aynı puanı, örneğın 10 tam puan üzerinden 8'er puanı aldıklarında, bu iki öğrenci için, **matematik başarıları aynıdır** diyebilir miyiz?

Bir adım daha ileri giderek bir soru daha soralım.

Okulda öğretmenlerin yaptığı matematik yazılılarından hep tam puan alan, takdirlik başarılı öğrenciler, üniversite giriş sınavında neden aynı başarıyı gösteremiyorlar?

Ölçmede başlangıç noktası olan sıfırın anlamı

Ölçmede varlığın sahip olduğu özelliğin miktarını ifade ederken başlangıç noktası olarak kullanılan sıfırın anlamı her zaman aynı değildir. Ölçmede başlangıç noktası olan sıfır, üstlendiği anlama göre, **mutlak** ve **göreceli** sıfır olmak üzere iki türdür.

Ölçmede mutlak sıfır

Sıfır; ölçmeye çalıştığımız özellik miktarının **yokluğu**, yani özelliğin hiç olmadığını gösteriyorsa buna mutlak sıfır denir. Örneğin; öğrencilerin kardeş sayılarını, aldığı seçimlik dersleri, sınıf tekrarı yaptıkları yılların sayılarını belirlemek için yapılan ölçmelerdeki sıfır mutlak sıfırdır ve ilgilenilen özelliğin olmadığını gösterir. Bir öğrencinin kardeş sayısı sıfır ise, bu sıfır hiç kardeşinin olmadığını; aldığı seçimlik ders sayısı sıfır ise hiç seçimlik ders almadığını; sınıf tekrarı yaptığı yıl sayısı sıfır ise öğrencinin hiç sınıfta kalmadığını gösterir.

Ölçmede göreceli sıfır

Hava sıcaklığı sıfır derece dediğimizde, buradaki sıfır sıcaklığın hiç olmadığı anlamına gelmez. Buradaki sıfır, sıcaklık için, suyun donma derecesinin referans alındığı bir başlangıç noktasıdır. Gece yarısı saat 00.00'da zamanın olmadığı da söylenemez. O da, zamanı ölçerken kullanılan bir başlangıç noktasıdır. Varlığın ilgilenilen özelliğinin olmadığı anlamına gelmeyen, sadece ölçmenin başlangıcı olan bir referans noktasını gösteren sıfıra da göreceli (izafi) sıfır denir. Bir sınavdan alınan sıfır puan, davranışı ölçülen bireyin hiçbir davranışı olmadığı anlamına gelmez. Sınavdaki sıfır, sadece o sınavda, ilgilenilen davranışın ölçülmesinde bir başlangıç noktasını ifade eder⁷⁰.

70 Bazı eğitimciler bu sifira "keyfi" sıfır da derler.

Bu nedenle, okullardaki sınavlarda çok başarılı olan öğrencilerin birlikte girdikleri üniversite sınavı gibi standart sınavlarda başarısız olmalarını yadırgamamak gerekir. Okullarda yapılan her bir sınavın sıfırı göreceli sıfırdır, dolayısıyla bu sınav puanlarının birbiriyle kıyaslanması çok doğru ve anlamlı değildir.

Deđerlendirme ve türleri

Deđerlendirme, ölçme sonucunun bir ölçütle kıyaslanarak sonuç hakkında bir yargıda bulunulmasıdır. Ancak deđerlendirme, ölçme sonucunun kıyaslanacağı ölçütün belirlenme şekline göre, **mutlak** ve **bađıl** deđerlendirme olarak ikiye ayrılır.

Mutlak deđerlendirme

Bir deđerlendirmede ölçüt, örneđin bir sınavda başarılı sayılmak için yeterli olan puan, ölçme sonuçları ortaya çıkmadan önce belliyse, yani sınava girecekler, örneđin sınavdan 50 ve üzeri puan aldıklarında başarılı sayılacaklarını biliyorlarsa, bu tür deđerlendirme **mutlak deđerlendirmedir**. Mutlak olarak adlandırılır, çünkü bu tür deđerlendirmede, önceden belirlenmiş, deđişmeyen “mutlak” bir ölçüt vardır. Bu öngörölmüş ölçüt bir sınav için 50 puan da olabilir 70 puan da olabilir. Önemli olan önceden belli olmasıdır. Ayrıca mutlak deđerlendirme türüyle deđerlendirilen bir sınavda kimse başarılı olamayabilir veya herkes başarılı olabilir.

Bađıl deđerlendirme

Deđerlendirmede ölçüt, ölçme sonuçları ortaya çıkmadan önce belli deđilse ve **ölçüt, ölçme sonuçlarına bađılı olarak sonradan belirleniyorsa**, bu tür deđerlendirmelere de bađıl

değerlendirme⁷¹ denir. Örneğin, başarılı sayılma koşulunun sınıf ortalamasının üzerinde puan almak olan bir sınavda değerlendirme ölçütü, ölçme sonuçları ortaya çıkmadan bilinemez. Bu değerlendirme türü de bağıl olarak adlandırılır çünkü ölçüt, grubun sınav puanlarına yani grubun genel performansına bağlıdır. Görüldüğü gibi bu değerlendirmede ölçüt olarak kullanılacak önceden öngörölmüş mutlak bir standart yerine, grubun genel performansına bağıl olarak ortaya çıkan göreceli bir standart vardır.

71 Bağıl değerlendirmenin yaygın olarak kullanılan üç tipi vardır.

DAĞILIMDA BELİRLENEN ORANLARA GÖRE DEĞERLENDİRME: Sınıftaki öğrencilerin puanlarının normal dağılım eğrisinin öngöröldüğü biçimde dağıldığı varsayımıyla, sıralı puanların belli oranlarına belli notlar verilerek değerlendirme yapılır. Örneğin, A dan F' ye kadar A,B,C,D, ve F harf notlarının verildiği bir not sisteminde, sınıfın %40'ının ortalamada (C notu) yığılacağı, B ve D notlarına kaymanın %20'ser olacağı, kalan uç noktalarda (A ve F) ise %10'ar öğrenci olacağı düşünülerek değerlendirmeye esas notlar bu oranlarda dağıtılır. (Yüz kişilik bir grupta, yüksekten düşüğe doğru sıralanmış puanların, ilk 10'una (yani %10'una) A notu, izleyen 20'sine (%20'sine) B notu, izleyen 40'ına (%40'ına) C notu, ardından 20'sine (%20'sine) D notu ve son 10 puana (%10'una) da F notu verilmesi gibi.) Ancak bu tür değerlendirmenin, mevcudu az olan sınıflarda, normal not dağılımını göstermede yetersiz kalacağı için öğrencileri tam olarak sıralamada da yetersiz kalabileceği bilinmelidir.

DAĞILIM ARALIKLARINA (DOĞAL PERFORMANS GRUPLARINA) GÖRE DEĞERLENDİRME: Bir sınav sonunda, öğrencilerin aldıkları puanlar yüksekten düşüğe göre sıralandığında, sıralı puanların aralarındaki belirgin kesintiler (puan farkları) varsa, bunların performans gruplarının sınırlarını oluşturduğu düşünülür. (Örneğin, puanlar 100, 98, 97, 97, 80, 78, 77... şeklinde sıralandığında, 97 ile 80 arası kesinti grup sınırı olarak ele alınır) Değerlendirici, grup sayısına ve grup sınırını belirleyen puan aralıklarına göre gruplara not takdir eder. Ancak, bu puan aralıkları çok belirgin olarak ortaya çıkmayabilir veya grup sayısı çok ya da çok az olabilir.

STANDART PUANLARA GÖRE DEĞERLENDİRME: Bu değerlendirme biçiminde, ham puanlar, grup ortalamasından, kaç standart sapma farklılaştığı $[z=(\text{puan}-\text{ortalama})/\text{standart sapma}]$ hesaplanarak, standart z puanlarına dönüştürülürler. Farklı ortalaması ve standart sapması olan puan dizilerini standart z puanlarına dönüştürmek demek; ortalaması 0, standart sapması 1 olan puan dizileri elde etmek demektir ki bu durumda farklı gruplardaki puanlar birbirleriyle kıyaslanabilir hâle gelir. Çünkü farklı gruplardaki ham puanlar, grup genel performansı referans alındığında, hem grup ortalamaları hem de standart sapmaları eşit olduğundan birbiriyle kıyaslanabilir. Ardından bu standart z puanlarına göre veya (ortalamanın altında kaldığı için negatif olan puanlarla işlem yapmamak için) bunların ortalamalarının 50'ye standart sapmalarının 10'a ötelenildiği T puanlarına $(T=[10 \times z]+50)$ göre not verilir.

Bu noktada, bađıl deđerlendirmeye, eđitimin amaçları ve eđitimden beklentiler aısından bakıldıđında akıllara Őöyle bir soru gelebilir:

Eđitim programlarında program sonunda ulařılacak ama belirlemenin temel gerekesi, eđitilenleri öngörülen bir standarda ulařtırmaktır. Bu, eđitim aısından dođru bir önermedir. Ama, **eđitimle, mevcut durumdan hareketle süreç sonunda varılmak istenen bir standarda ulařmak varken, süreç sonunda ortaya çıkan mevcut durumu bir standart olarak kabul etmek ne derecede mantıklı olabilir ki?**

Bađıl deđerlendirmeyi tercih eden kültürlerde bile bu tip deđerlendirmeyele deđerlendirilemeyecek durumlara bakılınca, deđerlendirme türünün seçiminin eđitimin niteliđi aısından ne kadar önemli olduđu biraz daha belirginleŐecektir.

Öđrenilecek konular arasında ařamalılık iliřkisi varsa yani bir bölüm izleyen bölümün ön kořulu niteliđindeyse,

Aralarında yođun sayısal ve mantıksal iliřkiler olan konularda, konunun anlaşılabilmesi için bu iliřkilerin tümüyle kavranması gerekiyorsa,

Dil öđrenme bařta olmak üzere, öđrenilenlerin iře kořulabilmesi veya iřlevsel olarak kullanılabilmesi için belli bir düzeyde performans sergilenmesi gerekiyorsa,

Tıp, pilot yetiřtirme, askerlik gibi görevin tam olarak yapılamamasının telafisi mümkün olmayan olumsuzluk riski tařıdıđı kritik alanlarda, görevin yapılabilmesi için belli bir düzeyde becerinin mutlaka kazanılması gerekiyorsa,

Niteliđin yüksek tutulmasının istendiđi, uygulamaların öngörölmüř standartlara göre yürütöldüđu ve **nitelikten taviz verilmemesi gereken öđrenme ortamlarında,**

bađıl deđerlendirme kullanılmaz.

Örneğin, derslerin yabancı dille yürütüldüğü programı izleyecek öğrencileri yetiştirecek dil hazırlık okulunda, mezunlar programı izleyecek kadar dil öğrenemedikten sonra o dil hazırlık programını açmanın bir anlamı yoktur. Hekim yetiştiren bir programda, hekimlerin yetersizlikleri nedeniyle kimse insan sağlığını risk altına sokmak istemez. Kimse iyi yetişmemiş bir pilota uçak emanet edip insanların hayatını tehlikeye atmaz.

Bu şartlar altında yanıtı aranacak, “Bağlı değerlendirmeyi kullananlar bunu neden tercih ediyorlar?” sorusunun yanıtı biraz da başka faktörlere bağlıdır.

Sistemler, bazı ön koşullara ve varsayımlara bağlı olarak tasarlanırlar ve uygulamada da bu koşullar veya varsayımlar gerçekleştiği oranda kendilerinden beklenen işlevi **olması gerektiği şekilde** yerine getirirler. Tıpkı demokrasi gibi. Belli bir kültürel olgunluğa ulaşmış toplumların elinde demokrasi bir fazilet rejimiyken, toplum belli bir olgunluğa ulaşmamışsa kolaylıkla bir rezalet rejimine dönüşebilir. Mesajın daha net anlaşılması için, küçücük çocukların eğitildiği bir anaokulunda, sınıf öğretmeninin **ders süresinin ne kadar, teneffüs ya da oyun süresinin ne kadar olması gerektiği konusunda** bir oylama yaptığını ve çocukların iradeleriyle demokratik bir biçimde karar verildiğini düşünelim bir an...

Öğrenenlerin, öğrenme süreçlerinde daha iyi öğrenmek için samimi bir şekilde ellerinden gelenin en iyisini yapmaya çalışacakları, sınavlarda da soruları doğru şekilde yanıtlayabilmek için ellerinden gelen tüm çabayı samimi bir şekilde sarf edecekleri, bağlı değerlendirmenin temel varsayımdır. Böylelikle dinamik ve gelişen standartlara göre yapılan değerlendirmelerle, **nitelik sürekli yükselecektir.**

Sınavdan yüksek puan alanların ortalamayı yükselttikleri için eleştirildikleri hatta arkadaşları tarafından dışlandıkları, bazı sınavlardan önce sınıftaki öğrencilerin hep birlikte düşük puan almak üzere anlaşış ortalamayı düşür-

dükleri bir sınıf kültüründe bu varsayımın ne kadar doğru olduđu takdirlerinize kalmıřtır⁷².

Ölçme deęerlendirme yapmanın özel gerekçeleri

Eđitim öğretim ortamlarında, sadece süreç sonlarında öğrencilerin başarılı olup olmadıklarına karar vermek amacıyla yapılmıř gibi görünen ölçme deęerlendirme etkinlikleri, aslında **üç farklı temel gerekçeyle** yapılır.

1-Sürece başlamadan önce öğrencilerin durumlarını görmek, ön koşul öğrenmelerinin ne durumda olduđunu belirlemek için, kısacası eđitsel anlamda **öğrencileri tanımak** için ölçme deęerlendirme yapılabilir. Bu deęerlendirme daha kapsamlı şekilde, bir programa başlamadan önce, öğrencilerin o programı izleyip izleyemeyeceklerini öngörmek ve sonucuna göre öğrencileri o programa kabul etmek veya **yerleřtirmek** için de yapılabilir. Sürecin başında bu amaçlarla yapılan ölçme deęerlendirmeye “**tanıma ve yerleřtirme**” için ölçme deęerlendirme denir.

2-Öğretim süreci belli bir zaman alır ve bu zaman diliminin sonuna gelindiğinde yapılan ölçme deęerlendirmelerle saptanan “süreçten kaynaklanan” yetersizliklere müdahale olanađı kalmaz. Bu nedenle, öğretim süreci esnasında programın öngörüldüğü şekilde uygulanıp uygulanmadığını saptamak, öğrenci gelişimini ve öğrenme düzeyini izlemek için, **süreç devam ederken** ara ara ölçme deęerlendirmeler yapılır. Öğrencilerin planlanan şekilde **yetiřtirilip yetiřtirilmediđini, programlara bir müdahalenin**

72 Meslek hayatım boyunca çok arzu etmeme rađmen, kopyanın bir ölçme hatası olduđunu bile bile, bazı deęerleri yaşatabilmek için bu ölçme hatasını göze alma pahasına, yıllarca gözetmensiz sınav yapamayışımın tek nedeni, gene kendi öğrencilerimin bunun karşısında durmaları olmuřtur. Anonim cevaplarla yaptıđım anketlerle sabit olmak üzere, gözetmensiz sınav yapmama karşı olan öğrencilerin genel gerekçesi hep řuydu: “**Asla yapmayalım hocam, ben kopya çekmem ama arkadařım kopya çekeceđi için benim hakkım yenmiř olur.**” Kimsenin ben kopya çekerim demediđi sınıflarda kopya çekmesinden korkulan kimlerdir sorusu yanıtızdır.

gerekli olup olmadığını görmek için yapılan bu ölçme değerlendirmeye de “**yetiştirme ve biçimlendirme**” için ölçme değerlendirme denir.

3- Süreç sonunda öğretimin etkililiğini saptamak ve **nihai başarı ya da performans düzeyini** belirlemek için (sınıf geçme / diploma-sertifika ya da sınıf tekrarı gibi nihai kararlar verebilmek amacıyla) yapılan ölçme değerlendirmeye de “**sonuç görmeye yönelik veya toptan/nihai**” ölçme değerlendirme denir.

Ölçülmek istenen davranışlar ve bunları ölçen araçlar

Bireylerin ölçülmek istenilen davranışları, sergilenme nedenlerine göre “**tipik davranışlar**” ve “**maksimum yeterlik davranışları**” olmak üzere iki temel gruba ayrılır.

Tipik davranışlar

Kişilerin okudukları gazete tercihleri, seçtikleri ders çalışma saatleri, tercih ettikleri ulaşım araçları, belli derslere karşı tutumları ve ilgi duyduğu alanlar gibi günlük yaşamlarında sergiledikleri, kesin olarak doğru ya da yanlış şekilde nitelenemeyen ve çoğu kez sonuçları belli düzenlemeler için kullanılan davranışlarına tipik davranışlar denir. Tipik davranışları ölçmek için kullanılan araçlar da, genel olarak “ölçek” olarak adlandırılırlar ve başlıca örnekleri “**Kişilik Ölçekleri**”, “**İlgi Ölçekleri**” ve “**Tutum Ölçekleri**”dir. Bu tür davranışları ölçen ölçme araçlarına genel olarak “**Tipik Davranış Testleri**” de denir.

Maksimum yeterlik davranışları

Kişilerin, öngörölmüş belli bir performansı, belli bir durumda ya da koşulda, en iyi veya en doğru biçimde ortaya koyabilmek amacıyla sergiledikleri davranışlar da vardır. Bu tür davranışları sergileyen birey, söz konusu davranışı

istenildiđi ya da dođru Őekilde sergileyebilmek iin elinden gelen maksimum abayı sarf eder. Örneđin, bir sınavda kendisinden istenilen cevapları en dođru Őekilde vermeye alışır. Yüz metre koŐarken bu mesafeyi en kısa sürede tamamlamaya gayret eder. Bu tür davranıŐlar da maksimum yeterlik davranıŐları olarak adlandırılır ve bu davranıŐları ölçmek için kullanılan araçlara da “Maksimum Yeterlik Testleri” denir.

Maksimum yeterlik davranıŐları da **bu davranıŐların dođasına göre** alt kategorilere ayrılır.

Genel yetenek

Bireyin öğrenme ve sorun çözmeye kapasitesinin göstergesi sayılabilecek, sayısal ya da mantıksal işlemler yapabilme, sözcük bilgisi ve dili kullanabilme, Őekiller arası ilişkileri görebilme gibi sahip olması istenen davranıŐlarına **genel yetenek** denir. Bu davranıŐların düzeyini ölçmek için kullanılan araçlara da “Genel Yetenek Testleri” denir.

Özel yetenek

Eđer bireyin yetenek kapsamındaki davranıŐları spor, müzik, resim gibi daha dar ve özel bir alana ilişkinse, bu kez bu davranıŐlar **özel yetenek** olarak adlandırılır ve bunları ölçmek için geliştirilmiş araçlara da “Özel Yetenek Testleri” denir.

Başarı

Bir öğretim süreci sonunda, öğrenme ürünü olan ve bireye eğitim yoluyla kazandırılmış olan istenen davranıŐlar da **başarı** olarak adlandırılırlar. Başarıyı, yani öğretim süreçleriyle kazandırılmış davranıŐları ölçmek için geliştirilen ölçme araçlarına da “Başarı Testleri” denir.

Sınav ve test kavramları

Çoğu kez gündelik hayatta birbirinin yerine kullanılsa da, eğitim terminolojisinde sınav ve test kavramları arasında bir fark vardır. Başarıyı ölçmek için geliştirilmiş ve içinde alt bölümleri olan ölçme araçlarına **sınav** denir. Örneğin, üniversiteye giriş için kullanılan **lisans yerleştirme sınavı** gibi. Bir sınavın içinde yer alan belli konulardaki başarıyı ölçen alt bölümler de **test** olarak adlandırılır. Lisans yerleştirme sınavındaki **matematik testi** veya **fizik testi** gibi.

Başarı testleri

Başarı testleri de uygulamada **standart başarı testleri** ve **öğretmen yapımı testler** olmak üzere ikiye ayrılır.

Standart başarı testleri

Seçme, yerleştirme, eleme gibi amaçlar için hem konu alanı hem de ölçme değerlendirme uzmanları tarafından kapsamlı şekilde hazırlanan, sonuçları da geniş kapsamlı kıyaslamalara ve normatif değerlendirmelere olanak sağlayabilen testlerdir. LYS, TUS gibi sınavlardaki başarı testleri standart başarı testleri için örneklerdir.

Öğretmen yapımı testler

Çok yakından tanıdığımız, şimdiki kuşağın eski Hababam Sınıfı filmlerinde gördüğü “Çıkarın kâğıtları yazılı yapacağım.” repliklerinde geçen, daha sonraki dönemlerde teksir makineleriyle çoğaltılan ama artık çoğu fotokopi formunda olan testler ise öğretmen yapımı testler kategorisine girer. Ama öğretmenlerin hazırlayıp uyguladığı bu öğretmen yapımı testler her zaman yazılı şekilde olmaz.

Öđretmen yapımı testlerin türleri

Öđretmen yapımı testler, başımızdan çok geçtiđi için iyi bildiđimiz ve uygulanması hakkında oldukça deneyim sahibi olduđumuz test türleridir. Ancak bu bilinen test türlerinin burada tekrar açıklanmasının nedeni, okuyanlara, bu tür testlerin olması gereken ideal biçimleri ile yaşadıkları test deneyimlerini kıyaslayarak yapabilecekleri bireysel deđerlendirmeleri için ölçütler sunabilmektir. Bu ölçütlere dayalı deđerlendirmeler, bizlere, içinden geçtiđimiz eğitim sistemimizdeki ölçme deđerlendirme etkinliklerini daha adil yargılayabilme olanađı sunacaktır.

Sözlü testler

Bilinen en eski test türüdür. Bilinen uygulamasında, öğrenci tahtaya kaldırılır, bir dizi soru sorulur ve alınan cevaplara göre puan verilir. Sözlü testler, sınıfların dışında bir sınav kurulunun önünde de uygulanabilir.

Bazı eğitimciler; sözlü testlerin başarıdan çok, öğrencinin sözlü ifade becerisini veya konuşma yeteneđini ölçtüđünü, bu nedenle de başarının ölçülmesinde kullanılmaması gerektiđini savunurlar. Ancak sözlü testler sözlü ifade becerisinin, sözcüklerin sesletiminin (telaffuzunun), ana dilde ya da yabancı dilde konuşma becerisinin ölçülmesinde tek seçenektir. Görme özürlü ya da yazmaya engel özürlü gibi özel durumlarda da zorunlu olarak kullanılır. Yüksek lisans ve doktora tez savunmalarında ya da araştırma raporlarının sunumlarında da kullanılır. Eğitimde başarının ölçülmesi açısından genel kanı, sözlü testlerin bir tamamlayıcı olarak kullanılması yönündedir.

Sınıfta uygulanacak ideal bir sözlüden önce yazılı bir test planı ve uygulanmasına ilişkin bir de uygulama yönergesi hazırlanmalıdır. Yazılı planda, “testin amacı”, “hangi davranışların hangi içerikle yoklanacağıın (belirtke çizelgesi)”, “testte sorulacak sorular” ve “bu soruların kabul edilebilir

cevapları olacak anahtar kavramlar” yer almalıdır. Uygulama yönergesinde ise “sözlüde kaç soru sorulacağı”, “genel olarak verilecek düşünme süresi” ve “puanlama esasları” olmalıdır. Bu “uygulama yönergesi”, sözlüden önce öğrencilerle mutlaka paylaşılmalıdır.

Sözlü esnasında öğrencinin verdiği cevaplara ilişkin kısa notlar alınmalı, puanlamada ve öğrenciye verilecek dönütlerde bu notlardan yararlanılmalıdır. Öğrencinin dış görünüşü, konuşma akıcılığı ve ölçülmek istenen davranışla ilgisi olmayan konular asla puanlamaya yansıtılmamalıdır.

Eğiticiler, sözlü öncesinde ve esnasında öğrencinin heyecanını arttıracak, onu kaygılandırarak davranışlardan kaçınmalı, aksine güven telkin edici bir üslup sergilenmelidir. Sözlülerde, uzun cevap gerektiren az sayıda soru yerine, cevabı kısa olan fazla sayıda soru sorulmalıdır. Her soru sorulduktan sonra, yeterince düşünme süresi verilmeli, süre tanıma konusunda tüm öğrencilere adil davranılmalıdır. Mümkünse, birden fazla puanlayıcının olması, puanlamanın nesnellığı artıracaktır.

Uzun cevaplı yazılı yoklamalar

Hazırlanmasının kolay olması, puanlamada öğretmeni nesnellikle sınırlandırmaması ve başarılarının bu tür ölçme araçları ile ölçüleceğini bilen öğrencilerin daha ayrıntılı ve daha derinlemesine çalışacaklarına olan inanç gibi nedenlerle tercih edilirler. Ama hazırlanmalarındaki kolaylığın aksine, okunmaları ve puanlamaları oldukça zordur ve çok zaman alır.

Değişik alanlara ilişkin, “Kim?”, “Nerede?”, “Neden?”, “Nasıl?” ve benzeri soruların cevaplarının yanı sıra; betimleme, benzerlik, zıtlık, karşılaştırma, açıklama, tartışma, geliştirme, özetleme gibi konulardaki geniş bir yelpaze içinde, pek çok üst düzey bilişsel davranışı ölçebilir. En önemlisi de, öğrencilerin kendilerine özgü, orijinal görüş düşünce ve ürünlerini ortaya koymalarını mümkün kılar.

Diđer yazılı test tiplerine göre, řans başarısı, yani öđrencinin bilmediđi hâlde biliyormuř gibi görünmesi durumu oldukça düşüktür. Ayrıca kopya çekilmesi diđer testlere göre biraz daha zordur.

Diđer yandan, cevaplar “kesin doğru” veya “kesin yanlış” olarak nitelenemeyebilir. Bu nedenle puanlanmasının nesnelliđi tartışmaya açıktır. Hatta aynı sınav kâğıdı farklı öđretmenlerce, hatta aynı öđretmen tarafından farklı zamanlarda puanlandığında bile aynı puanı alamayabilir. Özellikle kalabalık grupların puanlanmasındaki zorluk, yorgunluk, konsantrasyon düşüklüğü, puanlamanın nesnelliliđini etkileyebilir. Sınav süresinin çođunun yazmaya ayrılması, birim zamanda yanıtlanacak soru sayısını sınırlar ve bu da testte ölçülecek davranıřların miktarını sınırlar.

Uygulamadan önce, mutlaka bir test planı hazırlanarak ölçülmesi amaçlanan kritik davranıřlarla sınav kapsamına giren konular listelenmeli, ardından ayrıntılı bir amaç-içerik matrisi (veya belirtke çizelgesi) hazırlanmalıdır. Çok soru verilip, aralarından öđrencinin istediklerini cevaplaması yerine, bütün soruların zorunu olarak cevaplanması istenen testler, öđrencileri birbirine göre daha doğru bir řekilde sıralar.

Ezberciliđi teşvik etmemek için kitaptaki konu başlıklarının aynen soru olarak yazılmasından kaçınılmalı, soruların dili açık seçik anlaşılır olmalı, önemli kısımlar altı çizilerek ya da koyu yazılarak vurgulanmalıdır.

Kaliteli soru sorabilmek için, sorulması planlanan soru sayısının en az üç katı soru hazırlanıp sorular bunların arasından seçilmelidir. Test için, en azından test süresini ve soruların puan deđerlerini içeren bir yönerge hazırlanarak sorulardan önce olacak řekilde sınav kâğıdının başına yazılmalı, öđrenciler de yönergeyi okumaları için uyarılmalıdır.

Her sorunun eşit ağırlıkta puanlandığı testlerde, sorulacak soru sayısı belirlenirken, soruların puan değerlerine ilişkin şu ayrıntıya dikkat edilmelidir: Sorunun puan değerleri, ya tam sayı olacak şekilde ya da kalanı devretmeyen ondalık sayı olacak şekilde ayarlanmalıdır. Örneğin, 100 tam puan üzerinden puanlanacak bir testte, 5, 8 veya 10 soru sorulmalıdır. Böylece her sorunun puan değeri tam sayılarla (5 soru için 20'şer puan, 10 soru için 10'ar puan) veya kalan kısmı devretmeyen sayılarla (8 soru için 12,5 puan gibi) ifade edilebilecektir. Aynı husus alt bölümleri eşit şekilde puanlanacak bir sorunun alt bölümleri için de göz önüne alınmalıdır.

Puanlamada nesneliği sağlamak için bir soru kâğıdı cevaplanarak, her sorunun cevabında puanlama için ölçüt teşkil edecek kavramlar, anahtar fikirler ve alternatif örnekler, daha da önemlisi bunlara kaçır puan verileceği açıkça belirtilmeli, puanlama bu esaslara göre yapılmalıdır. Nesneliğe katkı için sınav kâğıtları bütün olarak değil, soru soru puanlanmalıdır.

Ölçülmesi gereken nitelikler içinde değilse, yazı güzelliği, sayfa düzeni, kâğıdı kullanmadaki tertip düzen gibi konular puanlamayı etkilememelidir. Öğrencinin diğer özelliklerinin puanlamaya yansımaması için mümkünse kâğıtlar öğrencilerin isimleri kapatılarak veya isimlere bakılmadan puanlanmalıdır.

Kısa cevaplı testler

Cevabı bir sözcük, bir rakam, bir tarih, bir sözcük öbeği ya da en fazla kısa bir cümle olan sorulardan oluşan testlerdir. Bu tip testlerin soruları, soru cümlelerinin yanı sıra, boşluk doldurma tipinde de olabilir.

Kısa cevaplı testlerde kullanılacak sorular, aşağıdaki örneklerde görüldüğü şekillerde düzenlenebilir.

Güneş ışığını soğurarak bitkilerde karbon özümlemesini sağlayan ve bitkilere yeşil renklerini veren maddeye ne ad verilir?

[_____]

Güneş ışığını soğurarak bitkilerde karbon özümlemesini sağlayan ve bitkilere yeşil renklerini veren maddeye _____ adı verilir.

Fotosentez nedir?

.....

Aşağıdaki cümlelerin harflerle işaretlenmiş altı çizili öğelerinin cümlelerin hangi öğeleri olduklarını, öğeler için ayrılmış ve aynı harflerle işaretlenmiş boşluklara yazınız.

Öğretmen Oya'ya soru sordu.

A B C D

A: _____

B: _____

C: _____

D: _____

Aşağıda Türkiye'de yetiştirilen bazı ürünlerin adları verilmiştir. Bu ürünlerin en fazla yetiştirildiği coğrafi bölgelerin adlarını karşılarındaki boşluklara yazınız.

A. Fındık: _____

B. Tütün: _____

C. Buğday: _____

D. İncir: _____

E. Zeytin: _____

Bu tür testler, özellikle ilkelerin, kavramların veya adlandırılmalarının hatırlanmasını sınamak için tercih edilir. Cevabın kısa olması, çok soru sorulmasını mümkün kılar ve bu da testte yoklanacak davranışların sayısını artırır. Uzun cevaplı yazılı yoklamalara göre daha nesnelirler ve puanlanmaları da kolaydır.

Doğru bir şekilde uygulanabilmesi için, hatırlanması istenen bilgi, gerçekten hatırlanmaya değer kadar önemli bir bilgi veya kavram olmalıdır. Bu tür testleri hazırlayan öğretmenlerin sıklıkla yaptıkları iki yanlıştan birisi, kitap cümlelerinden rastgele bir kelimeyi çıkararak soru hazırlamaktır. Diğer yanlış da, soruyu hazırlarken düşündükleri cevabın tek doğru cevap olduğunu zannetmeleridir. Bu nedenle sorular olabildiğince kesin ve **tek bir doğru cevabı gerektirecek** şekilde ifade edilmelidir. Örneğin Ankara'nın başkent olduğu bilgisini sınama amacıyla, cevabı "*başkentidir*" şeklinde düşünülerek hazırlanmış ancak iyi ifade edilmiş aşağıdaki sorunun, farklı şekillerde ve her birisi de mantıken doğru olacak şekilde cevaplanması mümkündür.

*Ankara Türkiye'nin _____ [“bir ilidir”,
“İç Anadolu bölgesinde yer alır”, gibi]*

Cevap olarak doldurulacak boşlukların sayısı puanlama hesabı göz önüne alınarak belirlenmelidir. Ayrıca, sorulara eşit puanların verildiği testler; puanlamada, soru analizleri yapmada ve test geliştirme işlemlerinde önemli kolaylık sağlar.

Eşleştirmeli testler

Birbiriyle ilişkili iki grup halinde verilen bilgi öğelerinin, **istenen bir ölçüte göre doğru şekilde eşleştirilmesini gerektiren** sorularla hazırlanan testlerdir. Bu bilgi öğeleri ve aralarındaki ilişki örnekleri şu şekillerde olabilir:

“Terimler” ve “bunların ne olduklarının kısa açıklamaları”,

“Semboller” ve “bunların nelerin sembolü oldukları bilgisi”,

“İlkeler” ve “bu ilkelerin uygulanabileceği yerler”,

“Yazarlar” ve “bu yazarların eserleri”,

“Önemli olaylar” ve “bu olayların tarihleri”,

“Problemler” ve “bu problemlerin genel çözüm yolları”,

“Buluşlar” ve “buluşu gerçekleştiren kişiler”...

Bu tür testlere “cevabı hazır testler” denir. Çünkü cevaplar cevaplayanın önünde hazır olduğu için, cevaplayanın herhangi bir cevap üretmesine gerek yoktur; yapması gereken, sadece hazır olan cevaplar arasındaki doğru eşleşmeyi görebilmektir. Cevabı hazır testler aynı zamanda “nesnel testler” olarak da adlandırılır. Nesnel testlerde, puanlayıcının puanlamaya müdahale etme olanağı yoktur. Başka bir ifadeyle, nesnel bir testi kim puanlarsa puanlasın, cevapların gerektirdiği aynı puanı vermek zorundadır.

Örnek bir eşleştirmeli test sorusu aşağıdaki gibi olabilir:

Birinci sütundaki şehir adları, ikinci sütunda listelenen ülkelerin başkentleridir. Başkentlerin hangi ülkelere ait olduklarını, ülkenin önündeki harfi, başkent önündeki boşluklara yazarak belirtiniz.

- | | |
|---------------|-----------------|
| ___ Ankara | A. Avusturya |
| ___ Atina | B. Avustralya |
| ___ Budapeşte | C. Belçika |
| ___ Canberra | D. İtalya |
| ___ Ottawa | E. Kanada |
| ___ Roma | F. Macaristan |
| | G. Türkiye |
| | H. Yeni Zelanda |
| | I. Yunanistan |

Adlandırmada kolaylık olsun diye birinci sütuna “öncüller”, izleyen sütuna da “cevaplar” denir.

Bu tür testlerin sorularında, öncüllerde ve cevaplarda yer alan öğeler mutlaka benzeşik (aynı cinsten) öğeler olmalıdır. Yukarıdaki örnek üzerinden açıklamak gerekirse, öncüllerde yer alanların mutlaka “başkentler”, cevaplarda yer alanların da mutlaka “ülke adları” olmaları gerekir.

Doğru şekilde uygulanabilmesi için şu konulara dikkat edilmelidir: Cevaplayanın soruları, okuma süresini en aza indirmek için uzun ifadeler öncüllere yazılmalıdır. (Örneğin, açıklaması bir paragraf tutan olaylar ile bu olayların oluş tarihleri eşleştirilecekse; bir paragraf uzunluğunda olan olaylar öncüllere, çok daha kısa olan oluş tarihleri de cevaplara yazılmalıdır. Bunun nedeni, cevaplayanın önce bir öncülü okuyacağı, bu öncül ile eşleştireceği ögeyi bulmak için cevapların tümünü gözden geçireceği, bu işlemi eşleşen ögeyi bulana kadar tekrarlayacağı varsayımıdır. Öncülleri ve cevapları oluşturan sütunlar, sözcüklerden oluşuyorsa alfabetik sırayla, sayılardan oluşuyorsa büyüklük küçüklük sırasıyla verilmelidir.

Eşleştirmeli sorulardan oluşan testler, “cevabı hazır testler” veya “nesnel testler” kategorisindedir. Ancak cevabı hazır testlerin olumsuzluklarından birisi de, farklı oranlarda şans başarısına açık olmasıdır. Şans başarısı demek, cevaplayanın bir cevap üretmek zorunda olmaması nedeniyle, sorunun doğru cevabını bilmediği hâlde biliyormuş gibi davranabilme ihtimalidir. Konuya açıklık getirebilmek için, dört öncül ve dört cevap içeren eşleştirmeli bir test sorusunu ele alalım. Birinci öncülün, karşısındaki cevaplardan herhangi birisiyle doğru şekilde eşleşme ihtimali %25’tir. Bunun doğru biçimde eşleştirilmesinin ardından, ikinci öncülün kalan üç cevabın herhangi birisiyle doğru olarak eşleşme ihtimali %33.33’e çıkar. Bunun da doğru şekilde eşleştirilmesi durumunda üçüncü öncülün kalan iki cevapla doğru şekilde eşleşme ihtimali yarı yarıyadır. Bunun doğru eşleşmesi hâlinde, son öncülün soru olma değeri yoktur. Ancak şans başarısını azaltmak için eşleştirmeli testlerde cevap sayısı ile öncül sayısını aynı tutulmaz. Böylelikle şans başarısı biraz daha azaltılmış olur. Bu nedenle, bu tür testler için örnek olarak verilen, ülkelerle başkentlerinin eşleştireceği soruda cevapların (ülkelerin) sayısı, öncüllerin (başkentlerin) sayısından fazladır.

Doğru yanlış testleri

Soruları, doğru ya da yanlış olarak nitelenebilecek önermelerden oluşmuş testlerdir. Cevaplayandan, sorudaki önermenin doğru mu yoksa yanlış mı olduğuna karar vermesi istenir. Aslında, tipik bir doğru-yanlış testi sorusu; teknik olarak, iki seçeneği olan çoktan seçmeli bir test sorusudur. Bu testler de “nesnel” veya “cevabı hazır testler” kategorisine girdikleri için şans başarısına açık, hatta en açık testlerdir. Cevaplayanın, hiçbir şey bilmiyor olsa dahi, bir soruyu beklendiği şekilde cevaplama olasılığı %50 gibi gerçekleşmesi oldukça yüksek bir olasılıktır.

Bu tür soruların asıl kullanım amacı **kavram yanlışlarını sınamaktır**. Ders kitaplarından ya da derste kullanılan kaynaklardan kalıp halinde alınan cümlelerde yazım yanlışları yaparak, mevcut sayıların değerlerini değiştirerek, cümledeki yargıyı olumsuzla çevirerek sorular düzenlemek doğru değildir. Aşağıdaki örnek sorular, bu tür testlerde kullanılabilir soru tiplerine örnek olabilir.

<i>Birisi 1kg. diğeri 10kg. olan iki demir gülle, bir binanın balkonundan aynı anda bırakıldığında ve hava sürtünmesi de ihmal edildiğinde, ağır olan gülle yere daha önce ulaşır.⁷³</i>	<input type="radio"/> Doğru <input type="radio"/> Yanlış
<i>Bir kilogram demirin ağırlığı bir kilogram pamuğun ağırlığından fazladır.</i>	<input type="radio"/> Doğru <input type="radio"/> Yanlış
<i>Aracına sürekli olarak 100 liralık akaryakıt alan birisi, akaryakıt zamlarından hiç etkilenmez.</i>	<input type="radio"/> Doğru <input type="radio"/> Yanlış
<i>Ramazan günlerinde orucun açıldığı ilk ilimiz 01 plaka numaralı Adana, son oruç açılan ilimiz de 81 plaka numaralı Düzce'dir.</i>	<input type="radio"/> Doğru <input type="radio"/> Yanlış

73 Bu önemli bir kavram yanlışlığıdır. Bu soruyu DOĞRU olarak cevaplayan Fizik Öğretmeni adayları ile karşılaştığım için bu dipnotu eklemeye gerek duyuyorum. (Bu yaptığımı da eğitim terminolojisinde "fırsat eğitimi" derler.) Hava sürtünmesi ihmal edildiğinde, serbest düşen cisimler, kütlelerinden bağımsız bir şekilde aynı hızla yere düşerler. Hatta havası boşaltılmış bir cam boru içinde, metal bir para ile hafif bir kaz tütünün aynı anda borunun dibine düştüğünü, hatta çakıldığını izlemek, tütün de para gibi yere çarpmasını görmek pek çok kişiyi şaşırır. (Konuyla ilgili bir video için: <https://www.youtube.com/watch?v=tOOUH6rNQ3E>)

Bu tür testlerde cevaplama çok zaman almadığı için, sınavda sorulabilecek soru sayısı oldukça fazladır ve bu da sınavda yoklanacak davranış miktarını oldukça yükseltir.

Testteki her bir soru, açıkça tek bir fikri ve yargıyı belirtmeli ve bunun ifadesi de olabildiğince açık ve yalın olmalıdır. Her maddedeki yargı ya da fikir, “kesinlikle doğru” ya da “kesinlikle yanlış” olmalıdır. Yanlış olarak düzenlenen sorulardaki yanlışlık, önemsiz bir noktada değil; ilkede veya fikirde olmalıdır. Bu tür test sorularında olumsuzluk ifadeleri dikkatle kullanılmalı, “olmamış değildir” gibi iki olumsuz yapının yan yana kullanıldığı ifadelerden kaçınılmalıdır. Doğru ve yanlış yargı içeren soruların cümle uzunlukları arasında belirgin fark olmamalıdır. Testte yer alan doğru ve yanlış yargı içeren soru sayısı da, eşit ya da birbirine yakın olmalıdır.

Doğru olmayan ifadelerin veya yargıların sınavlarda kullanılmasının öğrenciler tarafından yanlış bilinenleri pekiştireceği görüşündeki eğitimciler tarafından bu tür testlerin kullanımı eleştirilmektedir.

Çoktan seçmeli testler

Cevabı, doğru cevabı da içeren bir dizi cevaplar listesi içinden seçilerek verilen sorulardan oluşan, sıklıkla kullanıldığı için de çok iyi bilinen testlere çoktan seçmeli testler⁷⁴ denir.

Soruların biçimsel yapısı nedeniyle, soruyu oluşturan bölümlerin kendine özgü adları vardır. Sorunun ne sorulduğunu açıkça belirten ya da sezdireni veya problemi sunan bölümüne **soru kökü** denir. Soru kökünün altında

⁷⁴ Bu testler ülkemizde kullanılmaya başlandıktan sonra, “test” kelimesi yeni bir anlama bürünerek öğretmen ve öğrenciler tarafından sanki özel bir ölçme aracının adını almış gibi kullanılmaya başlandı. Bu kelimenin zıt anlamını da “klasik” kelimesi üstlendi. Bu anlamlar, “Hocam final test mi olacak klasik mi?” diye soran üniversite öğrencileriyle, “İkinci yazılı test yapacağım” diyen ortaokul ve lise öğretmenleri tarafından iyice pekiştirildi. Oysa test kelimesi başarıyı ölçen tüm ölçme araçlarının **genel adıdır** ve burada sözü edilen testin asıl adı “**Çoktan seçmeli objektif** (veya nesnel) **testtir**”.

sıralanan, sorunun beklenen cevabı ile yanlış cevapları da içeren listeye **seçenekler** denir. Seçenekler listesinde yer alan ve sorunun beklenen cevabı olan seçeneğe **doğru cevap** denir. Seçeneklerde yer alan ve doğru cevap dışında kalan diğer seçeneklere **çeldiriciler** denir. Çeldiricilerin işlevi, sorunun cevabını vermek için yeterli bilgisi olmayanları yanıltmak ve doğru cevabın tesadüfen bulunmasını engellemektir. Bazı durumlarda; bir paragraf, bir şekil, tablo ya da grafik gibi ortak bir veri hakkında sorulmuş iki ya da daha fazla soru art arda bulunabilir. Birbiriyle ilişkili bu soru grubuna da **soru takımı** denir.

Çoktan seçmeli testlerin sıklıkla kullanılmasının bir nedeni de, eşleştirmeli ve doğru yanlış tipi test sorularının da çoktan seçmeli test sorularına kolaylıkla dönüştürülebilmesidir. Örneğin, aşağıda sırasıyla, çoktan seçmeli test sorusuna dönüştürülmüş bir doğru yanlış, bir de eşleştirmeli test sorusu örneği vardır.

Aşağıda, sizin doğru ya da yanlış şeklinde değerlendirmeniz istenen 5 yargı mevcuttur. Bunların sırasıyla doğru ya da yanlış oldukları, hangi seçenekte yer almaktadır?

- ① Dünyadaki tüm kıtaların adı aynı harfle başlayıp biter.
- ② 1 kg limondaki şeker miktarı 1 kg çilekteki şeker miktarından azdır.
- ③ Dünyanın en hızlı büyüyen bitkisi bambudur
- ④ Salatalığın yarısından fazlası sudan oluşur
- ⑤ Dünyanın en zengin 3 ailesinin serveti, en fakir 48 ülkenin toplam servetinden fazladır.

- a)① Y ②Y ③D ④Y ⑤D
- b)① D ②Y ③D ④D ⑤D
- c)① D ②D ③Y ④Y ⑤Y
- d)① D ②D ③Y ④D ⑤Y
- e)① Y ②D ③D ④D ⑤Y

Çağdaş bir öğretim programında olması gereken dört temel boyut ile bu boyutlara ilişkin açıklamalar hangi seçenekte doğru şekilde eşleşmiştir?

AÇIKLAMALAR		BOYUTLAR	
A	Öğretim süreci sonunda kazandırılması ya da değiştirilmesi öngörülen davranışlar	I	AMAÇ
B	Öngörülen davranış değişikliğinin ne derecede gerçekleştirilebildiğinin saptanması	II	İÇERİK
C	Davranış değişikliğinde sadece araç olarak kullanılacak konular	III	SÜREÇ
D	Davranış değişikliğinin ya da yeni davranış kazandırmanın fiilen nasıl gerçekleştirileceği	IV	DEĞERLENDİRME

- a) A-III B-IV C-II D-I
 b) A-II B-IV C-I D-III
 c) A-I B-IV C-II D-III
 d) A-I B-IV C-III D-II
 e) A-II B-IV C-III D-I

Çoktan seçmeli testler de cevabı hazır testler kategorisine girerler. Bu nedenle şans başarısına açıktırlar. Soruyu cevaplayan birisi, hiçbir şey bilmiyor olsa dahi, her bir soru için, 100/seçenek sayısı kadar yüzdeler bir ihtimalle doğru cevabı bulabilir. Örneğin bir çoktan seçmeli test sorusunda 5 seçenek varsa, doğru cevap hakkında hiç fikri olmayan bir öğrenci ($100/5=20$) %20 ihtimalle doğru cevabı biliyormuş gibi davranabilir. Buna rağmen, başta çok kalabalık kitleler için yapılan standart sınavlarda olmak üzere, tüm kademelerdeki okullarda da yaygın biçimde kullanılmasının iki önemli ne-

deni vardır: Bunlardan birincisi, sınavın nesnel olması, diđeri de çok fazla sayıdaki sınav kađınının, (optik okuyucularla veya şablonlarla) kısa sürelerde puanlanabilmesidir.

Ancak bu kadar yaygın olmasının, eđitsel açıdan neden olduđu önemli bir olumsuzluk da vardır.

Üniversite seçme sınavları, ortaöđretime geçiş sınavları, tıpta uzmanlık sınavları gibi; kişiler için önemli olan pek çok sınavda kullanılması, her kademedeki öđretmeni, **öđrencilerini çoktan seçmeli test maddesi cevaplama konusunda yetiştirmeye yöneltmekte, hatta bu durum bazı zamanlarda öđretimin amaçlarının önüne dahi geçebilmektedir.**

Bu tür testler için soru hazırlamak, konu hakkında teknik bilgi sahibi olmayı gerektirir. Özellikle de gerçekten çeldirici işlevi olan seçenekler üretmek oldukça zordur ve çok zaman alır. Ancak bu testlerin puanlanması nesnel ve çok kolaydır. Ayrıca, elde edilen puanlar üzerinde birçok istatistiksel işlemler yapmak mümkündür. Cevaplayan açısından bakıldığında da, okuma becerisi ve hızı önemli rol oynar.

İyi uygulanması için şu noktalara dikkat edilmelidir: Daha seçenekler okunmadan soru kökünde fark edilen tek ve temel bir fikir bulunmalı, bununla ne yapılacağı açıkça ifade edilmelidir. Seçeneklerde tekrarlanan ifadeler soru köküne alınarak gereksiz tekrarlardan kaçınılmalıdır. Mümkünse her bir seçeneğin uzunluğu eşit veya birbirine yakın olmalıdır. Seçeneklerde yer alanlar, sözcüklerden oluşuyorsa alfabetik olarak, sayılardan oluşuyorsa büyüklük küçüklük, tarihlerden oluşuyorsa geçmişten günümüze veya tersi şekilde sıralanmalıdır.

Öđretmen yapımı çoktan seçmeli testlerde yapılan ciddi bir hata da, seçenekler arasına “hepsi” ya da “hiçbirisi” ifadelerinin konmasıdır. Seçenekleri bu şekilde düzenlenmiş sorularda, seçeneklerin doğru olma ihtimalleri farklılaşır. Ayrıca uygulama sonuçları da, bu tür seçenekleri olan soruların bilenle bilmeyeni ayıramadığını ve daha fazla

ölçme hatasına yol açtığını göstermektedir⁷⁵. Doğru cevaplar farklı seçeneklere eşit olarak dağıtılmalı, bu dağılım da belli bir sistematığe göre değil, rastgele olarak yapılmalıdır. Soru köklerindeki olumsuz ifadeler de altı çizilerek ya da koyu yazılarak vurgulanmalıdır.

Eğitsel yetersizlikleri ve çözüm önerileri

Çoktan seçmeli testlerin iki önemli yetersizliği vardır. Bunlardan birisi şans başarısına açık olması, yani bir soruya verilen doğru cevabın gerçekte öğrenilmiş bir davranış mı yoksa tesadüfi olarak sergilenmiş bir davranış mı olduğunun bilinmemesidir. Diğer yetersizlik de, kısmi öğrenmeleri ölçmemesi yani, bir soruyu cevaplarırken beş seçenek arasında kalarak yanlış cevap veren bir öğrenci ile cevabı iki seçeneğe kadar indirip yanlış cevap veren bir öğrenciyi diğerinden ayıramamasıdır.

Düzeltilme faktörü

Şans başarısı sorununa çözüm olarak önerilen yol, her ne kadar eğitim terminolojisinde düzeltme faktörü olarak adlandırılrsa da bunun gerçekte düzelttiği bir şey yoktur. Sadece öğrenciyi bilmediği soruyu cevaplamaktan caydırmaya yarar. Bu caydırmayı da öğrenciyi tehdit ederek yapar. Öğrenciye şu mesajı verir: “Bilmediğin soruyu rastgele bir seçeneği işaretleyerek, bilindik ifadesiyle ‘kafadan atarak’ cevaplamaya kalkarsan ve o da tutmazsa seni cezalandırırım. Bunu da senden puan kırarak yaparım.”

Ancak düzeltme faktörü bu cezalandırmayı, yani puan kırmayı adil biçimde yapmaya çalışır. Şöyle ki, beş seçenekli bir soruda, hiçbir şey bilmeyen bir öğrencinin doğru cevabı işaretleme ihtimali ($100/5=20\%$) yüzde yirmidir. Yani beş seçenekli bir test sorusunda rastgele bir seçeneğin işaretlenmesi 20% ihtimalle doğru cevap olabilir. Bu durumda, bir öğrenci, bilemediği bir soruyu “attığında” ve o da tutmadı-

⁷⁵ Haladaya, T.M. and Downing, S.M. (1989) **A taxonomy of multiple-choice item writing rules**. Applied measurement in education. 2(1), 37-50

ğında, sorunun tam puanının %20'si kadar bir puanla cezalandırılması olağandır. Bu mantıkla bakıldığında beş soru kafadan atıldığında bunlar yüzünden kaybedilmesi olağan sayılan puan toplamı, bir tam sorunun puanına eşit olur ancak bu beş seçenektan birisi de zaten doğru cevaptır. O nedenle çoktan seçmeli testlerde seçenek sayısı kadar değil de (doğru cevap hariç tutulmak üzere) **seçenek sayısının bir eksiği kadar yanlış işaretleme, bir sorunun tam puan değeri kadar puanın kırılmasıyla cezalandırılır.** O nedenle “Dört yanlış bir doğruyu götürüyor mu?” klasik sorusu her zaman değil, sadece beş seçenekli çoktan seçmeli testler için doğru bir sorudur. Düzeltme faktörünün doğru ifadesi de, **“seçenek sayısının bir eksiği kadar yanlışın, bir doğru cevabı geçersiz kıldığı bir uygulamadır.”** Düzeltme faktörü uygulandığında, genellikle dört seçenekli soruların sorulduğu ilköğretimde üç yanlış bir doğruyu götürürken, beş seçenekli soruların sorulduğu ortaöğretim ve yükseköğretimde dört yanlış bir doğruyu götürür.

Özel puanlama sistemleri

Çoktan seçmeli testlerin, kısmi öğrenmeleri ölçmemesi (beş seçenek arasında kalan ile iki seçenek arasında kalarak kaybedeni ayıramaması) sorununun da birkaç çözüm yolu vardır.

En zor olanı, her bir sorunun seçeneklerinin, sorunun cevabının doğruluk derecesini yansıtacak şekilde ağırlıklandırarak düzenlenmesidir. Örneğin, beş seçenekli bir sorunun (a) seçeneği %100 doğru, (b) seçeneği %75 doğru, (c) seçeneği %50 doğru, (d) seçeneği %25 doğru ve (e) seçeneği de tamamen yanlış olacak şekilde düzenlenebilirse, öğrenci, o sorudan, sorunun tam puanının işaretlediği seçeneğinin ağırlığı kadarını alır. Örneğin, bu örnekte %50 ağırlıklı (c) seçeneğini işaretleyen bir öğrenci o sorudan alınacak tam puanın yarısını, %25 ağırlıklı (d) seçeneğini işaretleyen bir öğrenci, dörtte birini, %100 doğru (a) seçeneğini işaretleyen bir öğrenci de sorunun puanının tamamını alır. Ancak böy-

lesi testler hazırlamak çok zordur ve muhtemelen böyle bir test ile yapılan ölçme, hazırlanmasındaki zorluklar yüzünden, kısmi öğrenmeleri ölçmemekten kaynaklanan ölçme hatalarından daha fazla ölçme hatası içerebilir.

Biraz daha kolay olan diğer yol da, cevaplayanın, doğru olduğunu düşündüğü seçeneği işaretlemesine ek olarak, bir de **emin olma derecesi** eklemesidir. Bu amaçla, her soru için iki işaretleme alanı düzenlenir, ilk alana, (a), (b), (c)... gibi seçenek işaretlenirken, hemen yanındaki alana da %100, %75, %50, %25 gibi emin olma derecesi işaretlenir. Eğer işaretlenen seçenek doğru cevapsa, o sorudan soru tam puanının emin olma derecesi kadarı alınır. Yani, verilen cevabın yanına emin olma derecesi %50 olarak işaretlenmişse ve o cevap da doğruysa, o sorudan alınacak puan, sorunun tam puanın %50'si, yani yarısı kadardır. Bu uygulamada düzeltme faktörü kullanılacaksa, göze alınan riskle orantılı olarak uygulanır. İşaretlenen seçenek doğru cevap değilse, bu kez sorunun tam puanının emin olma derecesi kadar puan, toplam test puanından düşülür.

Bu konuda çözüm olabilecek son yol ise, testi cevaplayana birden fazla seçenek işaretleme olanağı sunulmasıdır. Ancak bir koşulla. O koşul da, **eğer işaretlenen seçeneklerin arasında doğru cevap varsa, o sorudan alınacak puanının işaretlenen seçenek sayısına bölünerek verilmesidir**. Eğer cevaplayan eminse tek seçenek işaretleyerek o sorudan tam puan alırken, çok emin değilse, iki seçenek işaretleyecek ve doğru cevabı bulursa sorunun puanın yarısını alacaktır. Bu uygulama hakkında şöyle bir soru akıllara gelebilir. Eğer cevaplayan, beş seçenekli soruların olduğu bir testte, soruların hepsinde bütün seçenekleri işaretlerse, hiçbir şey bilmeseyse bile, o testten alınacak tam puanın beşte birini almaz mı? Evet alabilir. Ancak bu ölçme açısından bir sorun teşkil etmez. Çünkü davranış ölçümündeki sıfır, **mutlak sıfır değildir** ve sınavda **davranışın ölçülmeye başlandığı noktayı** gösterir. Bir testte ölçmeye sıfır yerine 20 puan noktasından başlanmış olması, tüm öğrenciler için ölçmenin başlangıç

noktasını aynı kıldıđından, yine de öđrenciler birbirlerine göre sorunsuz bir řekilde sıralanmıř olurlar.

En az hata ile ölçme

Davranıř ölçümünün sınavlar aracılıđı ile “**dolaylı ölçme**” řekilde yapılması, ölçmeye hata karıřmasına neden olmaktadır. Ölçmeyi yapan kiřiden, ölçme aracından, ölçmenin yapıldıđı ortamdaki davranıřları ölçülen kiřiden ve kullanılan ölçme yönteminin uygun olmamasından kaynaklanan hatalar göz önüne alındıđında, eđitimde davranıř ölçmenin hatasız bir řekilde yapılabilmesi neredeyse olanaksız görünmektedir.

Ölçmede **hata, varlıđın sahip olduđu gerçek özellik miktarı ile ölçme sonucu elde ettiđimiz özellik miktarı arasındaki farktır**. Bu fark, gerçekten bir hata ise, yani kasten yapılmamıřsa, kontrolümüzün dıřında geliřmiřtir. Bu nedenle, hem hatanın ne kadar olduđunu, hem de hatalı ölçme sonucunun, özelliđin gerçek deđerinden az mı, yoksa fazla mı olduđunu bilemeyiz. Bařka bir ifadeyle, elinden gelen tüm çabayı sarf ederek, dikkatli bir řekilde ölçme yapan bir eđitici; ne “Ben hatasız ölçme yaptım.”, ne de “Bu ölçmeye karıřan hata da řu kadardır.” diyebilir. Bir eđitim profesyoneli, yaptıđı davranıř ölçümünün sonuçlarına bakarken “ölçmeye bir miktar hata karıřtıđını” ve “bu hatanın hem yönünü (gerçek özellik miktarından fazla mı yoksa az mı olduđunu) hem de miktarını (ne kadar olduđunu) bilemediđini” bilir.

Hatasız ölçme yapmak neredeyse olanaksızdır ama ölçmeye karıřan hatayı en aza indirmenin bir yolu vardır. “Nasıl bir yol?” sorusuna yanıt bulmak için; bir anlıđına, çok hassas bir řerit metreyle bir masanın enini ölçmeye çalıřtıđımızı düşünelim. Ölçme sonunda bulunan deđer, ölçmeye karıřan hata nedeniyle, masanın gerçek eninden **farklı** olacaktır. Ancak hatanın hem büyüklüđünü hem de yönünü kontrol edemediđimiz için, ölçme sonucu elde edilen deđerin, masanın gerçek eninden az mı yoksa fazla mı olduđunu

ğunu da bilemeyiz. Bu noktada çözüm, art arda ölçümler yapmaktır. Aynı varlık üzerinde art arda yapılan ölçümlere, her defasında, kimisi gerçek değer üzerinde olacak şekilde, kimisi de gerçek değer altında olacak şekilde hatalar karışacaktır. Bu ölçümlerin ortalamasını aldığımızda elde edilen değer, yapılan art arda ölçmelerin sayısının çokluğu oranında, olabildiğince hatadan arınmış bir değer olacaktır.

Nasıl mı? Yönü ve miktarı rastgele biçimde ortaya çıkan hataların bir kısmı (olduğundan fazla ölçümler şeklindeki) pozitif hatalarken, bir kısmı da (olduğundan az ölçümler şeklinde) negatif hatalardır. Art arda yapılan ölçümlerin ortalaması alındığında, yönleri ve büyüklükleri farklı hatalar birbirini götürüleceği için; ortalama değer, gerçek değere **en yakın değer**⁷⁶ olarak ortaya çıkacaktır.

76 Eğer aynı varlığa ilişkin çok sayıda art arda ölçmeler yapar ve bunların sonuçlarının ortalamasını alırsanız, elde ettiğiniz bu değeri, hatadan olabildiğince arındırılmış “**ölçmenin gerçek sonucu**” varsayabilirsiniz. Bu, size art arda yaptığınız ölçmelerin her birisinin sonucu ile ortalama olarak ulaştığınız **ölçmenin varsayılan gerçek sonucu** arasındaki farkları hesaplama imkanı sağlar. Bu farklar dizisi, yaptığınız her bir ölçmeye karışan hataların miktarlarını ve yönlerini gösterir. Her bir ölçmeye karışan hatalar hesaplandıktan sonra, bunların da ortalaması alınarak, ölçmelere karışan **ORTALAMA HATA** miktarı bulunabilir. Ama bir sayı dizisinin ortalama değeri, o dizinin denge merkezi gibidir, yani, ortalama, ortalamayı gösteren noktanın üzerinde ve altında, **değerlerinin toplamı eşit olan negatif ve pozitif sayılar bırakır**. O nedenle ortalamadan farkların ortalamasını hesaplamaya kalktığınızda, bir kısmı negatif bir kısmı pozitif ama değerleri eşit olan bu **farkların toplamı daima sıfır çıkar**. Ama bu durumda da ortalama hatayı hesaplamının matematiksel bir yolu vardır ve sonuç $\sqrt{16}$ gibi **bir sayının karekökü** olarak hesaplanır. Bunun anlamı şudur; ölçmeye karışan hata ± 4 'tür. Yani ortalama hata 4 birimdir ve bu değer pozitif de olabilir negatif de. Çünkü ölçme hatalarının bir kısmı **gerçek değerden fazla**, bir kısmı **da gerçek değerden azdır**. Bu şekilde hesaplanan **ölçmeye karışan ortalama hataya “ÖLÇMENİN STANDART HATASI”** denir. Üstelik sınıflarda uygulanan testler için, aynı ölçme aracıyla, 50 kişiye bir defada uygulanan test, sanki aynı ölçme aracı kullanılarak art arda 50 ölçme yapılmış gibi ele alınarak, ölçmeye karışan standart hata hesaplanabilir. Ancak bunun için, o ölçme sonuçlarının **güvenirlik katsayısının** ve **standart sapmasının** bilinmesi gerekir. Sonra da belli ihtimaller dahilinde, bir öğrencinin hatadan arındırılmış puanı hesaplanabilir. Hatanın yönü bilinemediği için, hatadan arındırılmış puan, öğrencinin test puanının, ortalama hata kadar altında ve üzerinde olmak üzere **bir aralık** şeklinde ifade edilir ve buna da **güven aralığı** denir. Konunun ayrıntısı ölçme değerlendirme veya temel istatistik kitaplarında bulunabilir.

Art arda yapılan ölçmeler, ölçme sayısının çokluğu oranında, farklı nedenlerden kaynaklanan hatalardan arınacaktır. Bir oturumda yapılan ortaöğretim kurumları geçiş sınavı ile üç oturumda yıllara yayılmış şekilde yapılan geçiş sınavlarını kıyaslamak konunun önemini belirginleştirecektir.

Sonuç olarak, olabildiğince hatadan arınmış ölçme, sürece yayılmış ölçmedir. Bir öğrencinin değerlendirilmesinde ideal ölçme, mümkünse her ders saatinin sonunda yapılan ölçmelerin ortalaması olan ölçmedir.

Ölçmeler sürece yayılabildiği ve tekrarlanabildiği oranda; üniversitelerde bir vize bir final, ortaöğretim kurumlarında üç yazılı şeklinde yapılan ölçmelere karışan hatalar azalacaktır. Daha açık bir ifadeyle, bir vize bir final şeklinde yapılan ölçmelerde, final günü yaşadığı kişisel bir sorun nedeniyle başarısız olan bir öğrenci, art arda yapılan ölçmelerin olduğu durumda, böyle bir olumsuzluktan çok etkilenmeyecektir.

Ölçmenin sürece yayılması, ölçme hatalarının etkisini azaltmasının yanı sıra, ölçülecek davranışların miktarını da etkileyerek, daha kapsamlı ve daha ayrıntılı, dolayısıyla daha doğru davranış ölçmeyi de mümkün kılar.

Bir dönem içinde, öğrenenlere, 250 farklı davranış kazandırılmasının amaçlandığını varsayalım. İdeal bir ölçme için, bu 250 davranışın 250'sinin de tek tek yoklanması gerekir.

Bir an için bütün davranışları kapsayan, her davranış için dört soru hesabıyla, 1000 soruluk yazılı bir sınav yapmaya kalktığımızı düşünelim. Her şeyden önce, kişilerin odaklanma sürelerindeki sınırlılıktan kaynaklanan güçlükler yaşanacaktır. Ayrıca, bu kadar çok soruyu içeren ve yapılması da bir o kadar zaman alacak olan böylesi bir sınavın okunması ve nesnel olarak puanlanması da zor olacaktır. Bu yüzden, uygulamada bu tür sınavlar yapılmaz.

Uygulamada, **kazandırılan davranışların tümüyle değil, tüm davranışları temsil ettiği varsayılan bir davra-**

nış grubuyla ölçme yapılır. Tüm davranışları temsil ettiği varsayılan bu seçilmiş davranış grubuna “örneklem” denir. Buradaki temel sorun, seçilen davranış örnekleminin, kazandırılan tüm davranışları ne kadar temsil edebildiği, tüm davranışlara ne kadar genellenebildiğidir. Öyle ya, üniversitelerde her yarıyıl bir vize, bir final sınavı yapıp sorulan 15-20 soruya verdiği cevaplarla, öğrencinin bir dönemde öğrendiği her şeyi ölçüp değerlendirdiğimizi varsayıyoruz. Ama öte yandan “Her konuya çalıştım, sadece son konu kalmıştı, hoca da tuttu hep oradan sordu, diğer konulardan sorsaydı kesin geçerdim.” gibi öğrenci yakınmalarını da çok duyuyoruz.

Sonuç olarak; ölçmenin zamana yayılması, sürecin içinde sık sık yapılması, ölçülecek davranışların kapsamını arttırırken ölçmeye karışacak hataları da azaltarak ölçmenin kalitesini yükseltir.

Ölçmede yeni yaklaşımlar

Günümüzde, eğitim uygulamalarındaki yenilikleri ölçmeye yansıtma ve mevcut ölçme sorunlarını en aza indirmek, daha kapsamlı, daha hatasız ve daha işlevsel ölçmeler yapabilmek için pek çok yeni yaklaşım önerilmiştir. Bu yeni yaklaşımlara eğitim terminolojisinde “Alternatif Değerlendirme”, “Özgün (Ötantik) Değerlendirme” veya “Performans Değerlendirmesi” gibi değişik adlar verilmektedir.

Alternatif Ölçme-Değerlendirme ile geçmişten günümüze kadar kullanılagelen sözlüler, yazılı yoklamalar, çoktan seçmeli testler gibi geleneksel değerlendirme araçlarının dışında kalan araçlarla yapılan ölçme değerlendirme kastedilmektedir. Örneğin, bireylerin kendi kendilerini değerlendirmeleri için “**öz değerlendirme formu**”, grup arkadaşlarını değerlendirmeleri için “**akran değerlendirme formu**”; bireyin çok kapsamlı bir şekilde değerlendirilmesine olanak tanıyan ve eğitsel değeri olan her türlü ürününün

bir araya getirildiđi “**ürün dosyası**”, bireyin eđitsel açıdan önem taşıdığı için kayıt altına almaya deđer önemli yaşantılarının kaydedildiđi “**anekdot kaydı**” gibi yeni ölçme araçlarıyla yapılan ölçmeler, bu kapsamda düşünülebilir.

Özgün (Otantik) Ölçme-Deđerlendirme ise okulda öğrenilenler ile gerçek yaşamdaki olaylar arasında ilişkiler kurulmasına yönelik eđitsel çalışmaların, yine bu bakış açısıyla deđerlendirilmesi anlayışıdır.

Performansın Deđerlendirilmesi de, olabildiđince gerçekçi bir deđerlendirme yapabilmek için ölçmeyi, bir bağlam içinde yapmaktır.

Performansa dayalı ölçme yaparken, öğrenenin sadece bilgisini deđil, bunun yanı sıra becerisini ve tutumlarını da ölçebilmek için ona konuyla ilgili bir görev verilir. Öğrenenin, verilen görevi yerine getirirken kullandığı bilgisi, sergilediđi becerisi ve tutumları; uygun ölçme araçlarıyla ölçülür. Performans deđerlendirmesinden beklenti, öğrenenlerin okulda öğrendiklerini yaşama aktarmalarını sağlamak; problem çözme, eleştirel düşünme ve yaratıcılık gibi yaşamın her alanında kullanılacak zihinsel süreçlerini geliştirmektir.

Ölçmede geçerlik ve güvenilirlik

Geçerlik ve güvenilirlik, davranış ölçme alanında sıklıkla karşımıza çıkan iki terimdir. Bunlar her ne kadar pek çok kaynakta **ölçme araçlarının özellikleri** olarak açıklansa da, “ölçmenin geçerli olması”, “ölçmenin güvenilir olması” anlamında, **yapılmış olan bir davranış ölçümünün özellikleri** olarak da düşünölmelidir.

Ölçmenin tanımı, “varlığın ilgilenilen özelliđi” şeklinde başlar. Yani, ölçmeyi yapan varlığın sahip olduđu pek çok özelliđi arasından varlığın ilgilendiđi bir özelliđini seçerek onun miktarını, derecesini ortaya koymaya çalışır. Ama, oldukça zor ve karmaşık bir süreç olan davranış ölçümünde, çođu kez ilgilenilen özelliđe başka özellikler de karışır. Hatta bu durum; bazı

öğretmen yapımı testlerde, davranış ölçümünü, ilgilenilen özelliğin hiç ölçülemediği bir hâle bile sokabilir.

Bir tarih öğretmeni, öğrencisine, “1923 yılında ilan edilen Cumhuriyetimizin, bu yıl kaçınıcı kuruluş yıldönümünü kutluyoruz?” şeklinde bir soru soruyorsa, **geçerlik** konusunu hiç dikkate almıyor demektir. Çünkü bu soruya verilen cevap olarak sergilenen davranışın tarih konusuyla değil, sadece dört işlem bilgisiyle, daha ayrıntılı olarak ifade etmek gerekirse, matematik dersinde öğrenilen çıkarma işlemiyle ilgisi vardır. **Ölçmede geçerlik; ölçülmek istenen davranışın, başka davranışlarla karıştırılmadan ölçülmesidir.** Daha kısa ve anlaşılır ifadesiyle, geçerli ölçme, **ne ölçülmek isteniyorsa, sadece onu ölçbilmektir.**

Güvenirlik, **ölçme hatalarıyla** ilgili bir kavramdır. Pek çok ölçme kitabında, güvenilirlik “**art arda yapılan ölçümlerin tutarlılığı**” şeklinde tanımlanır. Bunun nedeni de “Yapılan ölçmeye hata karışmıyorsa, aynı varlığın aynı özelliğine ilişkin, aynı ölçme aracıyla yapılacak art arda ölçümlerin sonuçları da benzer olacaktır.” önermesidir. Ancak ölçmeye karışan hata türlerinden birisi de “sabit hata” olarak adlandırılan bir hata türüdür. Sabit hata, ölçülen özelliğin miktarından bağımsız olarak, her ölçmeye aynı **miktarda karışan** bir hatadır. Örneğin, eşit kollu terazinin bir kefesi delindiği için, deliğe yapılan bir kaynak, o teraziy-le yapılan bütün ölçümlerde, “o kaynağın kütlesi kadar” bir hataya neden olacaktır. Bu teraziyle 1 kilogramlık pirinç de tartılsa, 5 kilogram pirinç de tartılsa hata miktarı sabit kalacaktır. Bu durumda da, **sabit hata karışan ölçmelerde,** aynı ölçme aracıyla aynı varlığa ilişkin tekrarlı ölçümlerin sonuçları benzer olacaktır. Bu nedenle güvenilirliğin, “**ölçmelerin, hatadan arınık olması**” şeklinde tanımlanması daha doğrudur.

Geçerlik ve güvenilirlik kavramlarını daha anlaşılır kılmak için; bir ölçme aracıyla bir davranışın ölçülmesi, bir silahla bir hedefe ateş etmeye benzetilir.

Davranış ölçmede, hatasız ölçme yapılıyorsa, art arda yapılan ölçmelerin sonuçları aynı ya da birbirine çok yakın olacaktır. Bu, hem ölçme aracının **hatasız çalıştığının**, hem de ölçmeyi yapan kişinin **ölçme işini olması gerektiği gibi yaptığının** göstergesidir.

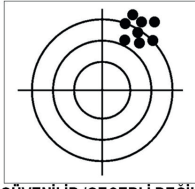
Şimdi de bir silahla, bir hedefin tam ortasına nişan alınarak art arda ateş edildiğini düşünelim. Eğer silahın nişangâh düzeninde ve diğer aksamında bir sorun yoksa (yani silah hatasız çalışıyorsa) ve ateş eden kişi de profesyonel bir nişancıysa, yani her defasında **aynı noktaya doğru şekilde nişan alıp ateş edebiliyorsa**, art arda yapılan atışlarda mermilerin hedefi vurduğu yerler aynı veya **birbirine çok yakın** olacaktır.

Bu örnek durumlar, **ölçmelerin** ve **atışların** hata içermediğinin, yani **güvenilir olduğunun** göstergesidir. Ancak **bu yeterli değildir**.

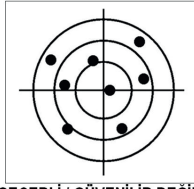
Doğru ölçüm yapan bir ölçme aracıyla, hatadan arınmış ölçümler yapmak, **ölçmek istediğimiz davranışı ölçüyor olduğumuz anlamına gelmez**. Matematik bilgisini ölçen tarih öğretmeni örneğinde olduğu gibi, tarih dersinde matematik bilgisini ölçmek, bu ölçme hatasız da olsa, eğitimin amaçlarının değerlendirilmesi açısından hiç de anlamlı değildir.

Benzer şekilde, belli bir hedefi vurmak için yapılan art arda atışlarda, mermiler; vurulmak istenen yerin dışında bir alanda toplansa bile, hedef vurulamadığı için o atış güvenilir dahi olsa gereksiz bir atıştır.

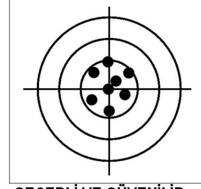
Ateş eden birisinin, **bir hedefi vurmak için** ateş etmesinde olduğu gibi, eğitimde de, ölçme **başta belirlenen amaçların gerçekleştirme derecesini görmek için** yapılır. Art arda yapılan atışlarda merminin **hedefin istenilen yerinde olması** o atışın geçerliğini gösterir. Davranış ölçümünde de, ölçülen davranışların, **ölçülmesi hedeflenen davranışlar** olması o ölçmenin geçerli bir ölçme olduğunu gösterir.



GÜVENİLİR / GEÇERLİ DEĞİL



GEÇERLİ / GÜVENİLİR DEĞİL



GEÇERLİ VE GÜVENİLİR

Geçerlik ve güvenilirlik konusunda şu çıkarımlarda bulunmak mümkündür:

Bir ölçümün güvenilirliğinin düşük olması geçerliği sınırlandırabilir ama yüksek güvenilirlik geçerliği garanti etmez.

Bir ölçümün güvenilirliği düşükse, geçerliği ne kadar uğraşılsa da belli bir değerin üzerine çıkamaz.