

Ölçme-Değerlendirmede Temel İlkeler

İlgi ŞEMİN (Değerlendirme Kurulu adına)

Dokuz Eylül Üniversitesi Tıp Fakültesi Fizyoloji Anabilim Dalı

Değerlendirme Kurulu Üyeleri:

P.Akan, B.Değirmenci, M.Gümüştekin, A.Karacı, B.Musal, İ.Şemin, C.Taşkıran, C. Terzi

ÖZET

Ölçme-değerlendirme eğitimin can alıcı noktalarından biridir. Bu özellik eğitimde öğrencilerin kazanımlarının yani çıktılarının beklenen düzeyde olup olmadığını göstermesinden ve öğrencilerin neyi ve nasıl öğreneceklerini etkilemesinden kaynaklanmaktadır. Ölçme ek olarak öğrenim planının değerlendirilmesini de sağlar. Bu yazıda eğitim parçalarından biri olan ölçme-değerlendirmenin temel özelliklerinden söz edilecektir. Bu amaçla önce ortak terminolojiyi sağlamak açısından kullanılan terimlerin tanımları yapılacak, daha sonra ölçme yöntemlerinden, bu yöntemlerin genel ilkeler içindeki yerlerinden ve analiz yöntemlerinden kısaca bahsedilecektir.

Anahtar sözcükler: Ölçme, değerlendirme

SUMMARY

Assessment evaluation is one of the most important parts of education. It shows the result of the curriculum whether it has reached the goals or not and also it drives the student's learning. Although assessment evaluation give chance to assess the curriculum plan. In this article this important part will be revised in basic terms. First the basic terminology, than the methods and their importance and the methods of analysis will be defined.

Key words: Assessment, evaluation

İlgi ŞEMİN

Dokuz Eylül Üniversitesi

Tıp Fakültesi

Fizyoloji Anabilim Dalı

Tel: 232 2595959 / 4451

Öğrenim planı yapılırken önce hedefler belirlenir, sonra bu hedeflere ulaşmak için hangi eğitim yöntemlerinin kullanılacağı saptanır ve sonuçta hedeflerin nasıl ölçüleceğine karar verilir. Ölçmede en temel ilke hem hedeflere hem öğrenme yöntemine uygundur.

Eğitimde ölçülecek hedefler bilişsel, psikomotor ve tutum-davranış olmak üzere üç başlıkta toplanır. Hedeflere yönelik içerik uygun araçlar kullanılarak ölçülmelidir ilkesi ile üç ana başlık için farklı araçların kullanılması gerekmektedir. Tercih edilecek olan ölçme aracının şu özelliklere sahip olması gerekir:

Geçerlilik: Bir ölçme aracının ölçmeyi amaçladığı özelliği, başka herhangi bir özellik ile karıştırmadan, doğru olarak ölçebilme derecesidir. Yani aracın amaca hizmet etmesidir.

Kapsam Geçerliliği: Ölçme konusu evreni yeterli ve dengeli olarak örnekleyen ve kapsadığı maddelerin her biri ölçmek istenen davranışı gerçekten ölçen araç, kapsam geçerliliğine sahiptir. Hazırlanan aracın ölçülen konuya ilişkin hedefleri, ağırlıkları oranında kapsamı gerekmektedir.

Uygunluk ya da ilgililik: Ölçme aracının ölçmek için düzenlendiği hedeflerle ilgili olmasıdır.

Güvenilirlik: Ölçmenin tutarlılığı ile ilgilidir. Bir testi değişik zamanlarda alan kişilerin, o testten aldıkları puanlar veya sıralamaları değişmiyorsa o test güvenilirirdir.

Ayrırdedicilik: Bilen ve bilmeyen öğrencileri ayırt edebilme gücüdür.

Nesnellik: Ölçme farklı kişiler tarafından yapıldığında da aynı sonucun alınmasıdır.

Uygulanabilirlik: Somut koşullara uygun olan ölçme aracının kullanılmasıdır.

Bilişsel Hedeflerin Ölçülmesi

Bilişsel alanda bilgi ve bilgiyi kullanma becerisi ölçülür. Bloom'un taksonomisine göre öğrenmede ilk basamak anımsamadır. Daha sonra kavrama, uygulama, analiz, sentez basamaklarından geçilir ve en sonunda, tam anlamıyla öğrenme olarak kabul edilen, edinilmiş bilgiyi farklı bir alanda kullanabilme basamağına ulaşılır. Anımsama'dan başlayarak yukarıdaki basamaklara çıkabilmek, gitikçe daha farklı bir düzeyde gelişkinlik gerektirir. Özellikle üniversite eğitiminde salt bilgi ölçen sınavlar yerine, daha üst taksonomik düzeyde gelişmeler hedeflendiği için, bilgiyi kullanma soruları tercih edilir. Düz bilgi anımsama "bu nedir?" soruları ile ölçülürken; kavrama, ilişki kurmayı gerektiren sorular ve grafik sunumları ile; analiz, sentez ve değerlendirme ise problem tarzı sorularla ölçülür.

Bilişsel hedeflerin ölçülmesinde ya çoktan seçmeli testler veya yazılı yanıt verilecek sorular kullanılır. Çoktan seçmeli test sorularının tipleri şunlardır:

- Tek doğru yanıtlı
- Çok doğru yanıtlı
- Doğru/yanlış
- Eşleştirme

Bu grubun geçerlilik ve güvenilirliği diğer yöntemlerden yüksektir. Bunun nedenleri şunlardır:

- Soru sayısı fazla olabildiğinden kapsam geçerliliğini sağlamak mümkündür.
- Yanıt tek olduğu için okuyan kişiye bağlı olarak ölçme sonucu değişmez, nesnelidir.
- Geçerli, uygun, bilişsel hedeflere yönelik soru hazırlama çok vakit alsa da sonuçta testin okunma süresi kısadır.
- Bu grupta seçenek sayısı arttıkça güvenilirlik de artar, dolayısıyla en çok tercih edileni beş seçeneikli çoktan seçmeli soru tipidir. Doğru-yanlış tarzı soruların ise olasılıkla bilinme yüzdesi büyük olduğu için güvenilirliği düşüktür.

Yazılı yanıt verilmesi istenen soru tipleri ise;

- Kısa yanıtlı sorular
 - Uzun yanıtlı sorular (essay)
 - Modifiye essay soruları
 - Problem çözme sorularıdır.
- Bu grupta soru sayısı kısıtlı olabildiği için kapsam geçerliliğini sağlamak güçtür.
 - Kağıdın değerlendirilmesinde nesnellik sağlamak da zordur, bu nedenle geçerlilik ve güvenilirlikleri düşüktür.
 - Bu grubun avantajı ise bilgiyi kullanma becerisinin ve tutum hedeflerinin değerlendirilmesinde sağladığı kolaylıktır.

Bilişsel hedeflerin ölçülmesi için hazırlanan araçta yukarıda sayılan yöntemler karma olarak da kullanılabilir.

Psikomotor ve tutum-davranış hedeflerinin ölçülmesi

Bu hedeflerin değerlendirilmesi gözlemle yapılacak ölçüm araçlarına gereksinim gösterir. Bu amaçla kullanılan yöntemler şunlardır:

- Uzun olgu
- Kısa olgu
- Klinik sözlü
- Objektif yapılandırılmış sınavlar
 - ❖ Nesnel örgün klinik sınav (Objective structured clinical examination=OSCE)
 - ❖ Nesnel örgün pratik sınav (Objective structured practical examination=OSPE)
 - ❖ Klinik nedenselleştirme sınavı (Clinically oriented reasoning exam= CORE)
 - ❖ Objektif yapılandırılmış uzun rapor yazımı (OSLER), vb
- Ödev hazırlama
- Proje hazırlama
- Karne ve profillemeye
- Gözlemsel yöntemler
 - ❖ Kendisini değerlendirme
 - ❖ Arkadaşları tarafından değerlendirilme
 - ❖ Eğitimci tarafından değerlendirilme

Beceri, tutum ve davranış ölçmede kullanılan araçların da ana ilkelere sayılan özelliklere sahip olma

oranı geçerlilik ve güvenilirliğini belirler. Bir değerlendirilmenin gözlemine dayanan ölçümlerde beklenen yanıtların somutlanması, skorlama için net tanımların yapılması nesnellığı artırır. Bu tip araçlarda çeşitli yöntemlerle standartları oluşturma ve buna bağlı olarak geçerlilik ve güvenilirliği arttırmaya yönelik çabalar sürdürülmektedir.

Bu yöntemler içinde nesnel örgün klinik sınavlar giderek yaygınlaşan biçimde kullanılmaktadır. Bu ölçme yönteminde hedeflere uygun sayıda istasyonlar kullanılarak öykü alma, fizik muayene, yorumlama, iletişim ve girişim becerileri ölçülebilir. İstasyonlarda hedefe uygun olarak gönüllüler, standardize hastalar, gerçek hastalar, maketler kullanılabilir. Her bir istasyonda öğrencilerin yapması istenen basamaklar tanımlanmıştır ve puanlama bu basamaklar üzerinden yapılır. Böylece birçok beceri çok sayıda istasyonla değerlendirilebildiğinden kapsam geçerliliğine sahiptir. Bütün öğrencilere aynı soruların sorulması adaleti sağlar. Puanlamada beklenen kriterler somut olarak belirtildiği ve bütün öğrencilere aynı kriterler uygulandığı için de nesnedir. İstasyondaki ölçme tamamlandığında öğrencilere geri bildirim verilerek eksiklerini fark etmesi ve öğrenmesine de olanak sağlanır.

Analiz

Kullanılan ölçme yöntemleri uygulamadan sonra analizlenerek geçerlilik, güvenilirlik, ayırtedicilik özellikleri açısından değerlendirilir ve sonraki uygulamalarda iyileştirme aşaması için veri elde edilir.

Test tipi sınavların analizi sırasında öğrencilerin aldığı puanlar, bu puanların dağılımı, başarılı ve başarısız öğrencilerin yanıtları ve benzeri bir çok ayrıntı testin amacına ne ölçüde ulaştığını gösteren ip uçları verir. Testten alınan en yüksek ve en düşük puan arasındaki fark, puanların dağılımı, testin güçlük derecesi gibi bilgiler test edilen alanda hedefe ne kadar ulaşıldığı, yani öğrencilerin konuyu ne kadar öğrendikleri hakkında da bilgi verir. Her bir test maddesinin analizi sonunda ise:

- 1) sorunun bağlı güçlük derecesi,
- 2) sorunun iyi öğrenci ile zayıf öğrenciyi birbirinden ayırt edebilme derecesi ve

- 3) sorunun çeldiricilerinin özellikle zayıf öğrencilere cazip gelip gelmediği hakkında bilgi verir. Test maddelerinin özellikleri ne kadar ideale yakınsa bütün test de o kadar geçerli ve güvenilir olacaktır.

Soru güçlüğü 0 ile +1 arasında bir değer olarak bulunur. Ölçülmek istenen özelliğe bağlı olarak, farklı güçlük derecelerinde sorular bir araya getirilerek testler düzenlenir.

Başlıca işlevi iyi öğrenci ile zayıf öğrenciyi ayırmak olan başarı testleri uygun hazırlandığında, iyi öğrencilerin soruya, zayıf öğrencilere göre daha yüksek oranda doğru yanıt vermesi beklenir. Eğer bir soruya tüm öğrenciler doğru veya tüm öğrenciler yanlış yanıt verdi ise, bu sorunun ayırt etme gücü hiç yoktur. Ayırt etme indeksi -1 ile +1 arasında değerler alır. İyi bir sorunun ayırt etme gücünün 0.30 üzerinde olması beklenir. Ayırt etme gücü 0.20 altında olan sorular düzeltilmelidir. Eğer değer 0'dan küçük ise, soru ters çalışmıştır. Bu tür soruların bilimsel olarak gözden geçirilmesi ve bir hata olup olmadığının araştırılması gerekmektedir.

Soru güçlüğü ve ayırt etme gücü sorunun yalnız doğru yanıtı ile ilgilidir. Ancak çeldiriciler de iyi soru düzenlenmesinde önemlidir. Eğer 5 seçenekli bir soruda 3 çeldiriciyi hiç bir öğrenci seçmemiş, tüm yanıtlar iki seçenekte toplanmış ise, bu soru "ikiden" seçmeli bir sorudur. İdeal olarak bir sorunun yanlış yanıtlarının çeldiriciler arasında eşit dağılımı beklenir. Çalışmayan çeldiricilerin üzerinde düşünülerek yeniden kullanılmadan önce düzeltilmesi yararlı olur.

Analiz sonuçları sorular hakkında matematiksel sonuçlar verse de, sorunun değerlendirilmesi bir bütündür. Maddenin yazımı, ifadesi, anlaşılabilirliği, kurallara uygunluğu, hedefe uygunluğu, üst düzey bilişsel yeti ölçmesi, testin organizasyonu, kapsam geçerliliği gibi faktörler bütün olarak iyi sorulardan oluşmuş iyi bir testi meydana getirebilecektir.

Ölçmenin asıl amacı, hedeflere ulaşma derecesine göre öğrencilerin o dönemde başarılı olup olmadığının değerlendirmesini sağlayacak verinin elde edilmesidir. Değerlendirmede başlıca iki yöntem kullanılmaktadır.

Bunlardan mutlak değerlendirme, önceden belirlenmiş bir standart ile kıyaslanarak karar verilmesi esasına dayanır. Bu nedenle "kritere bağlı değerlendirme" de denilmektedir. Bağıl değerlendirme ise birlikte olduğu gruba kıyasla yapılan özel hesaplamalar sonucunda bir karara varılır.

Dokuz Eylül Üniversitesi Tıp Fakültesi'nde bilgi test tipi sınavlar, bu sınavlarda entegre ve bütünleştirici sorularla kullanarak bilgiyi kullanma becerisi ölçülmektedir. Uygulanan testler analiz edilerek değerlendirilmektedir. Beceri ve tutum ise objektif yapılandırılmış olarak hazırlanan klinik sınavlar ve entegre uygulama sınavları ile ölçülmektedir. Elde edilen ölçme sonuçları ile mutlak değerlendirme yöntemi kullanılarak, ilk üç yılda sınıf geçme esasına dayanan değerlendirme yapılmaktadır.

Özet olarak, uygulanan eğitimin hedefleri ve öğrenme yöntemlerine uygun olarak çeşitli ölçme araçları ile hedeflenen bilgi, beceri ve tutum ölçülerek değerlendirilir. Sonuç değerlendirme çıktı hakkında bilgi verir ve buradan elde edilen veri, öğretim planının gözden geçirilmesinde kullanılır. Her geçen gün artan çabayla süren iyileştirme çalışmaları aracılığı ile her bir basamak için ideale ulaşılmaya çalışılmaktadır. İdealde buluşmak umuduyla...

KAYNAKLAR

1. Evaluation methods: A resource handbook. McMaster University program for educational development, program for faculty development and educating future physicians of Ontario (EFPO) project, 1995.
2. Whitman NA, Cockayne TW. Evaluating medical school courses: A user-centered handbook. University of Utah School of Medicine, 1984.
3. Mennin Stewart P, Kalishman Summers. Student assessment. Academic Medicine, September 1998;73:46-54.
4. Case S.M., Swanson D.B. Extended-matching items: A practical alternative to free-response questions. Teaching and learnin in medicine, 1993;5:107-115.
5. AMEE Guide No.9: Assessment of clinical competence using the objective structured long

examination record (OSLER). Medical Teacher, 1997;19:7-13.

6. AMEE Guide No.18: Standart setting in student assessment. Medical Teacher, Mar 2000;22:120-125.
7. Heard JK, Allen R, Cason GJ, Cantrell M, Tank PW. Practical issues in developing a program for the objective assessment of clinical skills. Medical Teacher, 1998;20:15-21.
8. Dugdale Alan E. The pathway MCQ: a method for teaching and testing deeper knowledge. Medical Teacher, 1998;20:250-253.
9. Halit Tekin. Eğitimde ölçme değerlendirme. Yargı yayınları, 1996.
10. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. Second ed. National Board of Medical Examiners, Philadelphia, 1998.
11. Fourth Biennial Conference of the International Association of Medical Science Educators. Advances in medical science education. Learning modes and teaching strategies, workshop notları. Washington, DC, USA, July 17-19, 1999.
12. Boulet JR, Ben-David MF, et al. High stakes examinations: What do we know about measurement? Academic Medicine Vol 73 No.10/October sup.1998;94-96.
13. Cusimano MD, Rothman A, Keystone J. Setting standarts for performance assessment. Academic Medicine Vol 73 No.10/October sup.1998; 112-113.
14. Carmines EG, Zeller RA. Reliability and validity assessment series: Quantitative applications in the social sciences. A Sage University paper, 17. Sage publications, Newbury Park London New Delhi, 1979.
15. Select and/or design assessments that elicit established outcomes. Available from: <http://www.ncrel.org/sdrs/ar>
16. Match assessments to the purposes for assessment. Available from: <http://www.ncrel.org/sdrs/ar>
17. Kehoe Jerard. Basic item analysis for multiple-choice tests. ERIC/AE Digest. Available from: <http://ericae.net/db/edo/ED398237.htm>