

Tıp eğitiminde ölçme ve değerlendirme

Doç. Dr. Duygu Anıl



1973 yılında Ankara'da doğdu. Hacettepe Üniversitesi'nde Eğitim Bilimleri Bölümünden mezun oldu (1995) Yüksek lisans ve doktora eğitimlerini aynı yerde tamamladı. 2010 yılında doçent oldu. Hacettepe Üniversitesi Diş Hekimliği Fakültesi ölçme değerlendirme merkezi kurulumu ve soru bankası oluşturma, MEB ulusal sınavlarında ölçme değerlendirme danışmanlığı, Milli Eğitim Bakanlığı Temel Eğitime Destek Projesi, YEĞİTEK PISA 2012 Türkiye Ulusal Raporu hazırlanma koordinatörlüğü görevlerinde bulundu. Araştırma alanları bilgisayar ortamında soru bankası oluşturma, ölçek ve test geliştirme, uluslararası ve ulusal sınavların psikometrik özelliklerinin incelenmesi, ölçme ve değerlendirme teknikleri, psikometridir. Halen Hacettepe Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı'nda öğretim üyesi olarak görev yapmaktadır.

Değişen ve gelişen dünyada birey davranışlarındaki değişiklikleri kalıcı hale getirebilmek ve dünyadaki gelişmelere ayak uydurabilen, çağın beklentilerine cevap verebilen, araştıran, sorgulayan ve kendini gerçekleştirmiş, özgüven duygusu gelişmiş bireyler yetiştirmek, ancak eğitimle mümkün olmaktadır.⁽³⁾ Günümüzde eğitim, istendik insan davranışlarının geliştirildiği bir sistem olarak görülmektedir. Bu sistem girdiler, süreç, çıktılar ve değerlendirme olmak üzere dört öğeden oluşmaktadır.⁽⁶⁾ Eğitimi bir sistem olarak ele aldığımızda, öncelikle öğrencilerin programa başlamadan önce ön koşul niteliğindeki bilişsel davranış, duyuşsal özellik ve devinışsel becerilerinin ne olduğuna bakılması ve buna göre öğrencilere kazandırılması istenen özellikleri kazandırıcı uygulamalara yer verilmesi planlanmalıdır. Bu sistemin girdileri arasında, öğrenci ve öğreticilerin özellikleri, eğitim programları ve bunun içinde hedef ve davranışlar önemli bir yer tutar. Sistem içerisinde yer alan süreçte, istendik davranışların gerçekleşmesini sağlayacak yani eğitim programlarındaki hedeflere ulaşabilmek için öğretim faaliyetlerinde bulunulur. Ve süreç sonunda öğrencilerde görülen öğrenme ürünleri eğitim sisteminin çıktılarını oluşturur. Sistem içerisinde yer alan bu öğelerdeki herhangi bir değişiklik veya aksaklık sistemin işleyişini tümüyle etkiler. Eğitim sürecinin başında istendik davranışların gerçekleşmesi sadece beklenti halindedir. Eğitim, bireyin davranışlarında kendi yaşıntısı yoluyla istendik

davranış değişikliği meydana getirme süreci olarak tanımlansa da eğitim süreci sonunda her zaman istendik davranışlar oluşmayabilir.⁽¹¹⁾ Bazen istenmeyen veya istendik fakat yeterli düzeyde kazanılmamış davranışlar olabilir. Eğitim sistemi içerisinde bu yetersizliklerin ve istenmeyen davranışların belirlenmesi, bu aksaklıkların kaynaklarının ortaya konması yani sistem içerisindeki işleyişin izlenerek, kontrol edilmesi ve gelişiminin sağlanması ancak, ölçme ve değerlendirme ile mümkün olmaktadır.

Ölçme ve değerlendirme kavramları birbirini tamamlayan fakat ayrı anlamlar taşıyan kavramlardır. Ölçme; *herhangi bir niteliğin gözlemek ve gözlem sonucunu sayı ya da sembollerle ifade etmektir.*⁽¹⁷⁾ Değerlendirme kavramı ise, ölçmeyi de içine alan daha genel bir kavramdır. Ölçme bize sadece bir niteliğin niceliğine ilişkin bilgi verir, bu bilginin yeterli olup olmadığı hakkındaki bilgi ise değerlendirme sorunudur. Değerlendirme; ölçme sonuçlarının bir ölçütle karşılaştırılıp, ölçülen nitelik hakkında bir karar verme süreci olarak tanımlanır. Değerlendirme, ölçme sonucu, ölçüt ve karar öğelerini içermektedir. Bugün ülkemizde, ölçme ve değerlendirmenin yeterli düzeyde bilinmemesi, güvenilirlik ve geçerlikten uzak ölçme araçlarıyla ölçme yapılarak öğrenciler hakkında isabetsiz kararlar verilmesine sebep olmaktadır.

Eğitim sürecinde değerlendirme genellikle iki amaca yönelik olarak yapılır: 1) öğrencilerin başarısını değerlendirerek

bir dersin hangi öğrenciler tarafından tekrar edilmesi gerektiğine karar vermek, 2) eğitim programlarının etkililiği hakkında yargıda bulunmak ve programdaki aksaklıkların, programın hangi öğe ya da öğelerinden kaynaklandığını belirleyerek gerekli düzeltmelerin yapılmasına olanak sağlamak. Bunlardan birincisinde öğrenci değerlendirilirken, ikincisinde ise eğitim programı değerlendirilmektedir. Değerlendirme sadece, programdaki öğrencilerin başarı düzeylerini belirlemek için değil aynı zamanda hizmet içi eğitimle personeli yetiştirmek ve eğitimcilerin toplumla nasıl iletişim kurabileceği hakkında kararlar verebilmek içinde de kullanılır.⁽¹³⁾

Tıp eğitiminde olduğu gibi değerlendirmeye genel anlamda iki farklı açıdan yaklaşılmaktadır. Bunlar;

1. Düzey belirleme değerlendirmesi (Summative Evaluation)
2. İzleme değerlendirmesi (Formative Evaluation)

Düzey belirleme değerlendirmesi öğrenme düzeyini belirlemek, dönem sonlarında; not vermek, sertifika ve programın etkililiği üzerinde araştırma yapmak, öğrenci başarısını ve öğrenme düzeyini belirlemek amacıyla yapılmaktadır. Düzey belirleme amacıyla yapılan değerlendirmenin en önemli özelliği; öğretim ya da öğrenme bittikten sonra, öğretim ya da öğrenmenin etkisi konusunda öğretmen ve öğrenci hakkında yargılara ulaşılmasıdır. İzle-



me değerlendirmesi ise, programın değerlendirilmesi, öğretimin etkililiğinin ve öğrenme güçlüklerinin saptanması gibi amaçlarla kullanılmaktadır. Elde edilen verilerle not verilmez. Çalıştığı konu programın kendisidir. Öğrencilerin öğrenme düzeyleri hakkında bir yargıda bulunulmaz. İzleme değerlendirmesinde amaç; süreci geliştirmek, öğrencilerin öğrenme eksikliklerini belirlemek ve daha fazla öğrenmelerinin gerektiğinin farkına varmalarını sağlamaktır. İzleme değerlendirmesi amacımız doğrultusunda herhangi bir zamanda kullanılırken; düzey belirleme değerlendirmesi yapabilmek için belirli dönemlerin beklenmesi gerekmektedir.⁽⁷⁻¹⁴⁻⁴⁾ İyi bir değerlendirme sisteminde birden çok yöntem kullanılır.

Tıpta mesleki yetkinlik "iletişim, bilgi, teknik beceriler, klinik akıl yürütme, duygular, değerler; alışkanlık halinde, sağduyulu olarak bireyler ve toplum yararına kullanmak" olarak tarif edilirken, mesleki yetkinliğin klinik beceriler, bilimsel bilgi ve ahlaki gelişim zemininde oluştuğu belirtilmektedir.⁽⁹⁾ Mesleki yetkinlik öğrenme alanları taksonomisinden hareketle, öğrenenin bilmesi, öğrendiğini kullanması şeklinde de özetlenebilir. Profesyonel işlevlere yönelik üç temel alanda öğrenci performansı değerlendirilir: İletişim becerileri (*tutum, davranış alanı*), bilgi ve entelektüel beceriler (*bilişsel alan*) ve pratik beceriler (*motor*

alan). İnsan davranışı üzerinde pratik, bilişsel ya da iletişimsel keskin ayrımlar yapmak çok zor olduğu için; böyle bir sınıflandırmanın yegâne amacı, ancak öğrenme sürecinin analizine izin vermek ve öğretmenlere eğitime yönelik karar verme konusunda yardımcı olmaktır. İletişim Becerileri; hekim ve hasta arasındaki ilişki olarak tanımlanır. Hastaları uygun şekilde bilgilendirme, yönlendirme ve uyarılarda bulunma; hastaya kaygılarını ifade etme fırsatı tanıma; hastaya güven verme; hastadan gelen ancak sözlü olmayan mesajları fark edebilme; hastayı dinleme ve yanıtlama bu becerilere örnek olarak verilebilir. Bilgi ve Entelektüel Beceriler; Bloom sınıflamasında uyarılan Buckwalter-McGuie sınıflandırmasına göre *bilgilerin hatırlanması, verilerin yorumlanması ve problem çözme olarak* üç farklı seviyede ele alınabilir. Pratik beceriler ise; hekimlerin, hastayı entübe etmek ya da bir sıvı tedavisine başlamak gibi mesleki hayatlarında rutin olarak yaptıkları işlevlerde sahip oldukları yeterliliklerdir. Günümüzde tıp eğitimi, büyük bir değişim içindedir. Tıp öğrencilerinin toplumun ihtiyaçlarına cevap verecek düzeyde yetiştirilmesi amaçlanmış; bu doğrultuda tıp fakültelerinde yeni birimler kurulmaya ve müfredatta değişiklikler yapılmaya başlanmıştır. Son yıllarda dünyadaki tüm tıp fakültelerinde öğretim alanında "beceri eğitimi" yeni bir bölüm olarak ortaya çıkmıştır. ⁽¹⁾

Ölçme araçları, ölçülen büyüklüğün gözlenip sayılarla ifade edilmesini; böylelikle gözlemin daha duyarlı yapılarak objektif ölçümler elde edilmesini ve ölçümlerin standartlaştırılmasını sağlar. Fakat ölçme sonuçlarına yine de istenmeyen değişkenler, yani hatalar karışabilir. Ölçme araçlarımızın hatalardan arınık ölçme yapabilmesi ölçme araçlarının güvenilir olması ile doğrudan ilgilidir.

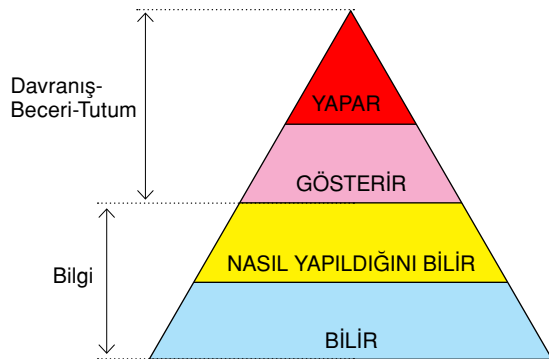


Tıp eğitiminde ölçme ve değerlendirmede kullanılan yöntemler

Ölçmenin doğasında mesleksi yeterliğe bir değer biçme süreci yatar. Uzman hekimden beklenen genel yeterlikler ise şöyle tanımlanmıştır;⁽¹⁾

- Hasta bakımı
- Tıbbi bilgi
- Uygulamaya dayalı öğrenme ve gelişim
- Kişilerarası ve iletişim becerileri
- Profesyonellik
- Sistemlere dayalı uygulama.

Son on yıl içerisinde tıp fakülteleri eğitim programlarına giren “iyi hekimlik” değerleri ile tanımlanan hasta, hasta yakını hekim arasındaki iletişim becerileri uygulamalarını, kanıt dayalı tıp uygulamalarını, öğrencilerde erken dönemde klinik ziyaretlerle farkındalık sağlanmasını, hekimlere farklı bakış açıları kazandıran tıpta insan bilimleri etkinliklerini, etik ve profesyonel değerlerin tartışıldığı etkinlikleri mezuniyet öncesi



Şekil: Miller'in Yeterlikler Piramidi

tıp eğitiminin önemli değişim - gelişim aşamaları olarak değerlendirmek gerekmektedir. Tıp eğitiminde küçük grup etkinliklerinin yaygınlaşması, probleme dayalı, göreve dayalı, çıktı temelli öğrenme stratejilerinin işe koşulması öğrenen merkezli uygulamaların yaygınlaşmasına olanak sağlamaktadır. Tıp eğitiminde bilgisayara, webe dayalı eğitim, eğitim materyallerinin kullanımını ve etkinliğini artırmaktadır. Uzaktan eğitim, bilgi paylaşımı için ulusal – uluslararası ortak çalışmaların altyapısını oluşturmaktadır.⁽¹⁵⁾

Tıp fakültelerinin ilk yıllarında öğrencilere kazandırılmaya çalışılan “iyi hekimlik” değerlerinin öğrencilerin kliniğe geçtiklerinde karşılığını bulabilmesi için temel bilimler ile klinik bilimlerin bilgi, beceri hatta tutum bakımından dikey entegre edilmesi gerekmektedir. Fakültenin ilk yıllarından başlanarak öğrencilerin birlikte çalışma, sorun çözme, grup içerisinde sorumluluk üstlenme, ekip çalışmasına uyum sağlama becerilerinin geliştirileceği eğitim ortamları sağlanmalı, öğrencilerin kendilerini, akranlarını değerlendirmeleri için çaba gösterilmelidir. Öğretim üyelerinde de öğrencilere geribildirim verme ve öğrencilerden geribildirim alma motivasyonu yaratılmalıdır.⁽¹⁵⁾

Tıp eğitiminde öğrencilerin klinik yeterlik düzeylerini belirlemek için Miller 1990 yılında, dört düzeyde tanımladığı bir yeterlikler piramidi önermiştir.⁽¹²⁾ (Şekil)

Öğrencilerin klinik yeterlik düzeylerini belirlemek ile ilgili olarak Miller'in tanımladığı yeterlikler piramidinin en alt basamağında bilginin değerlendirilmesi (temel bilgi ve kavramların bilinmesi),

üstte yeterliğin değerlendirilmesi (normal-anormal yapının, mekanizma ve fonksiyonların nasıl olduğunun bilinmesi ve bilinenlerin yeni durumlara uyarlanması), daha üstte performansın değerlendirilmesi (bilgi, beceri ve tutumların eğitim ortamında ve gözlem altında uygulayarak gösterilmesi) ve en üstte davranışın değerlendirilmesi (mesleğin gerçek yaşam koşullarında uygulanması) yer almaktadır. Miller piramidinin üst kısımları, yani öğrencinin bir duruma ait becerileri yapıp göstermesi performans değerlendirmeyi açığa çıkarmaktadır.⁽¹²⁾

Miller piramidinin tıp eğitiminde öğrencilerin klinik yeterlik düzeylerini belirlemek için tanımladığı dört düzey belirlenmesine göre öğrencilerin değerlendirilmesinde kullanılacak ölçme araçlarında “bilir”, “gösterir”, “yapar” boyutunda güvenilir ve geçerli öğrenci başarısını ölçebilen farklı ölçme araçlarının kullanılması önem arz etmektedir. Güvenirlik, bir ölçme aracının hatalardan arınık olarak ölçme yapabilme yeterliği olarak tanımlanır. Hazırladığımız ölçme aracının hatasını farklı güvenirlilik hesaplama yöntemleri ile belirleyebiliriz (test tekrar test yöntemi, iki yarı güvenirliliği, içtutarlık anlamında güvenirlilik hesaplama yöntemleri vb.) Geçerlik ise genel anlamda, bir testin, sadece o testle ölçülmek istenen değişkeni ölçmesi, başka bir değişkenle karıştırmaması olarak tanımlanır.⁽¹⁷⁾ Tıp eğitiminin beceri boyutunda uygulamalı bir eğitim olduğu düşünüldüğünde sadece bilginin değerlendirilmesinin yeterli olmayacağı açıkça görülmektedir. Bu nedenle; bilginin değerlendirilmesinde yazılı yoklamalar, çoktan seçmeli testler, sözlü yoklamalar, kısa cevaplı testler ölçme araçları olarak kullanılması gerekirken; bilginin gösterilmesinin

değerlendirildiği durumlarda uzun ve kısa vaka durumlarının ölçüldüğü sözlü sınavlar ya da yapılandırılmış objektif klinik sınavlar (OSCE) uygulanabilmektedir. Davranışın değerlendirilmesinde ise becerileri doğrudan gözlem, 360 derece değerlendirme, gizli standart hasta görüşmeleri, simülasyonlar ve modeller, asistan karnesi (logbook) ve gelişim dosyaları kullanılmaktadır.⁽²⁻⁵⁾ Tıp eğitiminde öğrenci başarılarını belirlemek için kullanılan bu ölçme araçları hakkında bilgiler Miller'in yeterlikler piramidinin tanımladığı dört düzey boyutu dikkate alınarak aşağıda açıklanmıştır.

Yazılı yoklamalar

Soruların yazılı olarak verildiği ve öğrencilerin yazılı olarak cevaplandırmalarının istendiği sınavlara yazılı yoklamalar adı verilir. Yazılı yoklamalarda, cevapların yazılı olarak verilmesi zorunluluğu, cevaplayıcı bağımsızlığı, puanlamanın sübjektif olması yazılı yoklamaların niteliklerini doğrudan etkileyen etkenlerdir. Az soruda soruyu içermesi ve hazırlanmasındaki kolaylıktan ötürü eğitimciler tarafından tercih edilen ölçme araçları arasında yer almaktadır. Özellikle üst düzey davranışların yoklanmak istendiği (örneğin, analiz veya sentez basamağında davranışların ölçülmesinde) veya sürecin adımlarının görülmesini gerektiren durumlarda en etkili kullanılması gereken ölçme araçlarından biridir.⁽⁵⁾ Şans başarısının olmaması yine bu aracın üstünlükleri arasında yer alırken; bu sınavlarda az sayıda soru sorulabilmesi kapsam geçerliğinin düşük olmasına ve puanlamanın sübjektif olması da güvenirliliği olumsuz etkileyen sınırlılıklarıdır.

Çoktan seçmeli testler

Bir sorunun cevabını verilen seçenekler arasından seçtiren sorulardan oluşmuş testlere çoktan seçmeli testler adı verilir. Çoktan seçmeli testler, madde kökü, seçenekler(cevap şıkları), anahtarlanmış cevap (doğru cevap) ve çeldiriciler olmak üzere madde formunu oluşturan dört öğeden oluşmaktadır. Bu testler geleneksel kâğıt testleri olarak uygulanabildiği gibi bireye uyarlanmış testler olarak bilgisayar ortamında da uygulanabilir. Tek uygulama ile çok sayıda sorunun sorulabilmesi, puanlamanın objektif olması, hazırlama süresinin uzun fakat uygulamasının kolay olması bu sınavların üstünlükleridir. Şans başarısının olması üst düzey kazanımların ölçülmesinin zor olması da bu testlerin sınırlılıklarıdır. Bu testler madde analizinin yapılmasına imkân verdiğinden, istenen güçlükte madde seçilerek istenilen güçlükte testler oluşturulmasına imkân tanır. Bu nedenle istenilen nitelikte geçerli ve güvenirliliği yüksek sınavlar hazırlamak mümkündür.⁽⁵⁾ Öğrencilerin yalnızca

ezbere dayalı bilgilerini değil bilginin bir bütün olarak anlaşılıp anlaşılmadığını ölçen bir değerlendirme yöntemidir. Her soru 4 veya 5 seçenek içerir. Sorular genelde hastaya ait bir bilgi, klinik bulguları veya grafik olarak bilgileri içerir. Resimleri ihtiva eden ayrı bir kitapçık da kullanılabilir.

Bu testler geleneksel kâğıt testleri olarak uygulanabildiği gibi bireye uyarlanmış testler olarak bilgisayar ortamında da uygulanabilir. Bilgisayar ortamında uygulanabilen bireye uyarlanmış test, klasik testlerden farklı olarak katılımcıların bilgisayarlarla yönetildiği ölçme aracıdır.⁽⁶⁾ Bilgisayar ortamında uygulanabilen bireye uyarlanmış testler, sadece Klasik test kuramına dayalı (KTT) değil Madde Tepki Kuramına (MTK) dayalı test puanlama yöntemleri ile bireylerin bir ya da birden çok maddeye cevaplarına dayanan yetenek düzeyi tahminlerine imkân tanımaktadır.

Sözlü sınavlar

Sözlü sınavların en belirgin özelliği soruların sözlü olarak sorulması ve cevapların sözlü olarak konuşma dilinde verilmesidir. Öğrencilerin sözlü anlatım becerilerini geliştirmek ve bu becerilere ne kadar sahip olduklarını ölçmede sıkça kullanılan sınav türüdür. Az sayıda soru sorulabilmesi nedeniyle kapsam geçerliğinin düşük olması ve puanlamanın sübjektif olması bu sınavların sınırlı yönleridir. Tıp eğitiminde teşhis, klinik bulguların yorumu, tedavi planları vb. işlemlerde ve uygulamalarda sözlülerin hasta başında uyguladıkları değerlendirme türüdür. Etkili bir şekilde hazırlanmış bir sözlü sınav 5-10 dakika kadar sürmektedir. Bu sınavlar klinik karar verme becerileri ve tıbbi bilginin kullanımı ve uygulamasının gerçekleşmesini sağlar. Genellikle uzmanlardan oluşan bir ekip gerçek hasta durumlarından çeşitli senaryolar hazırlar ve öğrencilere sunulur. Ardından soru havuzundan belli sayıda soru rastgele seçilerek sorulur. Cevaplar en azından iki bağımsız değerlendirici tarafından değerlendirilir. Bu sınavlar klinik karar verme ve tıbbi bilginin kullanıldığı ya da uygulandığı gerçek hasta ile yapılan sınavlardır. Sözlü sınavların, öğrencilerin problem çözme becerileri ve klinik bilgilerini değerlendirmek için uygun bir yöntem olduğu belirtilmektedir. Uzmanlar adaylara yaratacakları hasta ve hastalık durumlarını nasıl tasarlayacakları, soruların nasıl sorulması ve hazırlanması gerektiği ve adayın yanıtlarının puanlanması ve değerlendirilmesi açısından eğitilmelidir. Standardizasyonun sağlanmaması, az sayıda soru sorulması, puanlamanın sübjektif olması, uzman yanlılığı, yoğun kaynak ve deneyimli uzman gerektirmesi sınırlılıkları arasındadır.

Objektif Yapılandırılmış Klinik Sınavlar (OSCE)

OSCE (Objective Structured Clinical Examination) sınavı üst düzey düşünme becerilerini ölçmede birden fazla istasyondan oluşan ve her bir istasyonda farklı klinik becerilerin önceden belirlenen kriterler doğrultusunda puanlayıcılar tarafından değerlendirildiği performans sınavıdır. Bu sınav, 1977 yılında İskoçya Dundee Üniversitesi'nde Ronald Harden tarafından cerrahi bölümde yapılan sınavlar için kullanılmıştır.⁽¹⁰⁾ Her bir istasyonun 10-15 dk. aralığında sürdüğü bir ya da birden fazla değerlendirme aracının kullanılabilirdiği adayların farklı durumlarla karşı karşıya bırakıldığı farklı istasyonlardan oluşan sınavlardır. İstasyonlar arasında adaylar hasta notlarını tamamlayabilir ya da bir önceki istasyona yönelik özetler hazırlayabilir. Standart hastalar sıklıkla kullanılan değerlendirme araçlarından biridir ama bunun yanı sıra sınav verilerin yorumlanması, klinik durumların kullanılması, mankenlerle klinik senaryoların kullanılması, teknik becerilerin değerlendirilmesi gibi değerlendirme durumları da söz konusu olabilmektedir. OSCE bir değerlendirme yöntemi değil bir sınav organizasyon biçimidir. Değerlendirmenin temeli standartlara uygun gösterilen performansın gözlenmesidir. OSCE formatı standart anlamda fiziksel muayene ve hikâye alma, hasta ve yakınlarıyla olan iletişim becerileri, bulguların teşhisi, tedavi programı uygulaması, hasta notlarına dayanan klinik yargı, bilginin pratiğe dökülmesi gibi durumlarda sıklıkla kullanılmaktadır. Son yıllarda tıp fakülteleri genelinde kullanımı oldukça yaygınlaşmıştır. 16-18 aralığındaki bir istasyondan kurulan OSCE performansın değerlendirilmesinde güvenilir bir araç olabilir. Kritik davranışların skorlara yönelik önerilen bir puan ağırlığı bulunmamaktadır. Osce özel klinik beceri ve faaliyetlerin ölçülmesinde çok yararlı olmasına rağmen, oluşturulması oldukça zor, zaman alan ve ekonomik olmayan sınavlardır. Özel olarak oluşturulmuş hasta odalarına, kayıtların alındığı sistemlere, hasta dışı faaliyetlerin yapılabileceği mekânlara ihtiyaç vardır.

360 derece değerlendirme

Birçok tabakadan insanın değerlendirmesini içeren değerlendirme türüdür. Bu değerlendirmeye astlar, üstler, çiftler, hastalar, aileler vb. ranji oldukça geniş bir topluluk katılmaktadır. Anketlerle bireylerin birçok konuda ki (takım çalışması, iletişim, beceriler, karar verme, kontrol) performanslarına yönelik bilgilerin toplandığı süreçtir. Toplanan değerlendirmeler değerlendirilen bireye dönüt olarak verir. Toplamaya değil oluşum ve gelişmeye yönelik bir



değerlendirme yapılmak istendiğinde, bu değerlendirme türü daha açık ve doğru sonuçlar ve dönütler ortaya koyar. 360 derece değerlendirme iletişim becerileri, profesyonel tavırlar, hasta bakımı ve sağlık sisteminin gerektirdiği uygulamalarda kullanılabilir. Herkese uygun olan anketlerin hazırlanması ve Büyük bir çoğunluktan toplanacak olan verilerin derlenmesi ve çok doğru bir şekilde rapor edilmesinin zorluğu bu aracın sınırlılıkları arasında yer alır.

Canlı veya kayıta alınmış performansın değerlendirilmesi

Global değerlendirme değerlendirici uzmanlık öğrencisinin hasta bakımı, tıbbi bilgisi ve iletişim becerileri gibi genel yeteneklerini değerlendirir ve değerlendirme uzun süre zarfında elde edilen izlenim ve diğer öğretim üyeleri, uzmanlık öğrencileri ve hastalar ile çalışmalarının değerlendirilmesi ile toplanılan bilgiler ışığında örneğin klinik rotasyonun sonunda yapılır. Değerlendirme 1'den 5'e kadar derecelendirilerek yapılabilirdiği gibi değerlendirmesine göre yapılabileceği gibi 1=çok yeterli 2=yeterli 3= yetersiz şeklinde de yapılabilir. Değerlendiriciler iyi eğitilmediği takdirde sonuçlar çok sübjektif olmaktadır. Deneysel doktorların deneyimsizlere nazaran daha güvenilir ve yetenek düzeylerini iyi ayırt eden değerlendirmeler yaptığı araştırmalarca ortaya konmuştur. Aynı zamanda aynı bireyin farklı durumdaki değerlendirilmeleri de çok farklı çıkmıştır. Tekrarlanabilirliği bilginin derecelen-

dirilmesi yanında kolay görünürken, hasta bakımı ve kişiler arası iletişim becerileri yanında daha zordur. Çok kolay bir biçimde oluşturulabilir ve çabuk tamamlanabilir. Ancak performansların doğrudan gözlenmesi için zamana ihtiyaç vardır ya da değerlendirme için uzmanlarla iletişime geçilmesi gerekir. Bulguların tekrarlı olarak kullanılabilmesini geliştirmek için değerlendirilen bireylerin eğitilmesi şarttır.

Kontrol listeleri

Kontrol listeleri gerekli ve özel davranışlardan, aktivasyonlardan ya da karmaşık ve zor bir sürecin ya da yeteneğin adımlarından oluşur. Davranışın, adımın tamamlanmış olduğunu (tamamen, bir kısmı, hiçbir adımı) ya da doğruluğunu (doğru, kısmen doğru, doğru değil) gösteren ve tipik yanıtların bulunduğu formlardır. Bu listeler tüm performansın yeterliği hakkında yargıya ulaşmak için değil performans hakkında bilgi verir. Listelerin performansın yeterliği açısından kullanılması açısından standartların oluşturulmasına ihtiyaç vardır. Kritik bir davranışı oluşturan her tür yetenek ve yetenek bileşeninin değerlendirilmesinde oldukça kullanışlı olan bir değerlendirme aracıdır. Bu listelerin dokümanları; hasta bakım becerileri(hikaye alma, fiziksel muayene, prosedürel beceriler...), iletişim becerilerinin değerlendirilmesinde de kullanılmaktadır. Kontrol listeleri aynı zamanda PDÖ uygulamalarında(kanıt dayalı tıp alanında) öz değerlendirme amacıyla da kullanılmaktadır. Kontrol

listeleri bir görevi yerine getirirken basamakların detaylı bir biçimde belirtilmesi ve bu basamaklara ait değişimlerin kolay algılanması açısından en doğru dönütü sağlayabilmektedir.

Gelişim dosyaları (portfolyo)

Gelişim dosyaları adayların öğrenme planları ile bağlantılı çalışma ve ürünlerinin toplandığı bir dosyadır. Gelişim dosyaları; yazılı dokümanlar, video ve kaset kayıtları, fotoğraflar, bilginin yer aldığı diğer formlar... gibi her türlü dokümanı içerebilmektedir. Gelişim dosyasının oluşumundaki en önemli bölüm ne öğrenildiğinin yansıtılması gereğidir. Tıp eğitiminde bir gelişim dosyasında klinik deneyimler, bir tedavi yönteminin araştırılmasına yönelik kayıtlar, nicel bir proje planı ve araştırma sonuçları, etik ayrılıkların yer aldığı bir tartışma ve nasıl ele alındığı, hasta bakımı ile alakalı bir bilgisayar programı, hastalara yönelik hazırlanmış bir öneri reçetesi gibi kayıtlar bulunabilir. Hem izleme (formative) hem de karar verme (summative) amacıyla kullanılabilir. Süreç gerektiren yeteneklerin ölçülmesinde ya da daha üst düzeydeki yeteneklerin değerlendirilmesi için yararlı bir değerlendirme yaklaşımı olarak görülebilir. Öğretim deneyimleri, sabah raporları, hasta viziteleri, bireysel çalışmalar, araştırma projeleri, gelişim dosyası kullanılarak değerlendirilebilen aktivitelerdir. The Royal College of Physicians and Surgeons of Canada in the Maintenance of Competence Program (MOCOMPS)

asistanların değerlendirilmesine yönelik olarak hazırlanmış "PCdiary" isminde internet tabanlı bir gelişim dosyası geliştirmiştir. Gelişim dosyasının kapsamına yönelik ölçüt ve standartlar bakımından birlik sağlandığı takdirde reproduksiyon değerlendirmeler söz konusu olabilir.

Simülasyonlar ve modeller

Simülasyonlar klinik bir problemle birebir örtüşmemekle birlikte klinik performansın gerçeğine oldukça yakın bir biçimde değerlendirilmesine imkân sağlar. Simülasyonların anahtar tutumları gerçeğine benzer bir aşamallığı içinde barındırır; çok az yada hiç olmayan ipuçlarıyla klinik bir problemin sebebinin araştırılmasına izin verir; değerlendirilenlere yanıtları yaşayarak öğrenme fırsatını sunar. Dolayısıyla yanlış bir uygulamanın geri dönüşü vardır ve tekrar alıştırmayı mümkün kılar. Klinik durum simülasyonları (CCX), rol yapma (standart hastalar, klinik takım durumları), anatomik modeller; patolojik bulgulara erişmek amacıyla kullanılan vücut bölgelerinin modelleri ya da sayılanların bir kombinasyonu bu kısım içinde yer almaktadır. Mankenler ve modeller insan dokusuna çok benzeyen vinil ve plastikten yapılmış olup aynı zamanda elektronik özel devrelerle değerlendirilen hareketlerine göre tepkiler vermesi de sağlanmaktadır. Teknoloji ve bilgisayar destekli görsel simülasyon ve çevrelerin (VR) kullanılması ve bazen de bunların manken ve anatomik modellerle desteklenmesi değerlendirilene tıpkı gerçek bir insan üzerinde çalışıyor olma hissini yaşatabilmektedir. Bunlarla yapılan değerlendirmeler becerilerin ve karmaşık, zor klinik görevlerin diğer değerlendirme araçlarından daha net bir yordayıcı değerlendirme yapmasına olanak vermektedir. Araştırmalar bu tip araçların gerçeğe ne kadar benzer nitelikte tasarlanırsa yapı geçerliliğinin de o kadar yüksek olduğunu göstermiştir. Her bir simülasyon ve modeldeki uygulamalar için önceden uzmanlarca belirlenmiş puanlar bulunmaktadır. Değerlendirilenin puanı ve performansı tüm simülasyon ve modellerdeki puanlarının ve uygulamalarının bir kombinasyonu şeklindedir. OSCE sınavlarının içinde yer alan bu tip araçlarla ölçülen durumların güvenilirliği istasyon sayısı arttıkça artmaktadır.

Standart hasta sınavları

Standart hasta sınavları önceden belirlenmiş senaryolara hazırlanan standart hastalarla öğrencilerin 10-12 dakikalık süreçlerde yaptığı görüşmelerden oluşmaktadır. Standart hastalar önceden eğitilmiş hasta rolü yapan kimselerdir. Gerçek hasta durumlarını ortaya koyan standart hastalarla yapılan görüşmeler kontrol listeleri ile değerlendirilir. Ortaya konan her durumun kontrol

listeleri bulunmaktadır (bahsedilen bu durum; kimi zaman bir hasta hikâyesi alma, fiziksel muayene, tedavi planının uygulanması ya da verilen öneriler ışığında şekillenebilir. Kontrol listeleri de durumlara has davranışları kapsamaktadır). Kontrol listeleri ile performanstaki uygunluk, doğruluk, tamlik derecelendirilir. Performans ölçütleri ileri düzeyde ayarlanmıştır. Bu sınavlarda standart hasta olan kimsenin değerlendirmeleri de kullanılır. Standart hasta kullanımı OSCE sınavlarında da yer almaktadır. Standart hasta sınavları bireysel istasyonlarda güvenilir skorlar ortaya koyar ve toplam performansta geçme kalma kararları için kullanışlıdır. Gözlemciler arası güvenilir önemli olduğundan gerek doktor, gerek standart hasta gerek herhangi bir gözlemci değerlendirme konusunda eğitilmelidir. Bu sınavların kapsam ve uyum geçerliliğine yönelik yapılan araştırmalarda klinik becerilerin puanları, yazılı sınavların puanları ile bu sınavlardan alınan puanlar arasında oldukça yüksek ilişki tespit edilmiştir.

Sonuç

Eğitim ve psikolojinin de içinde yer aldığı davranış bilimlerinin inceleme alanına giren değişkenlerin çoğu doğrudan gözlenemediği için fiziksel nitelikleri bilinmeyen ve bu nedenle fiziksel boyutları tanımlanamayan değişkenlerdir. Fiziksel bilimlerdeki ölçmelerde ölçülen niteliği, ölçmenin yapıldığı süre içinde değişmez veya kontrol altında tutmak ve aynı niteliğin ölçülmesinde birbirinden bağımsız ölçme sonuçları elde etmek mümkündür. Fakat eğitim ve psikolojideki ölçmelerde, ölçme konusu olan nitelikler insanın özellikleri veya davranışları olduğundan, bir test aynı bireye birden çok defa uygulanamaz, uygulanırsa bile önceki uygulamalardan olacak hatırlamalar sebebiyle ölçmeler birbirinden bağımsız olmaz. Bu nedenle, sosyal bilimlerde kuramsal bir bağıntının doğruluğunun yoklanması, ancak, gözlem ve ölçümlerin duyarlılığı oranında mümkün olmaktadır.

Doğrudan gözlenemeyen değişkenleri ölçmede; hatayı en aza indirmek, ölçmek istediğimiz niteliğe başka değişkenlerin karıştırmamak ve daha duyarlı ölçme sonuçları elde etmek için diğer bilim dallarında olduğu gibi eğitimdeki ölçmelerde de araç kullanılmaktadır. Ölçme araçları, ölçülen büyüklüğün gözlenip sayılarla ifade edilmesini; böylelikle gözlemin daha duyarlı yapılarak objektif ölçümler elde edilmesini ve ölçümlerin standartlaştırılmasını sağlar. Fakat ölçme sonuçlarına yine de istenmeyen değişkenler, yani hatalar karışabilir. Ölçme araçlarımızın hatalardan arınık ölçme yapabilmesi ölçme araçlarının güvenilir olması ile doğrudan ilgilidir.

Sadece bilişsel değil uygulama ağırlıklı eğitim alanlarından birisi olan tıp eğitiminde beceri geliştirilmenin büyük önem taşıması nedeniyle kullanılacak ölçme aracını veya araçlarını belirlemeden öncelikle "ölçmek istediğim nitelik ne, neyi ölçüyorum" sorusunun cevabını aramak biz eğitimcilerin en temel düşünmesi gereken soru olması gerektiği inancındayım. Özellikle tıp eğitimi gibi öğretilemeyen ya da ölçülemeyen bir veya birkaç kazanımın hasta kaybına neden olabileceği gerçeği düşünüldüğünde; tıp eğitiminde kullanılacak ölçme araçlarının güvenilir ve geçerli olması diğer bilim alanlarına göre çok daha büyük hassasiyet gösterilmesini gerekli kılmaktadır.

Kaynaklar

- 1) ACGME Outcome Project. (2007) Accreditation Council for Graduate Medical Education (ACGME) & American Board of Medical Specialist (ABMS).
- 2) Amin, Z.; Chong, Y.S. & Khoo, H.E. (2006) Practical Guide to Medical Student Assessment, World Scientific Publishing Co. Pte. Ltd.
- 3) Anıl, D. (2009) Uluslararası Öğrenci Başarılarını Değerlendirme Programı (PISA)'nda Türkiye'deki Öğrencilerin Fen Bilimleri Başarılarını Etkileyen Faktörler. Eğitim ve Bilim (Education and Science), 34(152), 87-100 (2009).
- 4) Anıl, D. (2005) Ürüne Dayalı Program Değerlendirme Yaklaşımı, Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 5 (1), 67-89 (2005).
- 5) Anıl, D., (2013) Eğitimde Ölçme ve Değerlendirme. Ahmet Yesevi Uluslararası Türk-Kazak Üniversitesi online ders kitabı, TÜRTEP.
- 6) Baykul, Y. (2000). Eğitimde ve Psikolojide Ölçme: Klasik test Teorisi ve Uygulaması. ÖSYM Yayınları, Ankara.
- 7) Bloom, B.S., Hastings J.T. & Maduals, G.F. (1971) Handbook on formative and summative evaluation of student learning. NewYork: Mc.Graw Hill Book Co.
- 8) Embretson, S.E. & Reis, S.P. (2000) Item Response Theory for Psychologists. L. Erbaum Associates.
- 9) Epstein, R.M. & Hundert, E.M. (2002) Defining and Assessing Professional Competence. JAMA, January 9, 287(2).
- 10) Elçin, M., Odabaşı, O. ve Sayek, İ. (2005). Yapılandırılmış Objektif Klinik Sınavlar. Hacettepe Tıp Dergisi, 36, 1-2.
- 11) Ertürk, S. (2013) Eğitimde "Program" Geliştirme. Edge Akademi.
- 12) Miller GE. (1990) The Assessment of Clinical Skills/ Competence/ performance. Acad Med; 63-67.
- 13) Ornstein, C.A. & Hunkins, F.P. (1988). Curriculum: foundation, principles and issues. Prectice Hall, Englewood Cliffs.
- 14) Posner, G. J. (1995) Analyzing the curriculum. McGraw- Hill Companies.
- 15) TBB (2010) Türk Tabipleri Birliği Mezuniyet Öncesi Tıp Eğitimi Raporu, Türk Tabipleri Birliği Yayınları.
- 16) Thorndike, R.M. & Christ, T.T (2011) .Measurement and Evaluation in Psychology and Education (8th Edition) .
- 17) Turgut, M. Fuat (1997). Eğitimde Ölçme Değerlendirme Metodları, Yargıcı Matbaacılık (Onuncu Baskı), Ankara.