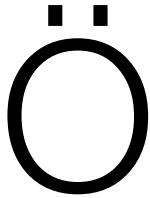


Tıp eğitimi: Ölçme ve değerlendirilmede becerilerin değerlendirilmesi

Dr. Ahmet Murt



1986 yılında İskenderun'da doğdu. Ortaöğrenimini İstiklal Makzume Anadolu Lisesi'nde tamamladıktan sonra İÜ Cerrahpaşa Tıp Fakültesi'ne girdi. Mayıs 2008'de Cerrahpaşa Tıp Fakültesi'nde öğrencilere yönelik Tıp Eğitimi Kongresi projesini hayata geçirdi. 2008-2009 akademik yılında TurkMSIC Genel Başkanlığı ve Uluslararası Tıp Öğrencileri Birliği Derneği Yönetim Kurulu Başkanlığı yaptı. Yine aynı dönemde Avrupa Tıp Öğrencileri Birliği'nde (EMSA) tıp eğitimi komitesinden sorumlu yönetim kurulu üyesi olarak görev yaptı. Kasım 2009'dan bu yana Avrupa Komisyonu destekli MEDINE (Medical Education in Europe) akademik işbirliği projesinin yönetim kurulu üyeliği görevini yürütmektedir. Cerrahpaşa Tıp Fakültesi'nden geçen yıl mezun olan ve zorunlu hizmetini Hakkâri, Yüksekova'da tamamlayan Dr. Murt, halen Cerrahpaşa Tıp Fakültesi Anatomi Anabilim Dalında asistan doktor olarak görev yapmaktadır.



ölçme ve değerlendirilmenin öğrenmeyi teşvik edici olduğu gerçeğinin altını önceden çizmiştik. Sınavların, öğrenmeyi olumlu yönde etkilemesi konusundaki başarısını arttırabilmek için ölçme ve değerlendirme unsurlarını eğitsel amaçlara daha

uygun hale getirmemiz önceliklerimiz arasında olmalıdır. Tıp fakültelerinde "müfredat"ın oluşturulması ve geliştirilmesi önem sırasında genelde ön sırada olsa da; öğrencilerin, sınavlarda sorulan bölümlere daha fazla önem verecek olmaları kaçınılmaz bir sonuçtur. Tabii; müfredatın seneler içinde "obezite" sorunu ile karşılaşmasının,

öğrencilere altından kalkamayacakları bir iş yükü oluşturduğu önermesi de yanlış değildir. Bu durumun hiç şaşırılmayacak doğal sonucu da öğrenciler tarafından sadece "işe yarayacak sınavdan daha fazla not alınmasını sağlayacak" bölümlerin önemsenmesi olacaktır. Bu durumda ölçme ve değerlendirmeyi, öğrenmeyi kamçıla-



yıcı bir dinamo olarak kabul etmek eğitsel amaçlarımızla daha fazla örtüşür.

Önceki yazılarımızda “Öğrenmek için değerlendirme” yaklaşımının öneminden bahsetmeye çalışmıştık. Bu yaklaşımın önemsenmesi gerektiğini düşünüyorum. Yalnız; şurası unutulmamalıdır ki; doktorların performanslarına ve yeterliklerine daha fazla odaklanan toplumsal gereklilikler “karar verdirici - summative” yöntemlerin önemini biraz daha artırır durumdadır. Yazılı sınavların öğrenciler hakkında verdikleri bilgiler çoğu zaman yetersiz kalabilmektedir. Bu yüzden gerek ülkemizde gerekse Avrupa'nın birçok önde gelen fakültesinde “sözlü sınavlar” becerilerin ve klinik yeterliklerin değerlendirilmesi için kullanılmaktadır. Burada şunu not etmemiz gerekir ki geleneksel sözlü sınavlar “güvenilirlik” konusundaki soru işaretleri nedeniyle Kuzey Amerika'da son 20 yıldır neredeyse hiç kullanılmamaktadır. Yetiştirdiğimiz doktorların klinik yeterliliğini doğrulukla ölçebilecek yöntemler konusunda ise geniş yelpazede bir arayış mevcuttur.

Öğrencilerimiz, gerçekten sınavlardan daha fazla puan almalarını sağlayacak bölümlere odaklanıyorlarsa; ölçme ve değerlendirme yöntemlerini, müfredatta ön gördüğümüz öğrenim çıktılarına uygun şekilde yapılandırarak fakültenin eğitim hedeflerine ulaşmasını sağlayabiliriz. Yani ilk bakışta soru gibi görülen bir durumu, eğitsel çıkarlarımızın lehine çevirebiliriz. Bu bağlamda, sınav içeriklerini dikkatlice tasarlayarak; sınavların öğrenim çıktılarıyla uyum içerisinde olduğu bir planlama yapmak oldukça değerli bir ilk adım olacaktır. Müfredatımızda belirlediğimiz hedeflerin bilgi, beceri ve yeterlikler olmak üzere farklı çerçevelerde tanımlanmış olduğunu hatırlatarak, tüm bunların tek bir sınav biçimi ile ölçülebilmesinin imkânsızlığını özellikle vurgulamak isterim.

Güvenilirlik ve geçerlilik kavramlarına önceden değinmiş olsak da becerilerin değerlendirilmesi ile ilişkilendirerek bazı açıklamalar yapmak yerinde olacaktır. *Güvenilirliğin* tanımının “*öğrencinin bir sınavdan alacağı puanın, aynı konuda (veya alanda) yapılacak bir başka sınavda alacağı puan ile hemen hemen aynı olması*” olduğunu önceden paylaşmıştık. Güvenilirlik bir başka deyişle; sınav sonuçlarının aynı şekilde tekrar edebilirliği yani tutarlı olmasıdır. Becerilerin değerlendirilmesinde güvenilirliği; sınavı yapanın yargıları, kullanılan olgular, adayın heyecanı ve fiziksel koşullar gibi faktörler birincil olarak etkilemektedir. Gelin burada bir adım daha ileriye gidelim ve güvenilirliğin iki alt başlığından; *değerlendirenler arası* ve *olgular arası* güvenilirlikten bahsedelim. *Değerlendirenler arası güvenilirlik*, sınavı yapan farklı

eğitmenlerce benzer notların verilip verilmediği ile ilgilidir. Bu konu, gerek öğrenciler arasında gerekse hocalarımız arasındaki yapılandırılmamış tartışmalarda çokça dillendirilmiştir. Sanıyorum sizler de benim gibi; öğrencinin sözlü sınavdan (veya beceri sınavından) aldığı notu sınavından sorumlu hocası ile ilişkilendirdiğine tanıklık etmişsinizdir. Gerçi öğrencilerimizin kimi zamanlar bu konuda haklı şikâyetlerde bulunuyor oldukları çoğumuza uzak gelen olasılıklardandır (!) Ne var ki, sınavı yapan her bir hocanın öğrenciye yaklaşımında veya sorduğu soruların kapsamında, diğerlerinden farklılık göstereceği su götürmeyecek bir gerçektir. Şurası net bir şekilde görülmüştür ki; her biri öğrenciye birer soru soracak şekilde 10 akademisyen tarafından yapılan sözlü sınav, 1 akademisyen tarafından yapılacak 10 soruluk sözlü sınava göre çok daha güvenilir (veya tutarlı) sonuçlar ortaya koymaktadır. Fakültelerimizin, verdiği eğitimin yanında hizmet sunumu ve bilimsel çalışmalarla içinde bulunduğu yoğunluğunu düşündüğümüzde; her bir öğrenciyi değerlendirmek için 10 akademisyeni hazır bulmamızın güçlüğü hissedilir olsa da buradan çıkarılması gereken ders oldukça basittir: “*Her bir öğrenci için karar verecek akademisyen sayısını imkânlarımız el verdiği ölçüde arttıralım.*” Öğrencilerimizin klinik yeterliğini test ederken, farklı olgulara yönelik yaklaşımlarını tutarlı bir şekilde değerlendirebilmek de önümüzde bir sorun olarak belirebilir. İşte bu sorun net olarak *olgular arası güvenilirlik* alt başlığının çalışma sahasıdır. Klinik yeterliğin daha güvenilir test edilebilmesi için kliniğe yönelik sözlü sınavlarda kullanılan olgu çeşitliliği artırılmalıdır. Bu çeşitliliğin nesnel bir şekilde sağlanabilmesi için farklı olguların yer aldığı istasyonlardan oluşan *OSCE (Objective Structured Clinical Examination)* metodu geliştirilmiştir. Burada olguların çeşitliliğinin sağlanmasına ek olarak unutulmaması gereken bağımsız faktör; öğrenciye her bir olguyla karşılaşmasında yeterli sürenin tanınmasının gerekliliğidir. Bu sınavlarda; öğrencilerimizin becerilerini gösterebilmeleri ve klinik yeterliklerini ispatlayabilmeleri için yazılı sınavlara göre daha fazla zamana ihtiyaçları olacaktır.

Öte yandan *geçerliliğin* tanımının ise; “*sınavın ortaya koymayı amaçladığı bilgi ve becerileri gerçekten ölçüp ölçmediğini değerlendirmek*” olduğunu belirtmiştik. Becerilerin ve klinik yeterliğin değerlendirilmesinde, her bir bileşenin doğru olarak ölçülmesini sağlayacak tek bir metottan bahsedebilmek neredeyse imkânsızdır. Miller'in piramidi, geçerliliğin değerlendirilmesindeki çerçeveyi belirlemesi anlamında bize yol gösterici olabilir (Şekil). 4 katlı bu piramit “Biliyor”, “Nasıl olduğunu biliyor”, “Nasıl uygulayacağını gösteriyor” ve “Uyguluyor” bölümlerinden oluş-

Şurası net bir şekilde görülmüştür ki; her biri öğrenciye birer soru soracak şekilde 10 akademisyen tarafından yapılan sözlü sınav, 1 akademisyen tarafından yapılacak 10 soruluk sözlü sınava göre çok daha güvenilir (veya tutarlı) sonuçlar ortaya koymaktadır. Fakültelerimizin yoğunluğunu düşündüğümüzde her bir öğrenciyi değerlendirmek için 10 akademisyeni hazır bulmamızın güçlüğü hissedilir olsa da buradan çıkarılması gereken ders oldukça basittir: “*Her bir öğrenci için karar verecek akademisyen sayısını imkânlarımız el verdiği ölçüde arttıralım.*”

maktadır. Bu 4 bölümü öğrencilerimizin herhangi bir beceriyi edinmesine uyarlayacak olursak; öğrenci önce bilecek, sonra onun bağlantılarını kurabilecek, daha sonra gözetim altında güvenli bir ortamda nasıl uygulayacağını gösterecek ve en sonunda gerçekten uygulama yetkisine kavuşacaktır. Gerek hekim adaylarını, gerekse kariyerinin farklı basamaklarındaki hekimleri eğiten tıp fakültesinin hedefi bu piramidin zirvesine bayrağını dikmek olmalıdır.

Geleneksel anlamda mezuniyet öncesi tıp eğitiminde en fazla yoğunlaştığımız bölüm, piramidin tabanınıymış gibi durmaktadır. Daha çok teorik bilgileri sorguladığımız yazılı sınavlarımız öğrencinin bildiklerini o kadar derinlemesine test ediyor ki, öğrenciye “biliyor” sertifikasını gönül rahatlığıyla teslim edebiliriz. Öğrencilerin bu bildiklerini

Bundan böyle uygulanması konusunda ciddi bir şekilde kafa yormamız gereken bir başka yöntem ise öğrencilerin birbirlerini değerlendirdikleri “*akran değerlendirme sistemi*”dir. Öğrencilerin eğitim sürecinde kazandıkları profesyonel yetilerin kendi arkadaşları tarafından değerlendirilmesi, her bir öğrenci hakkında değerli bilgilere ulaşılmasını sağlamıştır. Bu sistem, klinik yeterliklerin ve her öğrencinin gösterdiği akademik gelişmenin değerlendirilmesi için de kullanılabilir.

problem çözme veya klinik karar verme süreçlerinde ne kadar kullandığının (yani piramidin bir sonraki basamağı nasıl olduğunu bilip bilmediğinin) değerlendirilmesini sağlayabilmek için de sürekli bir çaba içindeyiz. Günümüzde bu aşama için en büyük silahımız sözlü sınavlardır. İşte burada bazı sorunlar karşımıza çıkmaya başlıyor. Çünkü sözlü sınavlar, yukarıda da değindiğimiz gibi sınavı yapanın yaklaşımına fazlasıyla bağımlı kalıyor. Belirli bir standarda kavuşturulmamış sorular ve sınava ayrılan vaktin çoğu kez yetersiz olması, sözlü sınavların güvenilirliğini zedeliyor. Ayrıca özellikle klinik yeterliği test ederken, bir veya en fazla iki olguya öğrencinin konuya hâkimiyeti ile alakalı genellemeler yapmak yanlış çıkarımlara neden olabiliyor. Çünkü öğrencinin farklı becerileri içeren her olguya yaklaşımı farklılık göstermektedir. Swanson, bir sözlü sınavda öğrencinin yeterliği hususunda gerçeğe en yakın genellemeyi yapabilmek için, -buraya lütfen dikkat-, 12 - 16 arasında farklı olguya yer verilmesi gerektiğini ortaya koymuştur. Sözlü sınavların bu karmaşık yapısı, piramidin ikinci basamağının da yazılı sınavlara konu olmaya başlamasına yol açmıştır. Yazılı sınavlarda kullanılan klinik senaryolarla öğrencinin klinik yaklaşımı değerlendiril-

mek istenmiş ve sınav soruları üzerine ayrıntılı çalışmalar sonucunda başarılı sonuçlara ulaşan soru kalıplarına da ulaşılmıştır. Ülkemizde de son yıllarda artan oranlarda; bu soruları yazılı sınavlarımızda görmekteyiz. Elbette yapılması gereken; çok daha iyilerini ortaya koymaktır. Nitekim dikkatsiz bazı yaklaşımlar birinci ve ikinci basamağın birbirleriyle karıştırılmasına sebep olabilmektedir. Yani siz öğrencinin “klinik bağlantıların nasıl olduğunu bilip bilmediğini” ölçmeye çalışırken, aslında sadece “bildiğini” ölçer durumda kalabilirsiniz. Vaka çözümünü sorularında sadece ipuçlarını verip öğrencinin tanıya ulaşmasını istemek yerine, vakanın klinik bağlantılarını sorgulayarak bu sorunu rahatlıkla aşabilirsiniz. Fakültelerde eğitimsel açıdan çok riskli yaklaşımlardan biri de sözlü sınavların, öğrencinin o dersten geçmesine veya kalmasına karar verecek statüye sahip olabileceğidir. Bu kadar büyük bir yetki, ancak güvenilirliğinden şüphe etmediğimiz metotlara verilmelidir.

Eğitim konusunda iddialı olan fakültelerin, bayrağı, piramidin en tepesine dikmeyi başarabilmesi gerekir. Burada durum çok bilinmeyenli denklem haline dönüşebilir. İkinci basamağın henüz istenilen seviyede olmadığından bahsederken, üçüncü ve sonra dördüncü basamakta yapılması gerekenlere değinmeye çalışmak kolay bir iş olmayacaktır belki. Açıkça söylemek gerekirse, işin bu kısmı ne kadar karmaşık görünse de umut vermektedir. Çünkü fakültelerimizin özellikle birinci aşamada (yani piramidin tabanında, temelinde) üzerine düşen görevi etkili bir şekilde yapabildiğini düşünüyorum. Piramidin tabanı bu denli sağlam olarak muhafaza edildiği sürece, üst parçaları oluşturmaya çalışmak anlamlı olacaktır. Sadece üst basamaklardaki eksiklerimize odaklanarak, onları geliştirmeye çalışırken birinci basamaktaki güçlü yapımızı kaybedersek o gelişmeler pek bir anlam taşımazlar. Tabanı çürüten ve bu sebeple birkaç vadeye kadar yıkılma tehlikesi ile yüzleşecek olan piramit kime ne fayda sunabilir ki?

Üçüncü basamak, öğrencinin “in vitro” performansınıdır. Yani öğrencinin, belirli bir beceriyi gerçek hayatta uygulamadan önce gözetim altında nasıl uygulayacağını göstermesidir. Kuzey Amerika’da çoktan terk edilmiş olsa da Avrupa’nın birçok fakültesinde ve ülkemizde eğitimsel standarda kavuşturulmamış gerçek hastalar klinik yeterliğin test edilmesinde kullanılmaktadır. Belirtmeden geçmeyelim; bu yaklaşımın doğruluğu ve güvenilirliği Avrupa’daki birçok fakültede tartışılmaktadır. Hasta başında öğrencinin alacağı anamnezi ve semptomlara yaklaşımını yapacağı fizik muayene ile ilişkilendirerek değerlendirilen bu yöntem, çoğu kez ikinci - üçüncü basamak arasında gidip gelmektedir. Bizler öğrencinin performansını

gözetim altında test etmeye çalışırken, sadece topladığı ipuçlarıyla kurduğu klinik bağlantıları değerlendirmekle kalıyor olabiliriz. Bu da, sınavı yapanların kendilerini geliştirerek aşmaları gereken bir sorundur. Yine, sadece bir olgu kullanarak öğrencinin klinik performansı ile ilgili genellemeler yapmak, başka bir yanlış kaynağıdır. Öğrencinin performansını doğrudan gözlemediğimiz ve daha fazla olgu bulunduracak şekilde zaman ayarlaması yapılmadıkça, bu yöntem güvenilirliğini her geçen gün daha fazla yitirecektir.

Üçüncü basamaktaki bu sorunun aşılması için bir alternatif olarak ortaya çıkan OSCE, bugün kabul görürlüğünü oldukça artırmıştır. Adaylar birçok olguyu barındıran farklı istasyonlardan geçmekte ve bu sayede becerileri geniş bir yelpazede değerlendirilmektedir. OSCE; birçok farklı olguya yer vermesi ve kişiye bağımlı olmayan yapılandırılmış değerlendirme yaklaşımı ile güvenilirliğini iyice artırmıştır. Ancak, bu yöntem simüle hastaların dahil edilmesi ve öğrencinin performansını doğrudan gözlemleyecek bir ekibin var olması gerekliliği ile yoğun bir iş gücünü gerektirmektedir. Ayrıca diğer yöntemlere göre nispeten daha pahalıdır. Güvenilirliği oldukça yüksek olmasına rağmen, bir defada fazla sayıda olgu ile karşılaşan öğrencinin bazı becerilerinin değerlendirilmesinin atlanması gibi durumlar, yöntemin yönetilmesinde dikkatli olunması gereken hususlardır. İstasyonlarda neden gerçek hastaların kullanılmadığı da yerinde bir tartışma konusudur. “*Simüle hastaları mı, yoksa gerçek hastaları mı kullanalım?*” sorusuna net bir cevap verebilmek de güçtür. Ne var ki, gerçekten nesnel bir sınav yapılandırma istiyorsak, hastanın (veya hasta rolünde olanın) her defasında (her öğrenciye) aynı hikâyeyi vermesi, aynı yakınmalarından bahsetmesi ve aynı ruh durumunu yansıtması gerekir. Bunu sağlayabilmek; eğitimden geçmiş simüle hastalarla çok daha mümkün olacaktır.

Bundan böyle uygulanması konusunda ciddi bir şekilde kafa yormamız gereken bir başka yöntem ise öğrencilerin birbirlerini değerlendirdikleri “*akran değerlendirme sistemi*”dir. Öğrencilerin eğitim sürecinde kazandıkları profesyonel yetilerin kendi arkadaşları tarafından değerlendirilmesi, her bir öğrenci hakkında değerli bilgilere ulaşılmasını sağlamıştır. Bu sistem, klinik yeterliklerin ve her öğrencinin gösterdiği akademik gelişmenin değerlendirilmesi için de kullanılabilir.

Oldukça yeni olan bir diğer yöntem ise 2000’lerin başında Charlin ve arkadaşları tarafından geliştirilen “*Script Concordance Test – SCT*”dir. Bu metot, hekimlerin kendilerine özgü bulgularını işlemekten geçirmeleri ve klinik problemin çözümünü sağlamaları sayesinde

uzmanlıklarını geliştirdikleri hipotezinden yola çıkan "script (kişisel senaryo)" teorisine dayandırılmıştır. Bu yöntemde, tanısı belirsiz bir klinik olgu yazılı bir şekilde eş zamanlı olarak öğrencilere ve uzmanlara sunulur. Öğrencilerin ve uzmanların verdikleri cevapların uyumu (concordance) karşılaştırmalı olarak değerlendirilir. Bu sayede öğrencinin klinik yorum kabiliyeti ortaya konmuş olur. Yapılan araştırmalar yöntemin güvenilirliğinin tatmin edici seviyede olduğunu göstermiştir.

Piramidin son basamağı ise; öğrencinin gerçek hasta ile kendi başına karşı karşıya geldiğinde becerilerinin ve yeterliliğinin ölçülmesidir. Simüle hastalarla oluşturduğumuz OSCE'nin güvenilirliği tatmin edici olsa da gerçek durumlar, üçüncü aşamada üstün başarı göstermiş öğrenciler için dahi zorlayıcı olabilmektedir. Hayalini kurduğumuz (en tepeye bayrağı dikmeyi başarabilen) tıp fakültesi bünyesinde bulundurduğu kariyerinin farklı noktadaki hekimlerin farklı olgulara yaklaşımını düzenli bir şekilde ölçülmeli ve her bir olgu içerisinde öğrencilerin (ve hatta uzmanlık öğrencilerinin) katkılarını değerlendirebilen bir klinik ortamı oluşturmalıdır. Piramidin tepesi, eğitimde kendini başarılı saymak isteyen bir tıp fakültesinin mutlak hedefi olmalıdır. Öte yandan bu sadece bireysel tıp fakülteleri ile sınırlı kalmamalı, konuya ülke çapında yapılandırılmış bir yaklaşım sergilenebilmelidir. Bilgi ve becerilere bu şekilde kademeli bir yaklaşım mezuniyet öncesi eğitimin bir dönemine uyarlanabileceği gibi aşamalar dikey olarak birden fazla seneye de yayılabilir. Fakülteler en uygun yol haritalarını kendi gelenekleri ve müfredat yapılarıyla çizebilirler.

Kaynaklar

Arnold L, Shue CK, Kalishman S, Prislun M, Pohl C, Pohl H, Stern DT. Can there be a single system for peer assessment of professionalism among medical students? A multi-institutional study. *Academic Medicine* 2007; 82(6): 578-586

Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The script concordance test: A tool to assess the reflective clinician. *Teach Learn Medicine* 2000; 12: 189-195

Dauphinee D. Determining the content of certification examinations. In: Newble D, Jolly B, Wakeford R. *The certification and recertification of doctors: issues in the assessment of clinical competence*. Cambridge: Cambridge University Press, 1994: 92-104

Harden RM, Gleeson FA. ASME Medical Education booklet no 8: assessment of medical competence using an objective structured clinical examination. *J Medical Education* 1979; 13: 41-54

Miller GE. The assessment of clinical skills/competence/performance. *Academic Medicine* 1990; 65: 563-67

Murt A. Tıp Eğitimi: Ölçme ve Değerlendirme; öğrenmek için değerlendirme. *Sağlık Düşüncesi ve Tıp Kültürü Dergisi* 18.Sayı

Murt A. Tıp Eğitimi: Ölçme ve Değerlendirme; öğrenmenin değerlendirilmesi. *Sağlık Düşüncesi ve Tıp Kültürü Dergisi* 19.Sayı

Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. *Medical Education* 1996; 22: 325-34

Swanson DB. A measurement framework for performance based tests. In: Hart IR, Harden RM, eds. *Further developments in assessing clinical competence*. Montreal: Can- Heal, 1987: 13-45

Anahtar konular	Önemli not
Karar verdirici/şekillendirici değerlendirme (summative / formative)	Sınavın amacı açıkça belirlenmiş olmalıdır.
Tasarlama (Blueprinting)	Sınav, öğretim çıktılarıyla uyumlu olarak tasarlanmalıdır.
Geçerlilik	Her bilgi, beceri ve yeterlik için en uygun sınav yöntemini seçtiğimizden emin olalım.
Güvenilirlik	Özellikle kliniklerde, farklı becerilerin ölçülmesini sağlayacak olgu sayısı çoğaltılmalıdır. Becerileri değerlendirirken sınav süresi de önemli bir konudur. Ayrıca becerileri değerlendirecek hoca sayısı uygun olarak belirlenmelidir.
Standartların belirlenmesi	Her beceri sınavı için, asgari gereklilikler ve sınavın tamamlanacağı seviye belirlenmiş olmalıdır.



Şekil: Miller'in piramidi