

---

## **Sözlü Sınavlarda Güvenilirlik: Geleneksel Ve Yapılandırılmış Sözlü Sınavların, Çoktan Seçmeli Sınavlardan Alınan Notlarla Gösterdiği Korelasyona Göre Karşılaştırılması**

Mustafa Kemal Alimoğlu\*, Yeşim Şenol\*, Sibel Çubukçu\*, Mehmet Aktekin\*

---

### **ÖZET**

**Gereke:** Geleneksel sözlü sınavların çoktan seçmeli sınavlara göre daha düşük geçerlilik ve güvenilirliğe sahip olduğu bilinmektedir. Yapılandırılmış sözlü sınavlarda ise öğrencilerin, çoktan seçmeli sınavlardan aldıkları puanlara yakın puanlar almaları beklenir. Çalışmanın amacı bu hipotezin doğruluğunu araştırmaktır.

**Yöntem:** 2001-2002 eğitim yılında Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı'nın ilk iki staj döneminde sınava giren 34 öğrencinin geleneksel sözlü ve yazılı notları arasındaki ilişki ile 2002-2003 eğitim yılında aynı anabilim dalının ilk iki staj döneminde sınava giren 31 öğrencinin yapılandırılmış sözlü ve yazılı notları arasındaki ilişki karşılaştırılırken ayrıca, değerlendiricilerin yapılandırılmış sözlülerde aynı öğrenciye verdikleri puanlar arasındaki ilişki araştırıldı. Yapılandırılmış sözlü sınavlarda ve yazılı sınavlarda benzer alanlar (bilgi, analiz, sentez) soruldu.

**Bulgular:** Geleneksel sözlü ile çoktan seçmeli sınavdan alınan puanlar arasında %37,6'lık bir ilişki saptanırken ( $p<0,05$ ); yapılandırılmış sözlü ile çoktan seçmeli sınavdan alınan puanlar arasındaki korelasyon %74,5 olarak hesaplandı ( $p<0,001$ ). Yapılandırılmış sözlülerde değerlendiriciler arası puan ilişkisi %95 bulundu ( $p<0,01$ ).

**Sonuç:** Yapılandırılmış sözlü sınavlar, geleneksel sözlü sınavlara göre daha güvenli ve geçerli biçimde bilgiyi ölçmektedir.

**Anahtar Kelimeler:** Geleneksel sözlü sınav, yapılandırılmış sözlü sınav, geçerlilik, güvenilirlik

---

Dr. Mustafa Kemal Alimoğlu, Akdeniz Üniversitesi, Tıp Fakültesi, Tıp Eğitimi Anabilim Dalı, 07040 Antalya

e-posta: mkemalalimoglu@yahoo.com

\* Akdeniz Üniversitesi, Tıp Fakültesi, Tıp Eğitimi Anabilim Dalı, 07040 Antalya

---

## GİRİŞ

Uzun yıllardır sözlü sınavlar, özellikle klinik alanlarda tıp öğrencilerini değerlendirmek amacıyla en fazla kullanılan yöntemler arasında yer almıştır ve halen de yaygın olarak kullanılmaktadır (1). Öğretim üyeleri arasında, sözlü sınavların sahip oldukları bazı avantajlar nedeniyle modern yöntemlere göre daha geçerli olduğu, yani ölçülmek istenenin bu yöntemle daha iyi ölçülebildiği kanısı hakimdir. Sözlü sınavın, eğiticiler tarafından çok tercih edilmesinin belki de en önemli nedeni yazılı sınavlardan farklı olarak öğrenciyle doğrudan aktif temas kurma olanağı vermesidir. Bu sayede öğrencinin bilgiyi analiz etme, yorumlama, düşüncelerini organize ve ifade etme gibi entelektüel becerilerini daha derinlemesine ölçme ve ölçerken de esnek davranabilme şansı doğmaktadır. Bu esneklik öğrencinin zayıf ve güçlü taraflarının daha belirgin olarak ortaya çıkarılabilmesine ve bu konuda öğrenciye anında geri bildirim verilebilmesine olanak tanımaktadır (2).

Sözlü sınavların, öğretim üyeleri tarafından tercih edilmesine rağmen, giderek daha az kullanılmasına ve yerini yapılandırılmış yazılı ya da klinik sınavlara bırakmasına yol açan çeşitli dezavantajları vardır. Geleneksel sözlü sınavlar genellikle plansızdır, öğretim üyesi karşısına gelen öğrenciye istediğini ya da o anda aklına gelen bir soruyu sorabilir. Belki de öğrenciden beklediği yanıt, kendi meslek yaşamında bile hiç kullanmadığı bir bilgiyi içeriyor olabilir. Düşük geçerlilikleri yanında sözlü sınavlar yazılı sınavlara göre daha fazla duygusal gerginliğe yol açan sınavlardır (3) ve bu gerginliğin öğrenci performansını olumsuz yönde etkilediği bulunmuştur (4). Geleneksel sözlü sınavlarda öğrencinin sözel performansı, iletişim kurmadaki ustalığı onu olduğundan daha çok ya da daha az biliyormuş gibi gösterip, değerlendiriciyi etkileyebilir. Öğrencinin tavrı, kılık kıyafeti, öğretim üyesinin geçmişte öğrenciyle ilgili edindiği izlenimler ya da ilişkiler sınav sırasındaki kararları etkileyebilir (5, 6). Geleneksel sözlü sınavların öğrenci, yöntem ya da değerlendiriciden kaynaklanan ve güvenilirliği zedeleyen ciddi eksikliklerinin olduğu yapılan araştırmalarla belirlenmiştir (1,5-10). American National Board of Medical Examiners tarafından yapılan bir çalışmada sözlü sınav yapanlar arasındaki korelasyonun şans eseri oluşabilecek herhangi bir korelasyondan çok farklı olmadığı gösterilmiştir (11).

Yine araştırmaların gösterdiğine göre tıp fakültelerindeki sözlü sınavlarda genellikle izole bilgileri ölçmekten öteye gidememektedir ki bu durumda objektif yazılı sınavların kullanılması çok daha güvenilir bir ölçüm yapma olanağı verecektir (8). Ancak öğretim üyeleri yapılandırılmış yazılı ya da klinik sınavlardaki pasif rollerinden rahatsızlık duymaktadır. Buna bir çare olarak yapılandırılmış sözlü sınav kavramı ortaya konmuştur (2).

Yapılandırılmış sözlü sınavlar, geleneksel sözlülerin geçerlik ve güvenilirliklerini azaltan faktörlerin ortadan kaldırılmaya çalışıldığı değerlendirme yöntemleridir. Bu sınavlarda önceden belirlenmiş klinik senaryolar kullanılır. Öğrenciye senaryo verilir ve buna ilişkin sorular sorulur. Sınava giren öğrenciler birbirlerini görmeyeceklerse tüm öğrencilere aynı sorular sorulabilir, eğer sınavdan çıkan öğrenci sınav bitmeden önce diğerleri ile karşılaşacaksa o zaman içerik ve zorluk açısından eş değer sorular hazırlanıp, kullanılır. Sınavda en az iki değerlendirici bulunur ve değerlendiriciler birbirlerinden bağımsız olarak puanlama yaparlar. Değerlendirici sayısı artıkça sözlü

sınavın güvenilirliğinin arttığını gösteren çalışmalar mevcuttur (12). Sınavın ve değerlendirilmenin nasıl yapılacağı konusunda öğretim üyeleri sınavdan önce tam olarak bilgilendirilir. Değerlendiricilerin verdiği puanlar arasındaki tutarsızlığı en aza indirmek için yapılandırılmış puanlama formları oluşturulur. Bu formlarda öğrencinin vereceği her türlü yanıtın puan olarak karşılığı bulunmakta ve öğretim üyesinden yanıtı göre bu puanları yazma ya da işaretlemesi beklenmektedir. Böylece değerlendiricilerin öğrenciye verdikleri puanlar arasındaki tutarlılık artmaktadır (11)

Araştırmalar yapılandırılmış sözlü sınavların klinik bilgi ve problem çözme becerilerini ölçmek için uygun ve güvenilir olduğunu, değerlendirici ve öğrencilerin sınavı daha tatminkar bulduklarını göstermektedir (2, 7, 13). Karşılaştırmalı çalışmalarda yapılandırılmış sözlü sınav, çoktan seçmeli sınav ve objektif yapılandırılmış klinik sınavlar arasında anlamlı korelasyon bulunmuştur (13, 14).

Bizim çalışmamızın amacı, yapılandırılmış sözlü sınavlarla geleneksel sözlü sınavları, çoktan seçmeli sınavlarla gösterdikleri korelasyona göre karşılaştırmak ve yapılandırılmış sözlü sınavlarda aynı öğrenciyi değerlendiren öğretim üyelerinin verdikleri puanlar arasındaki korelasyonu araştırmaktır. Yapılandırılmış sözlü sınavlar ve yazılı sınavlarda benzer alanlar(bilgi, analiz, sentez) sorulmuştur.

## GEREÇ-YÖNTEM

Akdeniz Üniversitesi Tıp Fakültesi 2001-2002 eğitim-öğretim döneminde öğretim elemanlarına yönelik ölçme ve değerlendirme kursları düzenlenmiştir ve tüm öğretim elemanları kurslara katılmıştır. Kurslar sonrasında klinik branşlarda yapılandırılmış sözlü sınavlar uygulanmaya başlanmıştır.

Tıp fakültesi 5. sınıf öğrencileri 3 haftalık bir eğitim sürecinden geçmek üzere gruplar halinde Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı'na gelmektedir. Bu eğitim süreci sonunda sözlü sınav ve ardından sözlü sınavdan başarılı olan öğrencilerin katılabildiği bir çoktan seçmeli sınav yapılmaktadır. Eski uygulamalarda sözlü sınavlara katılan öğrencilere ders başlıkları göz önüne alınarak herhangi bir konudan sorular sorulabilirken, 2002-2003 eğitim-öğretim döneminin başlangıcından itibaren yapılandırılmış tip sözlü sınavlar uygulanmaya başlanmıştır.

Çalışmamızda 2001-2002 eğitim-öğretim yılında Fiziksel Tıp ve Rehabilitasyon Anabilim dalının ilk iki staj döneminde geleneksel sözlü sınav ve ardından çoktan seçmeli sınava giren 34 öğrenci ile 2002-2003 eğitim-öğretim yılında Fiziksel Tıp ve Rehabilitasyon Anabilim dalının ilk iki staj döneminde yapılandırılmış sözlü sınava ve ardından çoktan seçmeli sınava giren 31 öğrencinin sonuçlarının çoktan seçmeli sınav sonuçları ile ilişkisine bakılmıştır. Tüm sınavlarda 100 puan üzerinden 60 ve üzerinde puan alan öğrenciler başarılı sayılmıştır.

Ayrıca yapılandırılmış sözlü sınavda aynı salonda görev yapan ve aynı öğrenciyi değerlendirirken birbirlerinden bağımsız olarak verdikleri puanların çalışmamızda kullanılmasına izin veren iki öğretim üyesinin verdikleri puanların birbiriyle korelasyonu araştırılmıştır.

Geleneksel sözlü sınav ve yapılandırılmış sözlü sınav puanlarının çoktan seçmeli sınav puanlarıyla ilişkisini ve öğretim üyelerinin aynı öğrenciye verdiği puanların ilişkisini ortaya koymak için korelasyon analizi uygulanmıştır.

## BULGULAR

2001-2002 eğitim-öğretim yılının ilk iki staj döneminde Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı tarafından düzenlenen geleneksel tipteki sözlü sınavlara toplam 37 öğrenci girmiş, üç tanesi başarılı olamayarak yazılı sınava girmeye hak kazanamamıştır. Sınavdan geçme oranı %91,9'dur. Kalan 34 öğrencinin geleneksel sözlü sınavdan aldıkları ortalama puan  $75,8 \pm 9,7$  (60-96 arası), çoktan seçmeli sınavdan aldıkları ortalama puan ise  $80,8 \pm 11,6$  (53-100 arası) idi. Bir öğrenci geleneksel sözlü sınavdan 60 puan alarak başarılı olmasına rağmen yazılı sınavdan 53 puan alarak kalmıştır. Geleneksel sözlü sınav ve çoktan seçmeli sınav puanları arasındaki ilişki 0,38 olarak bulundu ( $p < 0.05$ ) (tablo-1 ve şekil-1).

2002-2003 eğitim-öğretim yılının ilk iki staj döneminde Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı tarafından düzenlenen yapılandırılmış sözlü sınavlara toplam 33 öğrenci girmiş, iki tanesi başarılı olamayarak yazılı sınava girmeye hak kazanamamıştır. Sınavdaki başarı oranı %93,9'dur. Kalan 31 öğrencinin yapılandırılmış sözlü sınavdan aldıkları ortalama puan  $78,2 \pm 9,4$  (60-100 arası), çoktan seçmeli sınavdan aldıkları ortalama puan ise  $78,0 \pm 6,2$  (66-86 arası) idi. Yapılandırılmış sözlü sınavdan geçer puan alıp yazılı sınavda başarılı olamayan öğrenci yoktu. Yapılandırılmış sözlü sınav ve çoktan seçmeli sınav puanları arasındaki ilişki 0,74 olarak bulundu ( $p < 0.001$ ) (tablo-1 ve şekil-2).

Yapılandırılmış sözlü sınavlarda görev alan iki öğretim üyesinin birbirinden bağımsız olarak 26 öğrenciye verdikleri ortalama puanlar birinci öğretim üyesi için  $71,1 \pm 11,8$  (44-90 arası) iken ikinci öğretim üyesi için  $71,3 \pm 12,6$  (45-90 arası) idi. İki bağımsız değerlendiricinin verdikleri puanlar arasındaki ilişki 0,95 bulundu ( $p < 0.01$ ) (şekil-3).

Tablo - 1: Geleneksel sözlü sınavlar ile yapılandırılmış sözlü sınavlardaki başarı oranları ve bu sınavlardan alınan puanların çoktan seçmeli sınav puanları ile karşılaştırılması

	Eğitim-öğretim yılı	Sınavı yapan ABD	Katılan / Başarılı olan öğrenci sayısı	Başarı oranı (%)	ÇSS ile korelasyonu (r)	P*
Geleneksel Sözlü Sınav	2001-2002	FTR	37 / 34	91,9	0,38	<0.05
Yapılandırılmış Sözlü Sınav	2002-2003	FTR	33 / 31	93,9	0,74	< 0.001

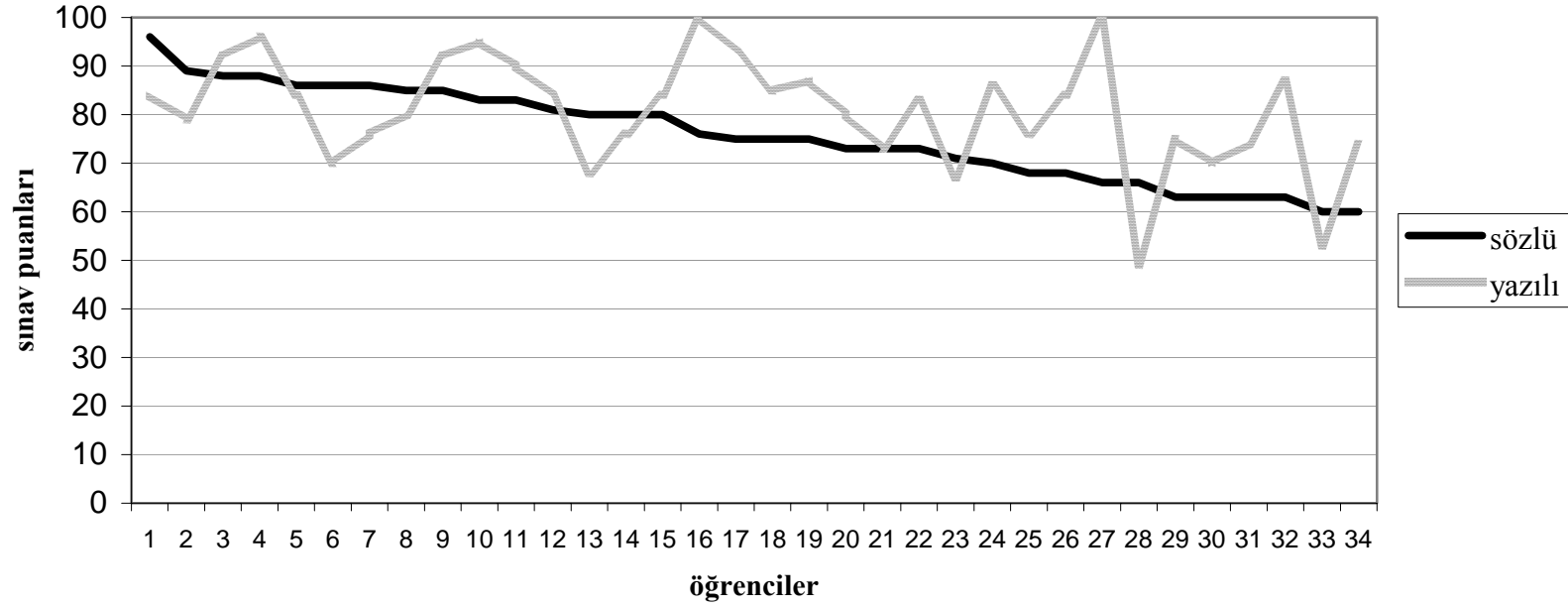
ABD Anabilim Dalı

FTR Fiziksel Tıp ve Rehabilitasyon

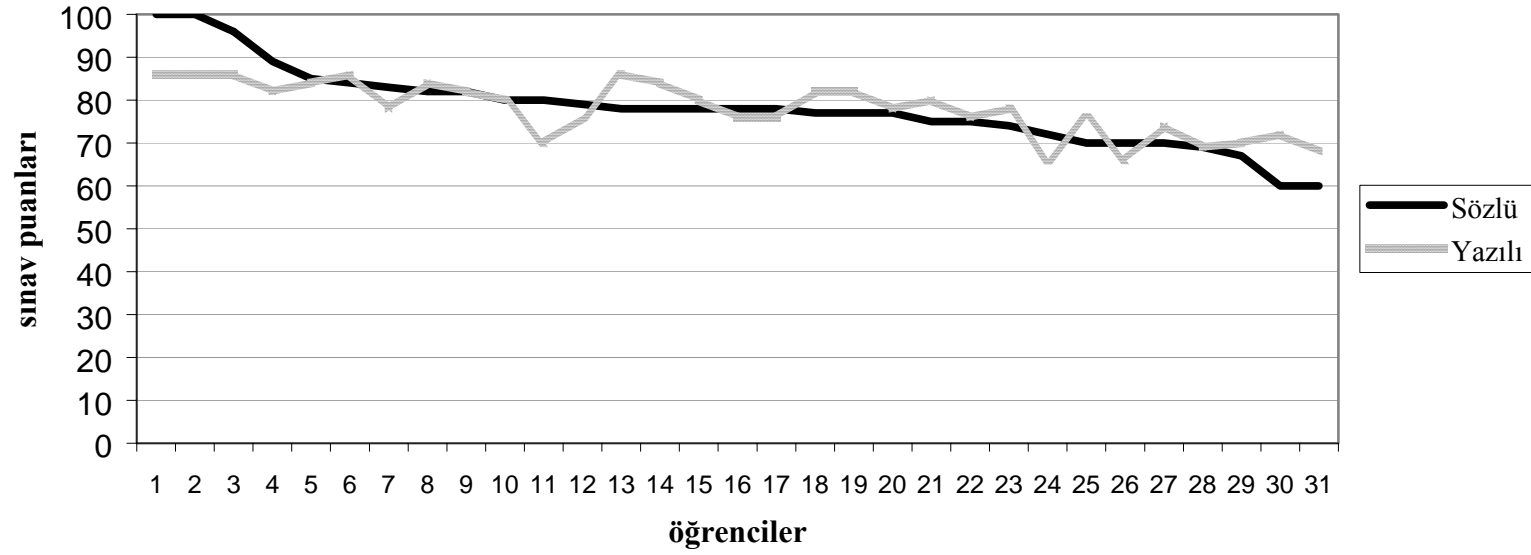
r Korelasyon katsayısı

\* Korelasyon analizi

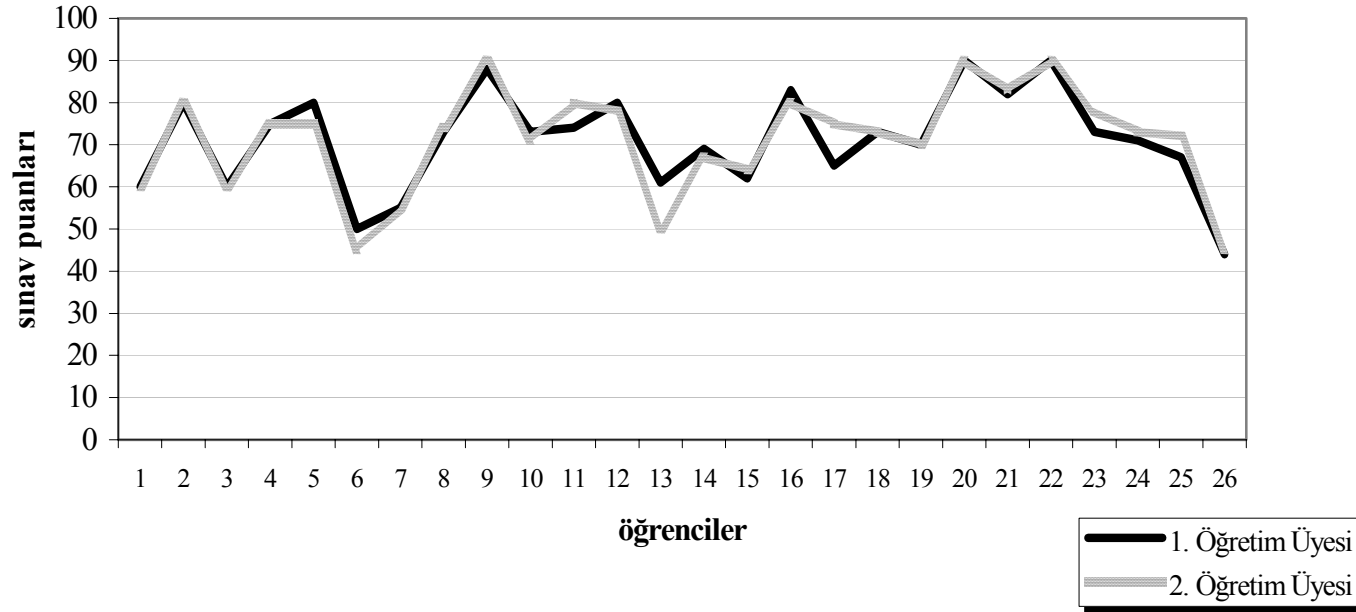
Şekil - 1 Geleneksel sözlü sınav ile çoktan seçmeli sınav ilişkisi ( $r= 0,38, p<0.05$ )



Şekil - 2 Yapılandırılmış sözlü sınav ile çoktan seçmeli sınav ilişkisi ( $r= 0,74, p<0.001$ )



Şekil - 3 Aynı öğrenciyi değerlendirirken birbirinden bağımsız puan veren iki öğretim üyesinin verdikleri puanların birbiriyle ilişkisi ( $r=0,95$ ,  $p<0,01$ )



## TARTIŞMA

Sözlü sınavların güvenilirlikle ilgili ciddi eksikliklerinin olduğu artık çok iyi bilinmekte ve bu eksiklikleri gidermek için bazı yeni yöntemler geliştirilmektedir. Bu yöntemler gerçek ya da gerçeğe yakın senaryolar üzerinde sınavların yapılandırılması ve objektifliği artırıcı somut kriterlerin ortaya konması esasına dayanmaktadır. Tıp fakültesi öğrencilerinin bilgi ve becerilerinin bu tür gerçeğe yakın senaryolar üzerinden ölçülmesi ilerideki meslek performansları hakkında daha güvenilir bir ön bilgi edinilmesine yardımcı olmaktadır.

Yeni geliştirilen yöntemlerin geçerlilik ve güvenilirliğini araştırmak için çeşitli çalışmalar yapılmaktadır. Allison ve Katona (14) geleneksel sözlü sınav puanlarıyla çoktan seçmeli sınav puanları arasındaki ilişkinin sözlü sınavların geçerliliğini ve güvenilirliğini belirlemede yararlı bir gösterge olduğunu bildirmiştir. Aradaki ilişki ne kadar belirgin olursa, sözlü sınavın güvenilirliğinin de o kadar yüksek olması beklenir. Çalışmamızda yapılandırılmış sözlü ve çoktan seçmeli sınav puanları arasındaki ilişki geleneksel sözlü ile çoktan seçmeli sınav puanları arasındaki ilişkiden daha yüksek bulunmuştur. Bu bulgu, yapılandırılmış sözlü sınavın klinik bilgi ve problem çözme becerilerini test etmede geleneksel sözlü sınava göre daha etkin ve güvenilir bir yöntem olduğu düşüncesini desteklemektedir. Bu görüşü destekleyen başka çalışmalar da mevcuttur (2, 7, 13).

Geleneksel sözlü sınavların düşük olan değerlendiriciler arası güvenilirliğinin yapılandırılmış sözlü sınavlarda yükseldiği bilinmekte olup, bizim çalışmamızda da değerlendiricilerin aynı öğrencilere verdikleri puanlar arasında yüksek korelasyon bulunması bu görüşü desteklemektedir.

Sözlü sınavlarla genellikle bilgiyi ölçmekten öteye gitmeyen soruların sorulması, bu sınavlar yerine yapılandırılmış yazılı sınavların kullanılmasının daha doğru olacağı kanısını doğurmaktadır. Eğer sözlü sınavda beceri de ölçülmek isteniyorsa o zaman tercih edilecek sınav türü objektif yapılandırılmış klinik sınavlar olabilir (8). Ancak öğretim elemanlarının bu tür yapılandırılmış sınavlardaki pasif rollerinden rahatsızlık duyması sözlü sınavları vazgeçilmez kılmaktadır (2). Ayrıca yer ve zaman kısıtlılığı, değerlendirici sayısının yetersizliği, öğrenci sayısının fazlalığı gibi nedenlerle objektif yapılandırılmış klinik sınavlar yaygın uygulama alanı bulamamakta, yapılandırılmış sözlü sınavlar tercih edilmektedir.

Fakültemizde yapılandırılmış sözlü sınavların uygulamaya konmasıyla öğretim üyeleri ortak senaryolar hazırlama, senaryolara ilişkin sorular belirleme çabası içine girmiş, sınavın her parçasının öğrenim amaç ve hedefleriyle uyumlu olması konusunda özellikle titiz davranmışlardır. Bu çabayla hazırlanan sınavların içeriği de eski geleneksel sözlü sınavların plansız içeriğine göre çok daha yüksek geçerliliğe sahip olmuştur.

Fakültemiz öğretim üyeleri, yapılandırılmış sözlü sınavların değerlendirici olarak kendi duygusal gerginliklerini de azalttığını ve öğrenci hakkında vicdanen daha rahat karar verebildiklerini de belirtmişlerdir.



Üniversitemizin gelecek yıllardaki hedefi gerekli alt yapı ve motivasyonu sağlayarak objektif yapılandırılmış klinik sınavlarla öğrencileri değerlendirmek, yapılandırılmış sözlüye ise bu sınavların içerisinde bir bölüm ayırmaktır.

### **Kaynaklar**

- 1) Daelmans HE, Scherpbier AS, Van Der Vleuten CP, Donker AS. Reliability of clinical oral examinations re-examined. *Med Teach* 2001 Jul;23(4):422-424
- 2) Amiel GE, Tann M, Krausz MM, Bitterman A, Cohen R. Increasing examiner involvement in an objective structured clinical examination by integrating a structured oral examination. *Am J Surg* 1997 Jun;173(6):546-9
- 3) Spangler G. Psychological and physiological responses during an exam and their relation to personality characteristics. *Psychoneuroendocrinology* 1997 Aug;22(6):423-41
- 4) Thomas CS, Mellsop G, Callender K, Crawshaw J, Ellis PM, Hall A, McDonald J, Silfverskiold P, Romans-Clarkson S. The oral examination: a study of academic and nonacademic factors. *Med Educ* 1993 Sep;27(5):433-9
- 5) Rowland-Morin PA, Burchard KW, Garb JL, Coe NP. Influence of effective communication by surgery students on their oral examination scores. *Acad Med* 1991 Mar;66(3):169-71
- 6) Burchard KW, Rowland-Morin PA, Coe NP, Garb JL. A surgery oral examination: interrater agreement and the influence of rater characteristics. *Acad Med* 1995 Nov;70(11):1044-6
- 7) Yang SC, Laube DW. Improvement of reliability of an oral examination by a structured evaluation instrument. *J Med Educ* 1983 Nov;58(11):864-72
- 8) Jayawickramarajah PT. Oral examinations in medical education. *Med Educ* 1985 Jul;19(4):290-3
- 9) Raymond MR, Webb LC, Houston WM. Correcting performance-rating errors in oral examinations. *Eval Health Prof* 1991 Mar;14(1):100-22
- 10) McCormick WO. A practice oral examination rating scale-inter-observer reliability. *Can J Psychiatry* 1981 Jun;26(4):236-39
- 11) Newble D, Cannon R. Assessing the students, oral in *A Handbook for medical teachers-3<sup>rd</sup> edition (reprinted version)*. Kluwer Academic Publishers, London 1996:134-5
- 12) Swanson R, Swanson S, Spooner J, Haight K, Ramsden V, Tan L. Inter-rater variability in an advanced cardiac life support course: a case study. *Med Teach* 1987;9(4):447-9
- 13) Anastakis DJ, Cohen R, Reznick RK. The structured oral examination as a method for assessing surgical residents. *Am J Surg* 1991 Jul;162(1):67-70
- 14) Allison R, Katona C. Audit of oral examinations in psychiatry. *Med Teach* 1992;14(4):383-9