

Tıp Fakültesindeki Çoktan Seçmeli Sınav Sorularının Değerlendirilmesi

Evaluation of Multiple Choice Questions (MCQ) Exams at Faculty of Medicine

Zeliha Cansever¹, Hamit Acemoğlu², Ümmü Zeynep Avşar², Salih Hoşoğlu³

¹Mevlana Üniversitesi Tıp fakültesi Tıp Eğitimi ve Bilişimi Anabilim Dalı, Konya

²Atatürk Üniversitesi Tıp fakültesi Tıp Eğitimi Anabilim Dalı, Erzurum

³Fatih Üniversitesi Mikrobiyoloji Anabilim Dalı

ÖZET:

Amaç: Tıp eğitiminde çoktan seçmeli sorular, yüksek kapsam geçerliliği olması, objektifliği, uygulama ve analizinin kolaylığı nedeniyle sık tercih edilir. Ancak, çoktan seçmeli soru hazırlamak ustalık gerektirir. Bu çalışmada Atatürk Üniversitesi Tıp Fakültesi'nde sınavlarda sorulmuş çoktan seçmeli soruların analizinin yapılması (zorluk ve ayıricılık indeksi) ve Modifiye Bloom Taksonomisi temel alınarak düzeylerinin belirlenmesi amaçlanmıştır. Ayrıca eğitici eğitimi alan ve almayanların soruları, indeks değerlerine göre sınıflandırılmıştır.

Anahtar Sözcükler:

Çoktan Seçmeli soru maddeleri, zorluk indeksi, ayıricılık indeksi

Key Words:

Multiple-choice test items, difficulty index, discrimination index

Gereç ve Yöntem: Kurul ve final sınavlarından rastgele yöntemle on ayrı sınavdan seçilen 945 soru çalışmaya alındı. Modifiye Bloom Taksonomisine göre düzeyleri belirlendi. Soruların zorluk ve ayıricılık analizleri yapıldı. Veriler, sayı, yüzde, ortalama ve standart sapma

olarak verildi. Sürekli verilerin normallik dağılımı Kolmogorov Smirnov ile test edildi. Verilerin analizinde Ki-kare, bağımsız gruplarda t testi ve tek yönlü ANOVA kullanıldı. Çoklu karşılaştırma Post Hoc LSD testi kullanıldı.

Bulgular: İyi ve çok iyi kalitedeki soru oranı toplam %44,1 (n= 416) idi. Doçentlerin hazırladığı sorular yardımcı doçent ve profesörlere göre ayıricılık indeksi daha yüksek ancak daha kolaydı. Modifiye Bloom Taksonomisine göre soruların büyük bir kısmı hatırlama ve anlama düzeyindeydi. Soruların düzeye göre zorluğunda fark yokken anlama ve değerlendirme düzeylerindeki sorular daha ayırıcı bulundu (p=0,021). Öğretim üyelerinin eğitici eğitimi alıp almamaları karşılaştırıldığında eğitim alanların soruları daha kolaydı ve ayıricılıklarında istatistiksel olarak fark bulunmadı (p>0,05).

Sonuç: Sınavlardan sonra madde analizlerinin yapılarak öğretim üyelerine bildirilmesi ile soru kalitesinin artırılması sağlanabilir. Eğiticilerin soru hazırlama için eğitim almaları önemlidir.

ABSTRACT

Background: MCQ's are often preferred due to ease of application and analysis, high content validity, and objectivity. But, skill required for preparing of multiple choice questions (MCQ). In this study we determined the difficulty and discrimination indices as well as classify based on Modify Bloom's Taxonomy (BT), by analyzing the multiple choice questions present in the examinations held at Ataturk University Medical Faculty. In addition, we classified the MCQ difficulty and discrimination indices between those received training course and not.

Methods: Randomly chosen from ten different committee exams and final exams of 945 questions analyzed. Levels were determined according to the Bloom Taxonomy and were performed difficulty of the questions and discrimination indexes. Data were expressed as the number, percentages, mean and standard deviation (SD). Continuous data normality determined by Kolmogorov-Smirnov test. Data Analyzed using chi-Square and t, one-way analysis of variance test. Post Hoc pairwise comparisons were performed using LSD multiple comparisons.

Results: The good quality questions were 44.1% (n= 416). Discrimination index of questions given to Associate professors were higher and easier than compare to Assistant professors and Professors. We assessed difficulty and discrimination index according to the BT level. Greater number of questions were in level of recall and comprehension. Apparently there was no difference in average index of difficulty level in statistically the discrimination index of questions found statistically higher in comprehension and evaluation levels (p=0,021). The difficulty level were found lower in who have received of training course when compared

to the group who have not received course. Meanwhile there was no difference questions of discrimination index in these groups.

Conclusions: Quality of questions can be improved by informing faculty members of the post-exam analysis of exam questions. It is therefore imperative that instructors be educated on preparing exam questions.

Sorumlu Yazar: Yrd. Doç. Dr. Zeliha Cansever
Mevlana Üniversitesi Tıp Fakültesi, Tıp Eğitimi
ve Bilişimi Anabilim Dalı, Konya
Tel no: 4444243(1609)
E-mail: zcansever@mevlana.edu.tr

Giriş:

Ölçme-değerlendirme eğitimin olmazsa olmaz bir parçası olup öğrenme ve öğretme döngüsünün önemli bir basamağını oluşturmaktadır. Tıp eğitiminde çoktan seçmeli sorular hem mezuniyet öncesi hem de mezuniyet sonrası dönemde yüksek kapsam geçerliliği olması, objektifliği, uygulama ve analizinin kolaylığı nedeniyle sık tercih edilir.

Bloom eğitsel etkinlikleri bilişsel, psikomotor ve duyuşsal olmak üzere üç temel alana ayırmıştır: 2001 yılında revize edilen Bloom taksonomisinde (BT) bilişsel alanda yer alan zihinsel etkinlikler alt basamaktan üst basamağa doğru giderek karmaşılaşan 6 düzeyde incelenir. Bu düzeyler sırasıyla; hatırlamak, anlamak, uygulamak, analiz etmek, değerlendirmek ve yaratmak düzeyleridir. Analiz etmek, değerlendirmek ve yaratmak gibi yüksek bilişsel düzeyi test eden soruların hazırlanması için eğitim ve tecrübe gerekir [1]. Öğretme, öğrenme ve ölçme-değerlendirme üzerindeki olumlu etkisi nedeniyle soruların, sınav öncesi gözden geçirilmesi ve sonrasında analiz edilmesi önemlidir.

Genellikle çoktan seçmeli soruların(ÇSS) kavrama ya da uygulamadan daha ziyade bilginin hatırlanmasına yönelik hazırlandığı

görülmektedir[2]. Miller Piramidinde belirtildiği gibi “bilir” ve “nasılı bilir” basamaklarında kuramsal bilgi temsil edildiğinden ÇSS lar bu basamakların sınanmasında uygun görünmektedir[3].

Öğrenme eksikliği ve başarısızlığa neden olan faktörlerin belirlenip iyileştirici adımların atılmasını sağlayan çalışmalara ışık tutması ve mezun olacakların en iyi şekilde değerlendirilmesini sağlamada sınamaya yöntemleri çok önemlidir [4]. Bu nedenle ölçme-değerlendirme yöntemlerinin çeşitlendirilmesi gerekir. En sık kullanılan ÇSS sınavlarında da, soruların hatasız hazırlanması ve kaliteli olması beklenir.

Çoktan seçmeli soru hazırlanmasında üç önemli nokta üzerinde durulmaktadır: a) öğrenim hedeflerinin yazılması, (b) her hedef için öğrenme düzeylerinin belirlenmesi ve (c) öğrenmeyi test etmede etkili çoktan seçmeli sorular hazırlanması. İdeal çoktan seçmeli sorular hazırlamak ustalık gerektirir [5]. Tıp fakültelerinde eğitimcilerin eğitim becerilerini geliştirmeye yönelik programlar uygulanmakta [6], bu programlar içerisinde ÇSS hazırlama teknikleri ile ilgili eğitimler de verilmektedir [7].

Bu çalışmada Atatürk Üniversitesi Tıp Fakültesi'nde sınavlarda sorulmuş çoktan seçmeli soruların analizi yapıldı (zorluk ve ayırıcılık indeksi) ve Bloom taksonomisine göre sınıflandırılması amaçlandı.

Gereç-Yöntem

Fakültemizde test otomasyonu sistemi olan Sinavmatik® 2010 yılından beri (<http://www.pilotltd.com/>) kullanılmakta olup soru bankasında yaklaşık 35.000 adet soru bulunmaktadır. Üniversitemizde eğitim dili Türkçe ve İngilizce olan 2 ayrı tıp fakültesi bulunmaktadır. 2012-2013 yılı içinde fakültemizin ilk 3 sınıfında uygulanan kurul ve final sınavlarından rastgele yöntemle on ayrı sınavdan seçilen soruların analizi yapıldı. Tıp

eğitimi anabilim dalı öğretim üyelerince eğitici eğitimi alanlar ve almayanların listesi çıkarıldı. BT'ye göre soruların düzeyleri belirlendi ve madde analizi yapıldı. Bağımlı değişken olarak soru kalitesinde “çıkarılmalı”, “geliştirilmeli”, “iyi” ve “çok iyi” kategorileri ile soruların zorluk ve ayırıcılık indeksleri yanında Modifiye BT düzeyleri (hatırlamak, anlamak, uygulamak, analiz etmek, değerlendirmek ve yaratmak yer almaktaydı. Bağımsız değişken olarak ise cinsiyet (K,E), sınıflar (1.,2.ve 3.), eğitim programının dili (Türkçe ve İngilizce), ünvanlar (Yrd. Doç., Doç.,Prof.), eğitici eğitimi alma ve almama durumu yer almaktaydı.

Soru bankası yazılımı ayırıcılık, güçlük ve kalite analizlerini aşağıda tanımlanan kriterleri göre yapmaktadır.

Ayırıcılık: Maddenin ayırıcılık indeksi, -1 ile +1 arasındaki bu sayı, maddenin yüksek puanlı ve düşük puanlı öğrencileri ayırmada ne derece etkili olduğunu ortaya çıkarır. İyi ayırabilen bir madde için bu değer +0,4'ten büyük olması gerektiği kabul edilir. Negatif ayırıcılıklı sorular ise görevini yapamamış kabul edilir ve testlerde kullanılmamalıdır.

Zorluk: Maddenin güçlük düzeyi, 0 ile +1 arasındaki bu sayı, maddeye doğru yanıt verenlerin oranını verir. Bu oran zor sorularda 0'a, kolay sorularda ise +1'e yakın çıkar. Maddenin testte kullanım amacına göre değişmekle birlikte, güçlük düzeyi 0,2'den küçük ve 0,8'den büyük maddelerin testlerde kullanılmaması gerektiği kabul edilir.

Kalite: Maddenin kalite yorumu, sistem, maddenin ayırıcılık indeksi ve güçlük düzeyine bakarak kullanılabilirliğine ilişkin bir kalite yorumu yapar. Bu yorum aşağıdaki dört seviyeden birini alır: 3 – Çok iyi madde: Ayırıcılık indeksi +0,4'ten büyük , güçlük düzeyi de 0,15 ile 0,85 arasında. 2 – İyi madde: Ayırıcılık indeksi +0,3 ile +0,4 arasında, güçlük düzeyi de 0,15 ile 0,85 arasında. 1 – Düzeltilebilir: Ayırıcılık indeksi 0,2'den büyük ancak güçlük düzeyi 0,15 – 0,85 aralığının

dışında kalanlar. 0 – Kötü madde: Testlerde kullanılmamalı. Ayrıcılık indeksi +0,2'den küçük. Veriler, sayı, yüzde, ortalama ve standart sapma olarak verildi. Veriler IBM SPSS 20.0 bilgisayar istatistik programına girilerek analiz edildi. Sürekli verilerin normallik dağılımı Kolmogorov Smirnov ile test edildi. Verilerin analizinde Ki-kare, bağımsız gruplarda t testi ve tek yönlü ANOVA kullanıldı. Çoklu karşılaştırma Post Hoc LSD testi kullanıldı. İstatistik önemlilik düzeyi $p < 0,05$ alındı.

Bulgular

Araştırmada toplam 945 soru incelenmiş olup, bunların 779 tanesi (% 82,4) erkek öğretim üyeleri tarafından hazırlanmıştı. Soruların 640 tanesi (%67,7) Türkçe eğitim programından olup bunların 399'u (%42,2) birinci sınıf sınavlarına aitti. İncelenen 26 farklı anabilim dalından 135 öğretim üyesinin hazırladığı soruların 458 (%48,5) tanesi profesör unvanına sahip öğretim üyelerine aitti. Eğitici eğitimi alan öğretim üyeleri tarafından hazırlanan soru sayısı 783 (%82,9) idi (Tablo 1).

Özellik		n	%
Öğretim üyesi cinsiyet	Erkek	779	82,4
	Kadın	166	17,6
Eğitim dili	Programının Türkçe	640	67,7
	İngilizce	305	32,3
Sınıf	1.sınıf	399	42,2
	2.sınıf	193	20,4
	3.sınıf	353	37,4
Unvan	Prof.Dr.	458	48,5
	Doç. Dr.	263	27,8
	Yrd. Doç. Dr.	224	23,7
Eğitici eğitimi	Almış	783	82,9
	Almamış	162	17,1
Sınavlar	1	34	3,6
	2	231	24,4
	3	116	12,3
	4	59	6,2
	5	56	5,9
	6	66	7,0
	7	44	4,7
	8	171	18,1
	9	82	8,7
	10	86	9,1
Total		945	100,0

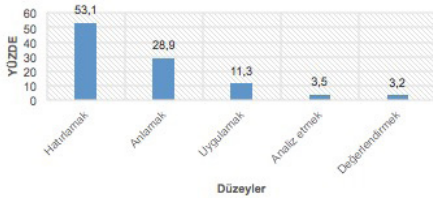
Soruların zorluk ve ayrıcılık indeksleri incelendiğinde 274 (%29) sorunun testten çıkarılması, 255 (%27,0) sorunun geliştirilmesi gerektiği bulundu (Tablo 2). Soruların ortalama zorluk ve ayrıcılık indeksleri sırasıyla $0,64\pm 0,25$ ve $0,29\pm 0,18$ bulundu.

Tablo 2. Soruların zorluk ve ayrıcılık indekslerine göre sınıflandırılması

	n	%
Çıkarılmalı	274	29,0
Geliştirilmeli	255	27,0
İyi	164	17,4
Çok İyi	252	26,7
Total	945	100,0

Soruların Modifiye Bloom taksonomisine göre düzeylerinin dağılımı incelendiğinde 502 (%53,1) soru hatırlamak, 273 (%28,9) soru anlamak, 107 (%11,3) soru uygulamak, 33 (%3,5) soru analiz etmek ve 30 (%3,2) sorunun ise değerlendirme düzeyinde olduğu bulundu (Şekil 1).

Şekil.1 Soruların Bloom Taksonomisine Göre Dağılımı



Soruları hazırlayan öğretim üyelerinin cinsiyeti,

unvanı ve eğitici eğitimi alma durumu, eğitim programının dili ve sınıflara göre soru kalitesinin dağılımı incelendiğinde erkek ve kadın öğretim üyelerinin hazırladıkları soruların “çok iyi” dağılımı sırasıyla %28,8 ve %16,9 bulundu ($p=0,001$). Öğretim üyelerinin unvanları ve eğitici eğitimi alıp almamaları karşılaştırıldığında istatistiksel olarak fark bulunmadı ($p>0,05$). Eğitim programı dili İngilizce ve Türkçede “çok iyi” madde dağılımı sırasıyla %37,7 ve %21,4 bulundu ($p<0,001$). 1., 2., ve 3. sınıflara göre “çok iyi” madde dağılımı sırasıyla %29,1 ve %36,8 ve %18,4 bulundu ($p<0,001$) (Tablo 3).

Soruları hazırlayan öğretim üyelerinin cinsiyeti, unvanı ve eğitici eğitimi alma durumu, eğitim programının dili ve sınıflara göre zorluk ve ayrıcılık indeksleri incelendi. Erkek ve kadın öğretim üyelerinin hazırladıkları soruların ayrıcılık indeks ortalamaları sırasıyla $0,30\pm 0,18$ ve $0,25\pm 0,16$ olup fark istatistiksel olarak anlamlı bulundu ($t=3,13$; $p=0,002$).

Doçent unvanına sahip öğretim üyelerinin hazırladıkları soruların zorluk ve ayrıcılık indeksleri profesör ve yardımcı doçent unvanına sahip öğretim üyelerinden daha ayrıcılık ve kolay bulundu ($p<0,05$).

Eğitim programı dili Türkçe ve İngilizcede hazırladıkları soruların ayrıcılık indeks ortalamaları sırasıyla $0,27\pm 0,16$ ve $0,33\pm 0,19$ olup fark istatistiksel olarak anlamlı bulundu ($t=4,87$; $p<0,001$).

Birinci sınıf sınavlarının zorluk indeksi diğer sınıflardan istatistiksel olarak daha düşük bulundu ($p<0,05$), ikinci sınıfın ayrıcılık indeksi ise diğer sınıflardan daha yüksek bulundu ($p<0,05$) (Tablo 4).

Tablo 3. Soruları hazırlayan öğretim üyelerinin bazı değişkenlere göre soru kalitesinin dağılımı

		Zorluk ve ayrıcılık indekslerine göre sınıflandırılması								χ^2	<i>p</i>
		Çıkarılmalı		Geliştirilmeli		İyi		Çok İyi			
		n	%	n	%	n	%	n	%		
Cinsiyet	Erkek	211	27,1	217	27,9	127	16,3	224	28,8	16,9	0,001
	Kadın	63	38,0	38	22,9	37	22,3	28	16,9		
Eğitim Programının dili	Türkçe	198	30,9	188	29,4	117	18,3	137	21,4	28,3	0,000
	İngilizce dili	76	24,9	67	22,0	47	15,4	115	37,7		
Sınıf	1.sınıf	121	30,3	86	21,6	76	19,0	116	29,1	41,1	0,000
	2.sınıf	33	17,1	53	27,5	36	18,7	71	36,8		
	3. sınıf	120	34,0	116	32,9	52	14,7	65	18,4		
Unvan	Prof.Dr.	142	31,0	124	27,1	78	17,0	114	24,9	8,7	0,194
	Doç. Dr.	66	25,1	70	26,6	41	15,6	86	32,7		
	Yrd.	66	29,5	61	27,2	45	20,1	52	23,2		
	Doc.Dr.										
Eğitici eğitimi	Aldı	226	28,9	222	28,4	129	16,5	206	26,3	5,5	0,142
	Almadı	48	29,6	33	20,4	35	21,6	46	28,4		

Tablo 4. Soruları hazırlayan öğretim üyelerinin bazı değişkenlere göre soruların zorluk ve ayırt edicilik indeksleri.

		Zorluk			Ayırtıcılık		
		Mean	SD	P	Mean	SD	P
Cinsiyet	Erkek	0,64	0,25	0,76	0,30	0,18	0,002
	Kadın	0,63	0,23		0,25	0,16	
Eğitim dili	Programının Türkçe	0,63	0,25	0,5	0,27	0,16	0,000
	İngilizce	0,65	0,25		0,33	0,19	
Sınıf*	1.sınıf	0,59 ^{ab}	0,25	0,000	0,29 ^a	0,20	0,000
	2.sınıf	0,69 ^a	0,23		0,34 ^{ab}	0,15	
	3. sınıf	0,66 ^b	0,24		0,26 ^b	0,14	
Unvan*	Prof.	0,62	0,25	0,000	0,28	0,17	0,001
	Doç.	0,69 ^a	0,23		0,32 ^a	0,19	
	Yrd. Doç.	0,61	0,26		0,28	0,16	

15

Eğitici eğitimi	Aldı	0,64	0,25	0,039	0,29	0,18	0,569
	Almadı	0,60	0,25		0,30	0,16	

SD=Standard Deviation

Karşılaştırma ANOVA ile yapıldı ve çoklu karşılaştırmalar için post hoc LSD kullanıldı.

a ve b için $p < 0,05$

Soruların Modifiye Bloom Taksonomisi düzeyine göre zorluk ve ayıricılık indeksleri incelendi. Düzeylerin ortalama zorluk indeksleri arasında istatistiksel olarak fark bulunmazken ($p>0,05$) ayıricılık indeks ortalamaları arasında istatistiksel olarak önemli fark olduğu bulundu ($p=0,021$). Yapılan post hoc LSD çoklu

karşılaştırma analizinde hatırlamak ve anlamak ile anlamak ve değerlendirmek grupları arasında istatistiksel olarak önemli derecede fark bulundu ($p<0,05$) (Tablo 5).

Sınavlarda sorulan bazı örnek soruların Modifiye BT'ye göre düzeyleri tablo 6'da verilmiştir.

Tablo 5. Soruların Modifiye Bloom Taksonomisi düzeyine göre zorluk ve ayıricılık indeksleri.

Düzyey	Zorluk			Ayıricılık		
	Ortalama	SS	p	Ortalama	SS	p
Hatırlamak (n=502)	0,65	0,25		0,28 ^a	0,18	
Anlamak (n=273)	0,63	0,24		0,32 ^{ab}	0,17	
Uygulamak (n=107)	0,62	0,25	0,095	0,29	0,16	0,021
Analiz etmek (n=33)	0,55	0,25		0,26	0,15	
Değerlendirmek (n=30)	0,67	0,25		0,25 ^b	0,15	

SS: Standart Sapma, a için $p=0,004$, b için $p=0,04$

*Karşılaştırma ANOVA ile yapıldı ve çoklu karşılaştırmalar için LSD kullanıldı.

Sınavlarda sorulan bazı örnek soruların Moffiye BT'ye göre düzeyleri tablo 6'da yerilmiştir.

Tablo 6. Sınavlarda sorulan bazı örnek soruların <u>Moffiye BT'ye göre düzeyleri.</u>	
<u>Bloom Düzeyi</u>	<u>Örnek</u>
Hatırlamak	Aşağıdaki oluşumlardan hangisi ön kolun <u>medialindeki</u> kemik olan <u>Ulna'da</u> bulunur? A) <u>Olecranon</u> B) <u>Fossa olecrani</u> C) <u>Incisura ulnaris</u> D) <u>Epicondylus medialis</u> E) <u>Capitulum</u>
Anlamak	Düz kaslarda kalsiyum, <u>kontraksiyonu</u> başlatıcı etkinliğini aşağıdakilerden hangisine <u>bağlanarak</u> yapar ? A) <u>Troponin C</u> B) Aktin C) <u>Miyozin</u> D) <u>Kalmodulin</u> E) <u>Tropomiyozin</u>
Uygulamak	25 mm/s hızla çekilen bir elektrokardiyografide eşit aralıklarla gelen iki QRS kompleksi arasında 25 küçük kare bulunmaktadır. Bu kişinin

	<p>dakikada nabız sayısı kaçtır?</p> <p>A) 25</p> <p>B) 40</p> <p>C) 60</p> <p>D) 80</p> <p>E) 100</p>
Analiz Etmek	<p>Belirgin skuam, papillomatöz ve akantozun görüldüğü hastalıktır?</p> <p>A) Liken Planus</p> <p>B) Eritema Multiforme</p> <p>C) Psöriazis</p> <p>D) Liken Simpleks Kronikus</p> <p>E) Hiçbiri</p>
Değerlendirmek	<p>45 yaşında erkek hasta, 3 gün boyunca ateş, nefes almada zorluk ve balgamlı öksürük şikayeti var. Fizik muayenede ateş 38.5 ° C'dir. Akciğer grafisinde sağ plevral efüzyon tespit ediliyor ve sağ torasentez yapılıyor. Elde edilen sıvının incelenmesinde bulanık bir görünüme sahip nötrofiller % 98 oranında ve mikrolitre başına 17. 000 olarak belirleniyor. Mevcut plevral inflamasyonu aşağıdaki terimlerin hangisi en iyi açıklar?</p> <p>A) Seröz inflamasyon</p> <p>B) Pürülan inflamasyon</p> <p>C) Fibrinöz inflamasyon</p> <p>D) Kronik inflamasyon</p> <p>E) Granümatöz inflamasyon</p>

Tartışma

Madde analizi sınavlardan sonra yapılan ve sınavların sorularının kalitesini belirleyen basit fakat etkili bir yöntemdir. İdeal bir sınavda zorluk indeksi %31-60 arasında, ayırıcılık indeksi de %25ten fazla olmalıdır[8]. Bizim sınavlarımızda da ayırıcılık indeksi ortalaması 0.25'in üzerinde bulundu.

Sınavlarımızda erkeklerle bayan öğretim üyelerinin sorularında zorluk yönünden fark yokken erkek öğretim üyelerinin sorularının ayırıcılık indeks değerleri daha yüksek bulundu. Fakültemizde İngilizce ve Türkçe tıp soruları bakımından İngilizce tıp sorularının ayırıcılık indeks değerleri daha yüksek bulundu.

Ayrıca doçent düzeyindeki öğretim üyelerinin soruları daha ayırıcı ve zorluk indeks değerleri daha yüksekti. Soru sorma bir sanattır. Aceleyle hazırlanan sorular soru kalitesini etkilemekte ve belirsizliklere, yanlış anlaşılmalara neden olabilmektedir. Hatta bazı kurumlarda soruların gözden geçirilmesi ile ilgili bölümler oluşturulduğundan ve soruların içeriği, teknik özelliği, gramer yapısı, şıkların içeriği editöryel değerlendirilmesinin yapıldığından bahsedilmektedir[9]. Bu süreç tıpkı atın koşu öncesi bakımının yapılmasına benzetilmiş ve "Vetting" kelimesi ile ifade edilmiştir. Bizim sınavlarımızda da soruların gözden geçirilmesi tıp eğitimi anabilim dalı, dekanlık ve eğitim koordinatörünün işbirliği ile yapılmaktadır. Ayrıca İngilizce yönünden de sorular tek tek gözden geçirilmektedir.

Bir çalışmada Mini essay sorulara göre ÇSS ların daha üstün olduğu belirtilmekte ve öğretim üyelerinin soru hazırlamada hatalarının olmaması için eğitim almaları gerektiğinden bahsedilmektedir[10]. Bizim çalışmamızda ise eğitici eğitimi alanların soruları daha zor idi ve ayırıcılık indeksleri arasında ise fark bulunmamaktaydı. Eğitici eğitimi almış olmak soru kalitesini etkilemiyordu. Bulgularımız eğitici eğitiminin bir kez alınmasının yeterli olmadığını ve soru kalitesi ile ilgili eğitimlere ağırlık verilmesi gerektiğini düşündürmektedir.

Test maddelerinin bilişsel öğrenme düzeylerinin belirlenmesi gerektiğinden bahsedilmektedir[11]. Sınavlarımızın değerlendirme sonrası madde analizleri yapılmış ve öğrenme düzeyleri de incelenmişti. Soruların büyük kısmı sırasıyla hatırlamak, anlamak ve uygulamak düzeylerinden oluşmuştu. En az kısmı ise analiz etmek düzeyindeki sorulardan oluşmaktaydı. Öğrencilerimizin farklı farklı öğrenme stilleri mevcuttur. Bilişsel alandaki hatırlamak (bilgi) düzeyini sınamaya yönelik sorular genellikle ezbere ve hatırlamaya dayalı sorulardır. Anlamak (kavrama) düzeyindeki sorular; hesaplama, gruplama ve seçme gibi alanları, uygulamak; yorumlama, kullanma ve çözme, analiz etmek; ayırt etme, ölçme, sınıflandırma, değerlendirme düzeyi karar verme, tayin etme, tahmin etme, eleştirme yaratmak (sentez) ise; tasarlama, geliştirme ve açıklama gibi konuları içermektedir[12]. İlk üç sınıfta çoğunlukla temel bilim dersleri verildiğinden dolayı daha çok hatırlama ve anlamaya yönelik soru sorulmuş olabilir. Fakat yine de üst bilişsel düzey sorularına da yer verilmesi gerektiği görülmektedir.

Sonuç

Çoktan seçmeli sorular hangi BT düzeyinde olursa olsun hala ölçme –değerlendirmede sık tercih edilen bir yöntemdir. Ölçme ve değerlendirme eğitimin önemli bir parçasıdır. Sınavlardan sonra madde analizlerinin yapılarak öğretim üyelerine bildirilmesi ile soru kalitesinin artırılması sağlanabilir. Eğiticilerin soru hazırlama için eğitim almaları önemlidir.

Çalışmanın Kısıtlılığı

Soruların düzeyleri konusunda görüş ayrılığı olması çalışmanın kısıtlılığını oluşturmaktadır. Ayrıca düzeyler bilgi boyutunda olgusal, kavramsal, işlemsel ve bilişüstü bilgi kısımlardan oluşmaktadır. Alt boyutlarda dahi görüş ayrılığı olabilmektedir. Burada revize edilen sınıflamanın esnek bir yapı içerdiği belirtilmektedir [13]. Çalışmada bu şekilde

bilgi boyutunda olgusal, kavramsal, işlemsel ve bilişüstü bilgi alt boyutlarıyla düzeylendirme yapılmamıştır.

KAYNAKLAR

1. Crowe, A., C. Dirks, and M.P. Wenderoth,

Biology in bloom: implementing Bloom's Taxonomy to enhance student learning in biology. CBE Life Sci Educ, 2008. 7(4): p. 368-81.

2. Harden, R.M., Laidlaw J.M, Essential Skills For A Medical Teacher. Written and Computer Based Assesment. 2012.

3. Dent J.A., Harden.R.M., A Practical Guide for Medical Teachers. 2013: p. 307.

4. Haladyna, T.M. and G.A. Kramer, The validity of subscores for a credentialing test. Eval Health Prof, 2004. 27(4): p. 349-68.

5. Collins, J., Education techniques for lifelong learning: writing multiple-choice questions for continuing medical education activities and self-assessment modules. Radiographics, 2006. 26(2): p. 543-51.

6. Cansever, Z., et al., What do trainers think about trainer training courses? J Pak Med Assoc, 2014. 64(5): p. 491-5.

7. Musal B., Eğitici Gelişim Programları: Dokuz Eylül Üniversitesi Tıp Fakültesi Deneyimi Ege Eğitim Dergisi, 2003. 3(1): p. 81-88.

8. Gajjar, S., et al., Item and Test Analysis to Identify Quality Multiple Choice Questions (MCQs) from an Assessment of Medical Students of Ahmedabad, Gujarat. Indian J Community Med, 2014. 39(1): p. 17-20.

9. Gopalakrishnan, S. and P.M. Udayshankar, Question vetting: the process to ensure quality in assessment of medical students. J Clin Diagn Res, 2014. 8(9): p. XM01-XM03.

10. Moeen Uz, Z. and A. Badr, Evaluation of mini-essay questions (MEQ) and multiple choice questions (MCQ) as a tool for assessing the cognitive skills of undergraduate students at the Department of Medicine. Int J Health Sci (Qassim), 2011. 5(2 Suppl 1): p. 43-4.

11. Downing, S.M. and T.M. Haladyna, Validity threats: overcoming interference with proposed interpretations of assessment data. Med Educ, 2004. 38(3): p. 327-33.

12. Durak H.İ.,Solakoğlu Z. Tıp Eğitiminin Temelleri. 2012.Sayfa: 69-83.

13. Eğitimde Program Geliştirme (Hedef Belirleme: Bilişsel, Duyuşsal ve Devinişsel Alanlar) Yrd.Doç.Dr Zeynel Kablan ,3.baskı, Ankara ,sayfa 127-159.