

Tıp fakültesi dönem I-II-III'de uygulanan testlerin betimsel açıdan incelenmesi: Mersin Üniversitesi Tıp Fakültesi örneği

Hüseyin Selvi

Mersin Üniversitesi, Tıp Fakültesi, Tıp Eğitimi AD, Mersin

Özet

Amaç; Bu çalışmada, tıp fakültesi dönem bir, iki ve üç öğrencilerine uygulanan sınavlar ve bu sınavlarda kullanılan testler ile maddelerin (soruların) psikometrik özelliklerinin incelenmesi ve ilgili sınavların oluşturulmasında/uygulanmasında rolü olan bireylerde farkındalık oluşturulması amaçlanmıştır. **Yöntem:** Dönem bir, iki ve üç öğrencilerine 2012-2013 ile 2014-2015 eğitim öğretim yılları arasında (son üç eğitim öğretim yılı) uygulanan toplam 69 komite ve final sınavı ile bu sınavlarda kullanılan 6900 madde incelenmiştir. Araştırma kapsamında elde edilen verilerin analizlerine yönelik olarak; madde güçlük değerleri, madde toplam test korelasyonları (çift serili korelasyon), güvenilirlik katsayıları (KR-20), basıklık ve çarpıklık değerleri, maksimum ve minimum toplam puanlar, test puanı ortalamaları ve standart sapma değerleri hesaplanmıştır. **Bulgular:** Elde edilen bulgular; sınavlardan elde edilen puanların iç tutarlık bağlamındaki güvenilirliklerinin 0.64 ile 0.92, ortalama madde güçlük indeksi değerlerinin 0.53 ile 0.78 arasında, madde-toplam test korelasyonu değerlerinin 0.16 ile 0.66 arasında, toplam puan ortalamalarının ise 53 ile 79 arasında yer aldığını göstermiştir. **Sonuç:** Sonuç olarak; elde edilen puanların güvenilir olduğu ancak madde toplam test korelasyonu (madde geçerliği) değerlerinin düşük olduğu ve artırılmasına yönelik bir dizi önlem alınmasının gerekli olduğu sonucuna ulaşılmıştır.

Anahtar sözcükler: Tıp eğitimi; psikometrik özellikler; madde analizi

Investigating descriptive properties for tests which are administered to medical faculty term I, II, III students: Sample of Mersin University Medical Faculty

Abstract

Purpose: In this study, the exams and the items administered to medical students studying in Term I, II & III at Mersin University, Faculty of Medicine were investigated for psychometric properties to build self awareness in those have role in the creation and application of relevant tests. **Method:** For this purpose, total of 69 committee-final exams and 6900 items administered during the academic terms of 2012-2013 and 2014-2015 (last three academic terms) were examined. The analysis for data; item difficulty indices, item total test correlations (i.e. biserial correlation), reliability indices (KR-20), kurtosis and skewness values, maximum and minimum total scores, mean test score values, and standart deviations were computed.

Yazının geliş tarihi: 17.12.2015

Yazının kabul tarihi: 12.02.2016

Yazışma adresi: Mersin Üniversitesi Tıp Fakültesi Tıp Eğitimi Anabilim Dalı, 33343 Mersin

Tel :0324 3610684-89, Dahili: 1055, E-posta : hsyn_selvi@yahoo.com.tr

Results: The findings obtained in this study indicated that the values in the context of internal consistency are between 0.64 and 0.92, item difficulty values are between 0.53 and 0.78, item total test correlations are between 0.16 and 0.66 and the mean of total scores are between 53 and 79. **Conclusions:** As a result; it can be stated that the data obtained in this study have high reliability values, but low item-total test correlation values (i.e. item validity). Therefore, it can be suggested that series of precautions should be taken to improve the low item total test correlation values.

Keywords: Medical education; psychometric properties; item analysis

Giriş

Tıp eğitiminin gerek hekim aday ve ailesi gerekse toplumsal açıdan oldukça pahalı bir eğitim olduğu, diğer taraftan da hekim adaylarına kazandırılmayan bilgi ve becerilerin doğrudan insan sağlığı ve hayatına yönelik uygulamaları içerdiği düşünüldüğünde, adaylar hakkında verilen/verilecek olan kararların isabetli olması konusu büyük önem kazanmaktadır.

Bu isabetlilik ise ancak verilen kararların nitelikli ölçme araçlarından elde edilen güvenilir ve geçerli veriye dayandırılmasıyla sağlanabilir.

Burada bahsedilen güvenilirlik kısaca kullanılan ölçme aracının hatasız ölçme yapabilme yeterliği olarak tanımlanmakta ve geçerlik için bir ön koşul oluşturmaktadır. Geçerlik ise genel olarak ilgili ölçme aracının, ölçülmek istenilen niteliği, başka niteliklere karıştırmadan ölçebilme yeterliği olarak tanımlanmaktadır.^{1, 2, 3}

Tıp fakültelerinde ise özellikle sınır yeterliklerin, ölçülmesine yönelik olarak gerçekleştirilen ölçme işlemlerinde ağırlıklı olarak çoktan seçmeli sorulardan (maddelerden) oluşan testlerin kullanıldığı bilinmektedir. Nitekim Ulusal Tıp Eğitimi Akreditasyon Kurulu (UTEAK) Mezuniyet Öncesi Tıp Eğitimi Ulusal Standartları (MOTES) arasında çoktan seçmeli maddelerden oluşan testlerin kullanılması yer almaktadır.⁴ Bu durumun nedenleri arasında; tıp eğitimiyle kazandırılması hedeflenen pek çok davranışın çoktan seçmeli maddelerle ölçülebilir nitelikte olması, sınıf mevcutlarının fazla olması, optik okuyucular kullanılarak puanlamanın oldukça hızlı yapılabilmesi ve puanlama güvenliğinin garanti altına alınabilmesi,

vb. hususlar sıralanabilir. Ancak bilindiği üzere çoktan seçmeli maddeleri hazırlamak diğer pek çok madde türünü hazırlamaktan çok daha güçtür.⁵ Bu güçlüğü eğitim vermekte olan akademisyenlerin nitelikli madde yazımı konusunda yaşadıkları sorunlar da eklendiğinde, ölçme sürecine karışan hata miktarı radikal biçimde artabilmekte ve verilen kararların isabetliliği bu durumdan olumsuz yönde etkilenebilmektedir. Bu açıdan bakıldığında yapılan sınavların da bir bakıma amacını yitirmeye başladığı söylenebilir.

Tıp fakültelerinde gerçekleştirilen sınavların niteliği ile ilgili yaşanan bu ve benzer sorunlar, sınavlara ve sınavları yapan kuruma olan güveni zedelemekte ve ayrıca eğitimde fırsat eşitliği ve adalet ilkesi açılarından da sorun oluşturmaktadır.

Bu sorunları önlemek amacıyla bilim insanları ve kurumlar (tıp fakülteleri); madde ve test istatistiklerini hesaplamak, madde bankası (soru bankası) oluşturmak, belirli zaman aralıklarında test geliştirme, çoktan seçmeli madde yazımında dikkat edilecek hususlar vb. konularda hizmet içi eğitimler vermek, vb. çeşitli faaliyetler sürdürmektedir.

Bu faaliyetler arasında yer alan madde ve test istatistiklerini hesaplamak; puanların güvenilir olup olmadığı, maddelerin güçlük düzeyleri, seçeneklerin işleyip işlemedikleri, maddelerde değişen madde fonksiyonu olup olmadığı, madde güvenilirlik ve geçerliklerinin sağlanıp sağlanmadığı, maddelerin şansa cevaplanma olasılıkları, anahtar cevap hatası, vb. konular hakkında bilgi vermekte ve testlerden elde edilen puanların güvenilirlik ve geçerliğinin sağlanabilmesi adına çeşitli önlemler alınmasını sağlamaktadır.^{6, 7, 8}

Ancak, gerek burada bahsedilen madde ve test istatistiklerini hesaplamak, gerekse diğer faaliyetleri nitelikli şekilde sürdürebilmek ölçme ve değerlendirme açısından belirli bir uzmanlık gerektirmektedir. Nitekim yapılan alanyazın taraması; tıp fakültelerinde uygulanan sınavların niteliklerinin incelendiği pek çok ulusal çalışmada ölçme ve değerlendirme işlem ve süreçleri ile ilgili kavramların birbirine karıştırıldığı ve yapılan hesaplamalarda hangi yöntemlerin kullanıldığını ya belirtilmediğini ya da hatalı kullanıldığını göstermiştir.^{9, 10, 11, 12}

Konu ile ilgili olarak; Özcan ve akr.⁹ Gaziosmanpaşa Üniversitesi Tıp Fakültesi'nde, 2013 - 2014 Güz öğretim yılında dönem bir, iki ve üç öğrencilerine uygulanan sınavları betimsel açıdan incelemiş ve inceleme kapsamında; anahtar cevap-seçenek dağılımı, madde güçlük indeksi (p_i), madde-toplam test korelasyonu (çift serili korelasyon, r_{cs}) ve çarpıklık katsayısı hesaplanmıştır. Çalışma sonucunda; anahtar cevabın seçeneklere benzer oranlarda dağıldığı, madde ayırt edicilik değerlerinin genel olarak düşük saptandığı, madde güçlük düzeylerinin ağırlıklı olarak 0.35 ve üzeri saptandığı, çarpıklık katsayısının ise 0 ile -1.68 değerleri arasında yer aldığı bulunmuştur.

Topal ve ark.⁴⁰ Pamukkale Üniversitesi Tıp Fakültesinde 2006-2007 öğretim yılında dönem I öğrencilerine uygulanan sınavları betimsel açıdan incelemişler ve çalışmada; madde güçlük ve ayırt edicilik değerleri ile iç-tutarlık katsayısı (KR-20) değerlerini hesaplanmaya çalışmışlardır.

Eskiocak ve ark.¹¹ ise, Trakya Üniversitesi Tıp Fakültesi, Biyokimya Anabilim Dalında Dönem bir ve iki öğrencilerine 1999-2004 yılları arasında uygulanan sınavları betimsel açıdan incelemişler ve inceleme kapsamında madde güçlük indeksi (p), madde-toplam test korelasyonu, güvenilirlik katsayısı ve çarpıklık katsayısını hesaplamışlardır. Çalışmada hangi güvenilirlik hesaplama yönteminin kullanıldığı belirtilmemekle birlikte elde edilen verilerin hemen hepsinde güvenilirlik değerlerinin düşük

olduğu, madde güçlük indeksi değerlerinin ağırlıklı olarak 0.50 ve üzerinde olduğu, ayırt edicilik değerlerinin ise ağırlıklı olarak 0.26 ve üzerinde olduğu bulunmuştur.¹¹

Mersin Üniversitesi Tıp Fakültesi'nde ise gerçekleştirilen sınavların ve bu sınavlarda kullanılan maddelerin test ve madde istatistiklerinin (psikometrik özelliklerinin) incelendiği bir çalışmaya rastlanılmamıştır. Bu nedenle bu çalışmada; Mersin Üniversitesi Tıp Fakültesi dönem bir, iki ve üç öğrencilerine son üç eğitim-öğretim yılı içinde uygulanan sınavlar ve bu sınavlarda kullanılan maddelerin psikometrik özelliklerinin incelenmesi ve sınavların oluşturulmasına paydaş rolü olan bireylerde farkındalık oluşturulması amaçlanmıştır.

Gereç ve Yöntem

Bu çalışmada, Mersin Üniversitesi Tıp Fakültesi Dönem bir, iki ve üç öğrencilerine son üç yıl içinde uygulanan teorik sınavların ve bu sınavlarda kullanılan maddelerin psikometrik özelliklerinin incelenmesi amaçlandığından, çalışmanın betimsel tipte bir çalışma olduğu söylenebilir. Araştırma için Mersin Üniversitesi, Sosyal ve Fen Bilimleri Etik Kurulu'ndan 26.10.2015 tarih ve 2015/22 sayılı yazıyla etik kurul onayı alınmıştır.

Bu araştırma; 2012-2013 ile 2014-2015 eğitim öğretim yılları arasında Mersin Üniversitesi Tıp Fakültesi, dönem bir, iki ve üç öğrencilerine uygulanan ve çoktan seçmeli 100'er maddeden oluşan kurul ve final sınavlarından elde edilen veriler üzerinden gerçekleştirilmiştir. Araştırmanın birey evrenini ilgili eğitim öğretim döneminde ilgili sınıflarda öğrenim görmekte olan toplam 2075 öğrenci oluşturmaktadır. Sınav ve madde evrenini ise ilgili yıllar arasında gerçekleştirilen 69 sınav ve bu sınavlarda kullanılan toplam 6900 madde oluşturmaktadır. Çalışma kapsamında bireyler, sınavlar ve maddelerin tamamına erişilmesi hedeflenmiş ve bu nedenle bir örnekleme işlemi yapılmamıştır. Araştırma verilerinin elde edildiği gruba ilişkin sayısal değerler Tablo 1'de yer almaktadır.¹³

Araştırma kapsamında öncelikli olarak; Tablo 1'den de anlaşılacağı üzere dönem bir ve iki için (1. ve 2. Sınıf) toplamda 1414 öğrenciye ait 30 komite ve 12 final sınavından elde edilen cevap kâğıtları (optik form) okunmuş ve buradan madde-seçenek matrisi elde edilmiştir. Bu

işlem daha sonra Dönem III için toplamda 661 öğrenci ve 21 komite ile 6 final sınavı için de tekrarlanmıştır. Bir sonraki aşamada ise elde edilen madde-seçenek matrisi, 1-0 matrisine dönüştürülmüş ve analizler bu matrisler üzerinden gerçekleştirilmiştir.

Tablo 1: Verilerin elde edildiği gruba ilişkin sayısal değerler

	Değişken	n	%	Toplam
2012-2013	Dönem I	207	33,38	620
	Dönem II	221	35,64	
	Dönem III	192	30,96	
2013-2014	Dönem I	243	34,12	712
	Dönem II	226	31,74	
	Dönem III	243	34,12	
2014-2015	Dönem I	273	36,74	743
	Dönem II	244	32,83	
	Dönem III	226	30,41	

Araştırma kapsamında elde edilen verilerin analizlerine yönelik olarak; madde güçlük indeksi değerleri, madde toplam test korelasyonları (çift serili korelasyon), güvenilirlik katsayıları (KR-20), basıklık ve çarpıklık değerleri, maksimum ve minimum toplam puanlar, test puanı ortalamaları ve standart sapma değerleri hesaplanmıştır.

Verilerin analizinde kullanılan; Madde güçlük indeksi (p_j), maddeyi doğru cevaplayan bireylerin toplam birey sayısına bölünmesiyle elde edilmektedir ve bu değer 0.00 ile 1.00 arasında değişim göstermektedir. Madde güçlük indeksi değerlerinin; 0.00-0.19 değerleri arasında yer alması, bu maddelerin 'çok zor', 0.20-.039 değerleri arasında yer alması 'zor', 0.40-0.59 değerleri arasında yer alması 'orta güçlükte', 0.60-0.79 değerleri arasında yer alması 'kolay', 0.80-1.00 değerleri arasında yer alması ise 'çok kolay' olduğu şeklinde

yorumlanmasını bir açıdan gerekli kılmaktadır.¹⁴

Madde toplam test korelasyonları (r_{cs}) ise, özellikle başarı testleri için maddenin bilen ve bilmeyen bireyleri birbirinden ne kadar iyi ayırt edebildiğinin bir göstergesi olarak kabul edilmektedir. Diğer ifadeyle madde toplam test korelasyonları, madde ayırt ediciliğini yansıtmaktadır ve elde edilen değerler korelasyon değerleri olduğundan -1.00 ile 1.00 arasında yer almaktadır. Alanyazında 0.30 değeri ise madde ayırt ediciliği için alt sınır kabul edilmektedir.⁵

Güvenirlilik katsayısı (KR-20), ilgili testin güvenilirliğini saptamak amacıyla hesaplanan bir katsayı olup, testin hatasız ölçme yapabilme gücünü yansıtmaktadır. Burada hesaplanan KR-20 güvenilirlik katsayısı 0.00 ile 1.00 arasında değerler almaktadır ve alanyazında bu değer 0.70'ten büyük olmasının gerektiği, 0.60

değerinin ise güvenilirlik için kabul edilebilir alt sınır olduğu ifade edilmektedir.¹

Basıklık ve çarpıklık değerleri ise, elde edilen verinin bir bütün olarak dağılımının değerlendirilmesinde kullanılmaktadır. Basıklık değerleri dağılımın sivriliği hakkında bilgi vermektedir. Pozitif bir basıklık değeri sivri, negatif bir basıklık değeri düz bir dağılım göstergesi olarak kabul edilmektedir. Çarpıklık değerleri ise dağılımın simetriklik durumu hakkında bilgi vermektedir. Negatif bir çarpıklık değeri sola, pozitif bir çarpıklık değeri ise sağa çarpık bir dağılıma işaret etmektedir. Çarpıklık değerinin -1.00 ile 1.00 arasında, basıklık değerinin ise -2.00 ile 2.00 arasında yer alması dağılımın normalliğine ilişkin göstergeler olarak kullanılabilir.¹⁵

Diğer taraftan aritmetik ortalama, standart sapma ve ortanca değerleri kullanılarak da dağılım yapısı hakkında bilgi edinilebilmektedir. Sağa çarpık bir dağılımda aritmetik ortalamanın medyan ve moddan büyük olması (aritmetik ortalama > medyan > mod), sola çarpık dağılımda ise modun medyan ve aritmetik ortalamadan büyük olması (aritmetik ortalama < medyan < mod) beklenmektedir.¹⁶ Buradan basıklık ve çarpıklık değerlerinin bir başarı sınavına yönelik hesaplandığı ve minimum ders geçme puanının (kesme puanı) aritmetik ortalamaya yakın bir değer aldığı düşünüldüğünde; sola çarpık bir dağılımın, pek çok bireyin ilgili sınavdan başarılı olduğunu gösterdiği kabul edilebilir.

Bulgular

Tıp Fakültesi dönem bir, iki ve üç öğrencilerine son üç yıl içinde uygulanan teorik sınavlar ve bu sınavlarda kullanılan testler ile maddelerin psikometrik özelliklerinin incelenmesi sonucunda elde edilen bulgular Tablo 2, Tablo 3 ve Tablo 4'de verilmiştir.

2012-2013 eğitim öğretim yılı için; dönem bir, iki ve üçe yönelik elde edilen toplam puan ortalamalarının 54 ile 79 puan arasında değiştiği ve genel olarak bir üst sınıfa geçmek için yeterli olan minimum puana (kesme puanına (60 puan)) yakın

değerler aldığı, hemen her testten elde edilen puan dağılımının sola çarpık olduğu, gerçekleştirilen toplam 23 sınavın iç tutarlık bağlamındaki güvenilirliklerinin 0.64 ile 0.93 arasında değiştiği ve bu sınavlardan 3'ü hariç diğerlerinin güvenilirliklerinin 0.70'den yüksek olduğu, ortalama madde güçlük indeksi değerlerinin 0.53 ile 0.78 arasında değiştiği ve bu değerlerden bu maddelerin 'orta güçlükte' ve 'kolay' olarak sınıflandırılabilceği, ortalama madde toplam test korelasyonu değerlerinin ise 0.16 ile 0.46 arasında olduğu ve %61'inin kabul edilebilir alt sınırın altında yer aldığı görülmüştür (Tablo 2).

2013-2014 eğitim öğretim yılı için; dönem bir, iki ve üçe yönelik elde edilen toplam puan ortalamalarının 57 ile 74 puan arasında değiştiği ve genel olarak bir üst sınıfa geçmek için yeterli olan minimum puana yakın değerler aldığı, hemen her testten elde edilen puan dağılımının sola çarpık olduğu, gerçekleştirilen toplam 23 sınavın iç tutarlık bağlamındaki güvenilirliklerinin 0.70 ile 0.92 arasında değiştiği, ortalama madde güçlük indeksi değerlerinin 0.57 ile 0.74 arasında değiştiği ve bu değerlerden bu maddelerin 'orta güçlükte' ve 'kolay' olarak sınıflandırılabilceği, ortalama madde toplam test korelasyonu değerlerinin ise 0.18 ile 0.56 arasında olduğu ve %43'ünün kabul edilebilir alt sınırın altında yer aldığı görülmüştür (Tablo 3).

2014-2015 eğitim öğretim yılı için; dönem bir, iki ve üçe yönelik elde edilen toplam puan ortalamalarının 54 ile 76 puan arasında değiştiği ve genel olarak bir üst sınıfa geçmek için yeterli olan minimum puana yakın değerler aldığı, hemen her testten elde edilen puan dağılımının sola çarpık olduğu, gerçekleştirilen toplam 23 sınavın iç tutarlık bağlamındaki güvenilirliklerinin 0.75 ile 0.91 arasında değiştiği, ortalama madde güçlük indeksi değerlerinin 0.54 ile 0.76 arasında değiştiği ve bu değerlerden bu maddelerin 'orta güçlükte' ve 'kolay' olarak sınıflandırılabilceği, ortalama madde toplam test korelasyonu değerlerinin ise 0.25 ile 0.46 arasında olduğu ve %26'sının kabul edilebilir alt sınırın altında yer aldığı görülmüştür (Tablo 4).

Tablo 2: 2012-2013 eğitim öğretim yılında kullanılan testler ve maddelere yönelik elde edilen psikometrik değerler

		2012-2013 Eğitim Öğretim Yılı								
		Aritm. Ort.	Std. Sap	Çarpıklık	Basıklık	KR-20	Maks.	Min.	Ort. p_j^*	Ort. r_{cs}^{**}
Dönem I	Komite 1	79	8.9	-1.6	3.54	0.84	91	44	0.78	0.38
	Komite 2	61	8.9	-0.04	-0.21	0.78	81	41	0.61	0.24
	Komite 3	55	11.1	-0.3	-0.41	0.87	89	38	0.68	0.33
	Komite 4	64	1.1	-0.02	-0.61	0.83	86	42	0.64	0.29
	Komite 5	59	11.9	-0.04	0.65	0.87	90	28	0.59	0.33
	Final 1	66	6.77	0.02	-0.3	0.65	80	52	0.66	0.17
Dönem II	Final 2	67	1.4	-1.6	5.85	0.86	91	22	0.67	0.37
	Komite 1	63	15.8	-0.72	0.86	0.93	95	14	0.63	0.46
	Komite 2	59	1.9	-0.29	-0.23	0.84	81	35	0.59	0.29
	Komite 3	62	8.8	0.42	0.26	0.77	86	42	0.62	0.22
	Komite 4	62	9.9	-0.84	1.81	0.83	86	33	0.62	0.28
	Komite 5	59	1.5	0.24	-0.04	0.83	87	36	0.6	0.28
Dönem III	Final 1	59	11.6	-1	0.67	0.86	80	27	0.59	0.33
	Final 2	53	9.5	0.05	-0.8	0.79	75	35	0.53	0.23
	Komite 1	62	1.5	-1	2.9	0.85	83	21	0.62	0.37
	Komite 2	56	11.1	-0.5	0.58	0.85	79	22	0.56	0.32
	Komite 3	58	1.9	-0.91	2.06	0.85	78	19	0.58	0.3
	Komite 4	65	8.3	-0.36	0.08	0.76	84	46	0.65	0.24
Dönem III	Komite 5	68	9.2	-0.56	1.19	0.81	85	36	0.68	0.28
	Komite 6	64	6.8	-0.51	1.04	0.64	78	43	0.64	0.18
	Komite 7	74	6.2	-0.08	-0.45	0.65	88	59	0.74	0.16
	Final 1	56	8.9	-0.58	0.08	0.77	71	33	0.56	0.21
	Final 2	54	8.1	-0.09	0.19	0.71	72	32	0.54	0.17

*: Madde güçlük indeksi, **: Madde-toplam test korelasyonu-madde ayırt ediciliği

Tablo 3: 2013-2014 Eğitim Öğretim Yılında Kullanılan Testler ve Maddelere Yönelik Elde Edilen Psikometrik Değerler

		2013-2014 Eğitim Öğretim Yılı								
		Aritm. Ort.	Std. Sap	Çarpıklık	Basıklık	KR-20	Maks.	Min.	Ort. p_j^*	Ort. r_{cs}^{**}
Dönem I	Komite 1	74	12.3	-2.6	8.32	0.92	88	19	0.74	0.56
	Komite 2	61	9.2	0.13	0.07	0.81	82	37	0.61	0.26
	Komite 3	66	9.8	-0.39	0.56	0.83	92	38	0.66	0.3
	Komite 4	62	11	-0.44	1.05	0.86	86	25	0.62	0.33
	Komite 5	62	11	-0.6	1.03	0.86	83	26	0.62	0.32
	Final 1	65	9.9	-0.44	2.3	0.83	92	28	0.65	0.3
	Final 2	65	9.3	-0.57	0.14	0.81	83	37	0.65	0.27
Dönem II	Komite 1	62	12.4	-0.06	0.38	0.88	91	28	0.62	0.35
	Komite 2	59	12.8	-0.17	0.26	0.89	88	26	0.59	0.36
	Komite 3	67	11.9	-0.56	0.68	0.89	93	29	0.67	0.38
	Komite 4	64	12.1	-0.61	0.2	0.89	90	31	0.64	0.39
	Komite 5	59	13.9	0.07	-0.34	0.91	92	23	0.59	0.4
	Final 1	65	11	-0.23	0.49	0.87	90	33	0.65	0.34
	Final 2	59	9.1	0.2	0.97	0.79	83	31	0.59	0.23
Dönem III	Komite 1	64	1.4	-0.5	1.15	0.85	87	27	0.64	0.31
	Komite 2	62	9.2	-0.13	0.2	0.8	84	39	0.62	0.24
	Komite 3	59	9.6	-1.5	4.17	0.8	78	24	0.59	0.28
	Komite 4	60	9.3	-0.01	-0.22	0.8	81	38	0.6	0.24
	Komite 5	66	7.2	-0.18	-0.5	0.7	80	47	0.66	0.2
	Komite 6	59	8.6	-0.38	1.9	0.77	86	30	0.59	0.23
	Komite 7	72	7.33	-1	2.2	0.73	88	47	0.72	0.25
	Final 1	57	7.7	0.23	1.11	0.72	83	39	0.57	0.18
	Final 2	59	1.5	-2.4	11.8	0.85	78	2	0.59	0.34

*: Madde güçlük indeksi, **: Madde-toplam test korelasyonu-madde ayırt ediciliği

Tablo 4: 2014-2015 Eğitim Öğretim Yılında Kullanılan Testler ve Maddelere Yönelik Elde Edilen Psikometrik Değerler

		2014-2015 Eğitim Öğretim Yılı								
		Aritm. Ort.	Std. Sap	Çarpıklık	Basıklık	KR-20	Maks.	Min.	Ort. p_j^*	Ort. r_{cs}^{**}
Dönem I	Komite 1	73	1.4	-2.4	7.65	0.88	87	27	0.73	0.46
	Komite 2	64	12.3	-1.28	2.48	0.89	87	21	0.64	0.42
	Komite 3	63	11.7	-0.87	1.38	0.88	86	21	0.63	0.37
	Komite 4	60	11.3	-0.33	1.59	0.85	93	21	0.6	0.32
	Komite 5	62	11.6	-0.78	1.31	0.87	83	21	0.62	0.32
	Final 1	57	1.8	-1.78	7.55	0.84	77	2	0.57	0.31
Dönem II	Final 2	63	13.7	-2.2	6.63	0.91	83	1	0.63	0.44
	Komite 1	54	1.3	-0.26	-0.47	0.82	74	29	0.54	0.26
	Komite 2	58	15.3	-0.87	1.01	0.91	90	16	0.58	0.42
	Komite 3	66	11.8	0.07	0.13	0.88	92	37	0.66	0.38
	Komite 4	62	8.8	0.06	-0.54	0.78	84	45	0.62	0.25
	Komite 5	60	1.3	0.43	-0.2	0.83	85	40	0.6	0.28
Dönem III	Final 1	55	9.2	0.47	1.48	0.78	87	33	0.55	0.2
	Final 2	56	8.5	0.13	-0.2	0.75	75	35	0.56	0.2
	Komite 1	60	12.3	0.34	-0.47	0.89	88	34	0.6	0.36
	Komite 2	57	13.5	-0.42	0.1	0.9	86	26	0.57	0.38
	Komite 3	66	1.8	-0.21	-0.72	0.85	87	42	0.66	0.31
	Komite 4	60	13.1	-0.95	1.82	0.89	84	16	0.6	0.39
Dönem III	Komite 5	69	11.5	-1.36	2.24	0.88	85	30	0.69	0.4
	Komite 6	63	11.1	-1.1	2.73	0.86	83	23	0.63	0.33
	Komite 7	76	9.5	-2.1	7.05	0.87	94	35	0.76	0.44
	Final 1	64	1	-0.71	-0.12	0.84	80	38	0.64	0.29
	Final 2	61	1.5	-0.28	0.57	0.84	85	29	0.61	0.31

*: Madde güçlük indeksi, **: Madde-toplam test korelasyonu-madde ayırt ediciliği

Tartışma

Tıp eğitimi programıyla hekim adaylarına kazandırılması hedeflenen davranışlar doğrudan insan sağlığı ve hayatına yönelik uygulamaları içermektedir. Bu nedenle hekim adayları hakkında verilen kararların isabetliliği konusu bazen hayati öneme sahip olabilmektedir. İsbetli kararlar ise daha önceden açıklandığı gibi ancak nitelikli ölçme araçlarından elde edilen puanlarla dayandırılarak verilebilir.

Bu çalışmada da Mersin Üniversitesi, Tıp Fakültesi dönem bir, iki ve üç öğrencilerine son üç yıl içinde uygulanan testler ve bu testlerde kullanılan maddelerin psikometrik özelliklerinin incelenmesi ve ilgili sınavların nitelikleri hakkında bilgi elde edilmesi amaçlanmıştır. Araştırma kapsamında elde edilen bulgular; son üç yıl içinde uygulanan sınavların genel olarak kesme puanına (60 puan) yakın ortalama değerler aldığını ve dağılım yapılarının hafif sola çarpık olduğunu göstermiştir. Buna ek olarak sınavlardan elde edilen ortalama madde güçlük indeksi değerlerinin %71'inin 0.60 ve üzerinde yer aldığı gözlenmiştir.

Alanyazında 0.20'den küçük ve 0.80'den büyük madde güçlük indeksi değerlerinin test içinde kullanımı genel olarak önerilmemektedir.⁵ Bununla beraber madde güçlük indeksi değerleri ortalamasının 0.50'ye yakın olması; puan değişkenliğinin sağlanması, ranjin ve varyansın genişlemesi, güvenirlüğün yükselmesi vb. değişkenler açısından istenilen bir durum olarak karşımıza çıkmaktadır.⁴ Diğer taraftan özellikle eğitim programları içinde gerçekleştirilen sınavlardan elde edilen madde güçlük indeksi değerlerinin yüksek çıkması, bu maddelerin yokladığı davranışların bireyler tarafından iyi öğrenildiğinin de bir göstergesi olarak da kabul edilebilir. Bu çalışmada ise madde güçlük indeksi değerlerinin genel olarak 0.60 ve üzerinde yer alması; tıp fakültesi öğrencilerinin tıp fakültesini kazanmak için lisans yerleştirme sınavında gösterdikleri başarı, yapılan sınavlara yönelik çalışma yoğunlukları, vb. durumlar da göz önünde bulundurulduğunda; öğrencilerin konulara hakim oldukları ve nispeten başarılı bir

grafik sergiledikleri şeklinde yorumlanabilir. Nitekim elde edilen bulgular Sayek ve ark.¹⁷ tarafından hazırlanan rapor ile Ogenler ve Selvi¹⁸ tarafından gerçekleştirilen çalışmanın bulgularıyla da örtüşmektedir.

Diğer taraftan elde edilen bulgular, incelenen 69 sınavın 63'ünün iç tutarlık bağlamındaki güvenirlüklerinin (KR-20) 0.70 ve üzerinde, 3'ünün ise 0.64-0.65 aralığında yer aldığını göstermiştir. Alanyazında güvenilir bir test için KR-20 değerinin 0.70'ten büyük olmasının gerektiği 0.60 değerinin ise güvenirlük için kabul edilebilir alt sınır olduğu ifade edilmektedir.¹ Yine alanyazında, KR-20 paralel maddeler arası varyans hesaplamalarına dayandığından 0.90'dan büyük bir KR-20 değerinin ise tek boyutluluk göstergesi sayılabileceği belirtilmektedir.¹⁹ Tıp fakültelerinde gerçekleştirilen sınavların ise birbiriyle bağlantılı (yatay ve dikey entegrasyon) birden fazla dersin birleşimiyle oluşturulduğu bilinmektedir. Bu nedenle 0.90'dan büyük bir KR-20 değerini beklemek bu sınavlar için mantıklı görünmemektedir. 0.70'ten küçük değerler ise testin güvenirlüğü ile ilgili sıkıntı olduğu şeklinde yorumlanabilir. Çalışma kapsamında elde edilen bulgular ise Mersin Üniversitesi Tıp Fakültesinde uygulanan sınavlardan elde edilen puanların güvenirlüklerinin oldukça yüksek olduğunu göstermektedir.

Madde-toplam test korelasyonu açısından duruma bakıldığında ise alanyazında bu değerlerin alt sınırının 0.30 olarak ifade edildiği gözlenmiştir.⁷ Araştırma kapsamında elde edilen bulgular ise madde-toplam test korelasyonu değerlerinin genel olarak düşük olduğunu göstermektedir. Bu duruma madde yazarı öğretim üyelerinin nitelikli madde yazma konusunda yaşadıkları sıkıntılar, konu kapsamının sınırlılıkları, testlerden elde edilen maksimum ve minimum puanlar arası ranjin düşük olması ve bu durumun varyansı daraltması, vb. sorunların neden olduğu düşünülmektedir.^{20,21,7} Bununla birlikte elde edilen bulgular, kabul edilebilir alt sınırın altında kalan madde-toplam test korelasyonu değerlerine ilişkin yüzdelerin sürekli azaldığını göstermektedir. Bu

gelişime o yıllarda madde yazarı öğretim üyelerine verilmeye başlanan nitelikli madde yazma konulu hizmet içi eğitimlerin katkı sağladığı düşünülmektedir.

Sonuç olarak yapılan bu çalışma; Mersin Üniversitesi, Tıp Fakültesi dönem bir, iki ve üç öğrencilerine son üç yıl içinde uygulanan sınavlardan elde edilen puanların iç tutarlık bağlamındaki güvenilirliklerinin oldukça yüksek olduğunu, öğrencilerin kesme puanına göre nispeten başarılı bir grafik sergilediklerini, bununla birlikte madde toplam test korelasyonu (madde geçerliği) değerlerinin ise bir miktar düşük olduğunu göstermiştir. Benzer bulgulara alanyazında gerçekleştirilen diğer çalışmalarda da rastlanması bu sorunun farklı fakültelerde de yer aldığı bir göstergesi olarak kabul edilebilir. Ancak burada hesaplanan madde-toplam test korelasyonu değerleri en nihayetinde korelasyon hesaplamalarına dayanmaktadır ve bu değerlerin; dağılımın çarpıklığından, ranj daralmasından, vb. etkilenebildiği bilinmektedir. Bu nedenle bu değerlerin hesaplanan diğer değerlerle birlikte yorumlanması ve alt sınır olarak kabul edilen 0.30 değerinden çok düşük çıkması durumunda artırılmasına yönelik madde yazarı öğretim üyelerini kapsayan hizmet içi eğitimler planlamak ve sürdürmek gibi bir dizi önlem alınmasının gerekli olduğu ifade edilebilir.^{5, 9, 10, 11}

Bununla birlikte bu ve benzer çalışmaların belirli aralıklarla tekrarlanması, yapılan sınavların niteliklerinin ve böylece verilen eğitimin kalitesinin artırılması hususunda yol gösterici olduğundan önerilebilir.

Kaynaklar

1. Murphy K.R. & Davidshofer, C.O. Psychological testing: principles and applications. Prentice Hall, New Jersey 2001.
2. Anastasi A. ve Urbina S. Psychological testing. Prentice Hall 1997.
3. Aiken L., Psychological testing and assessment. Allyn&Bacon 2000.

4. Ulusal tıp eğitimi akreditasyon kurulu mezuniyet öncesi tıp eğitimi standartları. (2014). Erişim Tarihi: 16.12.2015. Erişim Adresi: http://www.uteak.org.tr/uploads/belge/MOTE_STANDARTLAR_2014.pdf
5. Erkuş A. Sınıf öğretmenleri için ölçme ve değerlendirme: kavramlar ve uygulamalar. Ankara: Ekinoks Yayınları 2006.
6. Erkuş A. Psikolojide ölçme ve ölçek geliştirme-I: temel kavramlar ve işlemler. Ankara: Pegem Akademi 2012.
7. Crocker L. ve Algina J. Introduction to Classical and Modern Test Theory. Holt, Rinehart and Winston, Inc 1986.
8. Camili G., Shepard L.A. Methods for identifying biased test items. Sage Publication 1994.
9. Özcan K.V., Aydoğan Y., Bulut, İ. Gaziosmanpaşa üniversitesi tıp fakültesi'nde uygulanan çoktan seçmeli sınavların betimsel analizi. *Gaziosmanpaşa Üniversitesi Tıp Fakültesi Dergisi*. 2014;6(4):281-94.
10. Topal K., Aybek H., Kara C.O., Büke A., Aybek Z.. PAÜ tıp fakültesi dönem 1 öğrencilerine 2006-2007 eğitim ve öğretim yılında uygulanan çoktan seçmeli sınavların madde ve test analizleri. *Pamukkale Tıp Dergisi*. 2008;1(3):120-26.
11. Eskiocak S., Gökmen Süer S., Erbaş H., Çakır E., Gülen Ş. Trakya üniversitesi tıp fakültesinde son 5 yılda yapılan biyokimya sınav sorularının analizi. *Türk Biyokimya Dergisi* 2004;29 (4);273-76.
12. Şimşek Cankur N., Ediz B, Sırmalı Ş A. Entegre Eğitim Sisteminde Uygulanan sınavlardaki Doğru Yanıt Seçeneklerinin Dağılımı. *Uludağ Üniversitesi Tıp Fakültesi Dergisi* 2002;28(2):1-3.
13. 2014-2015 Mersin Üniversitesi, Tıp Fakültesi, Faaliyet Raporu 2015.
14. Baykul Y. Eğitimde ve psikolojide ölçme. ÖSYM yayınları 2000.

15. Şenocak M. Ş. Biyoistatistik & Araştırma yöntemleri. İstanbul medikal yayıncılık 2014.
16. Tekin H. Eğitimde ölçme ve değerlendirme, Yargı yayınevi, Ankara 1993.
17. Sayek İ., Odabaşı O., Kiper N. Türk tabipler birliği mezuniyet öncesi tıp eğitimi raporu-2010, Türk Tabipleri Birliği Yayınları. Ankara 2010.
18. Ogenler O. & Selvi S. Variables that affect medical faculty student's achievement: sample of mersin university. *IRCMJ* 2014;16(3):e14648. doi: 10.5812/ircmj.14648.
19. Hattie J. Methodology review: Assessing unidimensionality of tests and Items. *Applied Psychological Measurement* 1985;9(2):139-64.
20. Gulliksen, H. Theory of mental tests. New York: Wiley 1950.
21. Magnusson, D. Test Theory. Massachusetts. Addison -Wesley 1967.