

Özgün araştırma

Çoktan Seçmeli Sınavlarda Kullanılan Ölçme ve Değerlendirme Sisteminin Uygulanması: Madde Analiz Örneği

Tuğba Badat¹, Günseli Usgu², Elif Dinler³, Kezban Bayramlar⁴, Yavuz Yakut⁵

Gönderim Tarihi: 4 Ekim, 2019

Kabul Tarihi: 20 Aralık, 2020

Basım Tarihi: 31 Aralık, 2020

Öz

Amaç: Bu çalışma, madde istatistik analizleri kullanılarak çoktan seçmeli sınavlarda, soru ve cevapların güvenilirliğini test etmek amacıyla yapıldı.

Gereç ve Yöntem: Çoktan seçmeli sınavın standart ve kalite açısından değerlendirilmesinde madde analiz yöntemi kullanılır. Madde analiz yöntemiyle öğrencilerin maddelere verdiği yanıtlar incelenerek sınavın güvenilirlik ve geçerlik düzeyi hakkında bilgi sağlanır. Çalışmamızda örnek olarak Ortopedik Rehabilitasyon dersi Final Sınavı seçildi. 20 soruluk çoktan seçmeli sınav 79 öğrenciye uygulandı. Test sonuçları istatistik programına işlendi. Sorular için madde güclüğü ve madde ayırt ediciliği belirlendi. Kuder Richardson 20 (KR-20) ile testin güvenilirliği değerlendirildi.

Bulgular: Analiz sonucunda testten alınan en yüksek puanın 85, ortalama puanın ise 58,5 olduğu bulundu. Sorular için madde güclüğü incelendiğinde, değerlerin 0,11 ile 0,91 arasında değiştiği ve orta güçlükte (0,59) bir sınav olduğu görüldü. Madde ayırt ediciliklerinin -0,02 ile 0,56 arasında değiştiği ve sınavın orta ayırt edicilikte olduğu belirlendi. Testin ayırt edebilme derecesinin 0,30 olduğu tespit edildi. Testin güvenilirliği incelendiğinde KR-20 değerinin 0,56 olduğu görüldü.

Sonuç: Yapmış olduğumuz ortopedik rehabilitasyon dersi çoktan seçmeli final sınavının; öğrencilerimiz için orta güçlükte ve orta ayırt edicilikte bir sınav olduğu, güvenilirlik yönünden ise sınırda bir test olduğu ve soruların tekrar gözden geçirilmesi gerektiği belirlendi. Derse yönelik gerekli iyileştirme, düzeltme ve soruların tekrar gözden geçirilmesinin, derse yönelik verimi artıracak görüşüdeyiz.

Anahtar Kelimeler: Ölçme ve Değerlendirme, Çoktan Seçmeli Sınav, Madde Analizi, Madde Ayırt Ediciliği

¹**Tuğba Badat (Sorumlu Yazar).** Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon Bölümü/ GAZİANTEP – 03422118080- tugba.badat@hku.edu.tr - 0000-0002-0484-0221

²**Günseli Usgu.** Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon Bölümü/ GAZİANTEP – 03422118080- gunseli.usgu@hku.edu.tr – 0000-0003-4269-5210

³**Elif Dinler.** Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon Bölümü/ GAZİANTEP – 03422118080- elif.dokunlu@hku.edu.tr – 0000-0003-0198-1456

⁴**Kezban Bayramlar.** Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon Bölümü/ GAZİANTEP – 03422118080- kezban.bayramlar@hku.edu.tr – 0000-0001-6912-4405

⁵**Yavuz Yakut.** Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon Bölümü/ GAZİANTEP – 03422118080- yavuz.yakut@hku.edu.tr – 0000-0001-9363-086

Original Research

Use of Measurement and Assessment in Multiple Choice Questions (MCQ's): Item Analysis Sample

Tuğba Badat¹, Günseli Usgu², Elif Dinler³, Kezban Bayramlar⁴, Yavuz Yakut⁵

Submission Date: 4th of October, 2019 **Acceptance Date:** 20th of December, 2020 **Pub.Date:** 31th of December, 2020

Abstract

Objective: This study was conducted to test the reliability of questions and answers in MCQ's by using item analysis.

Methods: In the evaluation of MCQ's in terms of standard and quality, item analysis method is used. The students answers to the items are examined with item analysis method and information is provided about the reliability and validity level of the exam. In our study, the final exam of Orthopedic Rehabilitation course, which consists of 20 multiple choice questions applied to 79 students, was selected. Item difficulty and item discrimination were determined for the questions. Reliability of the test was evaluated with Kuder Richardson (KR) 20.

Results: As a result of the analysis, the highest score obtained from the test was found to be 85 and the average score was 58.5. When the item difficulty was evaluated for the questions, it was seen that the values ranged between 0.11-0.91 and there was an intermediate difficulty test (0.59). It was determined that item discrimination was between -0.02 and 0.56 and the level of item discrimination of the questions were moderate. The degree of discrimination was found to be 0.30. When the reliability of the test was examined, KR-20 value was found to be 0.56.

Conclusion: The orthopedic rehabilitation lesson of the multiple choice final exam which we have done; was determined that there is a medium difficulty and medium discrimination test for our students, a test on the border in terms of reliability and questions should be revised. We believe that the necessary improvement, correction and revision of the questions will increase the efficiency of the course.

Keywords: *Measurement and Evaluation, Multiple choice question, Item Analysis, Item Discrimination*

¹**Tuğba Badat (Corresponding Author).** Hasan Kalyoncu University Faculty of Health Sciences Department of Physiotherapy and Rehabilitation / GAZİANTEP – 03422118080- tugba.badat@hku.edu.tr - 0000-0002-0484-0221

²**Günseli Usgu.** Hasan Kalyoncu University Faculty of Health Sciences Department of Physiotherapy and Rehabilitation / GAZİANTEP – 03422118080- gunseli.usgu@hku.edu.tr – 0000-0003-4269-5210

³**Elif Dinler.** Hasan Kalyoncu University Faculty of Health Sciences Department of Physiotherapy and Rehabilitation/ GAZİANTEP – 03422118080- elif.dokunlu@hku.edu.tr – 0000-0003-0198-1456

⁴**Kezban Bayramlar.** Hasan Kalyoncu University Faculty of Health Sciences Department of Physiotherapy and Rehabilitation / GAZİANTEP – 03422118080- kezban.bayramlar@hku.edu.tr – 0000-0001-6912-4405

⁵**Yavuz Yakut.** Hasan Kalyoncu University Faculty of Health Sciences Department of Physiotherapy and Rehabilitation / GAZİANTEP – 03422118080- yavuz.yakut@hku.edu.tr – 0000-0001-9363-0869

Giriş

Değerlendirme, öğretme ve öğrenme siklusunun en önemli bileşenleridir. Değerlendirme yöntemi olarak kullanılan çoktan seçmeli sorular, öğrencilerin öğrenme performanslarını değerlendirmede objektif ve güvenilir ölçüm aracı olarak kullanılmaktadır ve çok sayıda katılımcıya kısa sürede uygulanabilmesi, sonuçların kısa sürede alınması, tarafsız ve adil olması nedeniyle günümüzde sıkça tercih edilen bir ölçme ve değerlendirme yöntemidir (Yılmaz, 2004; Durak, 2002). Çoktan seçmeli soruların hazırlanması karmaşık, zorlayıcı ve zaman alıcı bir süreçtir (Kolte, 2015). Çoktan seçmeli sınavlar kavramsal anlayış ve yorumlamadan ziyade gerçek bilgilerin hatırlanmasına vurgu yapmaktadır. Çoktan seçmeli soru yazmak, iyi sorular yazmaktan daha fazlasıdır. Düzgün bir şekilde yapılandırılmış maddeler; sadece izole edilmiş gerçeklerin hatırlanmasını test etmek yerine, *Bloom'un* taksonomisinde yer alan yorumlama, sentez ve bilgi uygulaması gibi yüksek bilişsel işlemleri değerlendirmektedir (Mehta, Mokhasi, 2014). Çoktan seçmeli sınavın standart ve kalite açısından değerlendirilmesinde madde analiz yöntemi kullanılır (Singh, Gupta, Singh, 2009). Öğrenme performanslarına hangi seviyede ulaşıldığını ölçmek için kullanılan sınavların adil, kullanışlı ve güvenilir olması gerekmektedir (Eskiocak, Gökmen, Erbaş, Çakır, Gülen, 2004). Madde analiz yöntemi ile öğrencilerin maddelere verdiği yanıtlar incelenerek, sınavın güvenilirlik ve geçerlik düzeyi hakkında bilgi sağlanır (Considine, Botti, Thomas, 2005). Eğitim süreci hakkında bilgi elde edilebilmesi için sürecin her aşamasında kullanılacak ölçme ve değerlendirme yöntemleri, bireyin hedeflere ulaşip ulaşmadığını ve bunun ne düzeyde gerçekleştiğini göstermesi açısından önemlidir (Tekin, 2000). Eğitimin aksayan ve eksik yönlerinin belirlenmesi, öğretimde geleceğe yönelik hedeflerin ve en genel anlamda öğretimi geliştirmeye yönelik önlemlerin alınabilmesi ve düzenlemelerin yapılabilmesi açısından önemlidir (Tekin, 2000).

Madde analizi; sınava uygun soruların seçilmesi, belirli nitelikleri taşımayan maddelerin düzeltilmesi ve uygun olmayan maddelerin sınavdan çıkartılması işlemlerinin yapılma sürecidir. Sınav sonrası yapılan madde analizi hatalı soruların tespit edilmesini ve bu soruların değerlendirme dışı bırakılarak sınavın adil bir şekilde yapılmasını sağlar (Çelik, 2000). Bu analizler sonucunda sınavlar üzerinde yapılan düzenlemeler, eğitimciler için de yol gösterici olmaktadır.

SABAK Öz Değerlendirme (ÖDR) hazırlık sürecinde bölümümüzde zayıf ve güçlü yönlerin tespit edilmesiyle ölçme değerlendirme sürecine katkısı bulunan madde analizi eğitimine ihtiyaç duyuldu. Üniversitemiz ilgili biriminden madde analiz eğitimi alındı.

Bu çalışma; Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon Bölümü'nün 2018-2019 Eğitim-Öğretim yılı Bahar Dönemi 6. Yarıyıl Ortopedik Rehabilitasyon dersi çoktan seçmeli final sınavının değerlendirilmesi ve soruların güçlük, ayırt edicilik ve güvenilirliklerinin belirlenerek; ilgili öğretim üyelerine dengeli, kaliteli, istenen bilgiyi ölçen soruların hazırlanması yönünde geri bildirim sağlanması amacıyla planlandı.

Gereç ve Yöntem

Bu çalışma Hasan Kalyoncu Üniversitesi Sağlık Bilimleri Fakültesi Fizyoterapi ve Rehabilitasyon Bölümü'nde gerçekleştirildi. Çalışmamızda, ortopedik rehabilitasyon dersi final sınavı madde analizi yapıldı. 3. sınıf öğrencilerimize uygulanan sınava 79 öğrenci katıldı. 20 sorudan oluşan çoktan seçmeli sınav için belirlenen süre 30 dk idi. Değerlendirme 100 puan üzerinden yapıldı. Tüm öğrencilerin sınav puanları değerlendirildi. Öğrenciler başarı durumlarına göre üst grup ve alt grup olarak sınıflandırıldı. Öğrencilerin yüksek not alan 1/3 lük kısmı üst grup, düşük not alan öğrencilerin 1/3'ünün alt grubu oluşturduğu kabul edildi (Kaur, 2017; Garg, Kumar, Maria, 2019). Üst grupta bulunan öğrencilerin sayısı 30, alt grupta bulunan öğrencilerin sayısı 26 olarak belirlendi. Orta grupta bulunan öğrenciler ise 23 kişi olarak ayrıldı.

Her soru; madde güçlüğü, madde ayırt ediciliği ve madde geçerliği-güvenirliği yönünden analiz edildi.

1. Madde Güçlüğü: Bir maddeye doğru cevap veren kişi sayısının testi alan tüm kişi sayısına oranı madde güçlüğü (p) verir. Madde güçlüğü (p) hesaplama formülü aşağıdaki gibidir (Garg, Kumar, Maria, 2019).

$$[(H + L) / N] \times 100$$

H: Üst grupta soruları doğru cevaplayan öğrenci sayısı

L: Alt grupta soruları doğru cevaplayan öğrenci sayısı

N: İki gruptaki toplam öğrenci sayısı

Madde güçlüğü, soruların kolay, orta ve zor sorular olma özelliğini belirtmektedir. Bu değer 1'e yaklaştıkça soru kolaylaşırken, 0'a yaklaştıkça soru zorlaşır; 0,50'ye yakın olduğunda ise orta güçlükte bir soru olduğu belirtilir. Bir maddenin % 30-70 arasında bir "p" değeri varsa, bu soru kullanılabilir olarak kabul edilir. Zorluk indeksi % 70' ten büyük olan maddeler kolay kabul edilir (Tablo 1). Bir gruptaki öğrencilerin hepsinin doğru yanıtladığı bir sorunun güçlük değeri 1 olarak bulunur. Güçlük değeri 1 olan sorunun tekrar gözden geçirilmesi gerekmektedir.

Tablo 1. Testlerin madde güçlüğü değerine göre yorumlanması

p değeri	Yorum
<%30	Çok zor
>%70	Çok kolay
%50-%60	İdeal
%30-%70	Kabul edilebilir

2. Madde Ayırt Ediciliği: Üst grupta soruyu doğru yanıtlayanlar ile alt grupta soruyu doğru yanıtlayanların sayılarının farkının, kişi sayısına oranı madde ayırt ediciliğini verir. Üst grupta olanlar ile alt grupta olanlar arasındaki sayısal fark arttıkça, bir maddenin ayırt ediciliği artar. Madde ayırt ediciliği (d) hesaplama formülü aşağıdaki gibidir (Garg, Kumar, Maria, 2019).

$$d = [(H-L)/N] \times 2$$

Maddenin ayırt ediciliği -1 ile +1 arasındadır. +1 ayırt edicilik değeri, maddenin maksimum ayırt edici güce sahip olduğu anlamına gelir. Madde ayırt ediciliğinin 0.35'ten daha büyük olduğu ($d > 0.35$) bulunan testlerin mükemmel ayırım gücüne sahip yani yüksek ayırt edicilikte olduğu, 0.2 ile 0.35 arasında ($0.2 < d < 0.35$) bulunan değerlerin ayırt ediciliğinin orta düzeyde olduğu ve ayırt ediciliği 0.2'nin altında olan ($d < 0.2$) testlerin düşük düzeyde ayırt etme özelliğine sahip olduğu bilinmektedir (Tablo 2). Ayırt edicilik değeri "0" olduğunda, maddenin konuyu bilen ve bilmeyen öğrencilerin ayırımını yapamadığını gösterir ve bu istenmeyen bir durumdur. -1 ila 0

arasında değişen negatif ayırt ediciliğe sahip bir madde, düşük not alan öğrencilerin o maddeyi doğru yaptığını ve tüm yüksek not alan öğrencilerin o maddeyi yanlış yaptığını gösterir. Negatif ayırt ediciliğe sahip olan maddeler sadece yararsız olmakla kalmayıp, testin geçerliliğini de düşürmektedirler (Gajjar, Sharma, Kumar, Rana, 2014). Bu nedenle negatif ayırt ediciliğe sahip olan maddelerin çoktan seçmeli sınavdan çıkartılması gerekmektedir (Obon, Rey, 2019)

Tablo 2. Testlerin madde ayırt ediciliği değerine göre yorumlanması

d değeri	Yorum
>0,35	Mükemmel
0,20-0,35	Kabul edilebilir
<0,20	Zayıf

3. Madde Geçerlik ve Güvenirliği: Çoktan seçmeli sınavın madde geçerliği ve güvenirliği; öğrencilerin başarı puanlarının dağılımına göre analiz edildi. Standart sapmanın küçük olması ortalama başarı puanından sapmaların ve riskin az olduğunu gösterirken, standart sapmanın büyük olması ortalama başarı puanından sapmaların ve riskin çok olduğunu gösterir. Farklı başarı düzeyindeki öğrencilerin ayırt edilmesini amaçlayan çoktan seçmeli sınavın standart sapması büyük olmalıdır. Bu durum testin güvenirliliğinin yüksek olduğunu ifade eder (Topal, Aybek, Büke, Kara, Aybek, 2008).

Çoktan seçmeli sınavların, güvenirlilik değerinin veya iç tutarlılığının göstergesi Kuder Richardson 20 formülüdür (KR-20). Sınav tekrarı durumunda ilk sınavın etkisinde kalmadan öğrencilerin aynı puanı alabilme derecelerini yansıtır. KR-20 formülü, testin ölçtüğü bilginin homojen olduğu varsayımından hareketle gerçekleştirilmiştir. KR-20 formülü ile güvenirlilik değeri 0 ila 1 arasında değişmektedir. Testin madde geçerlik ve güvenirlilik değerinin KR 20 indeksine göre yorumlanması Tablo 3’de gösterildi. Değer 1’e yaklaştıkça güvenirliliğin arttığı kabul edilmektedir. KR-20 indeksi; 0,9 dan büyük ise testin mükemmel güvenirlilikte olduğu, 0,8 ile 0,9 arasında güvenirlilik düzeyinin iyi olduğu, 0,8- 0,5 arasında ise sınırda test olduğu ve düzeltilmesi gerektiği, 0,5’ten düşüğe test sonuçlarının güvenilir olmadığı kabul edilir (Kheyami, Jaradat, Al-Shibani, Ali, 2018).

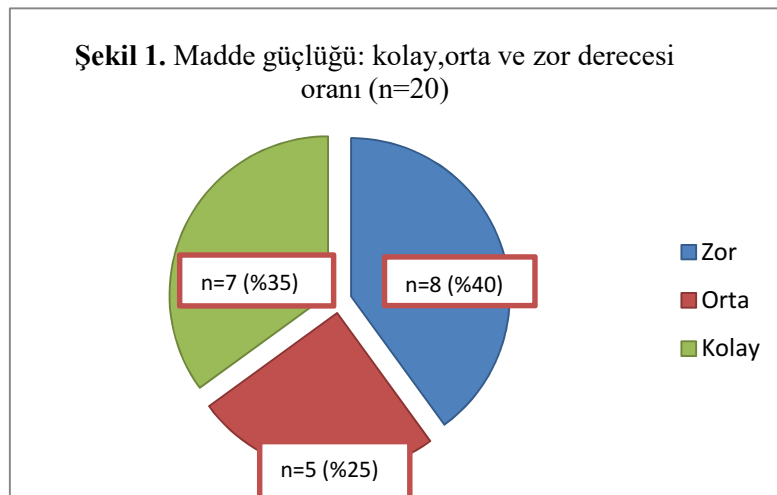
Tablo 3. Testlerin madde geçerlik ve güvenirlik değerinin KR 20 indeksine göre yorumlanması

KR-20 değeri	Yorum
>0,90	Mükemmel güvenirlik
0,80 – 0,90	Çok iyi
0,70- 0,80	İyi
0,60- 0,70	Değiştirilmesi gereken öğeler yer almaktadır
0,50- 0,60	Testin gözden geçirilmesi gerekmektedir.
<0,50	Şüpheli güvenirlik

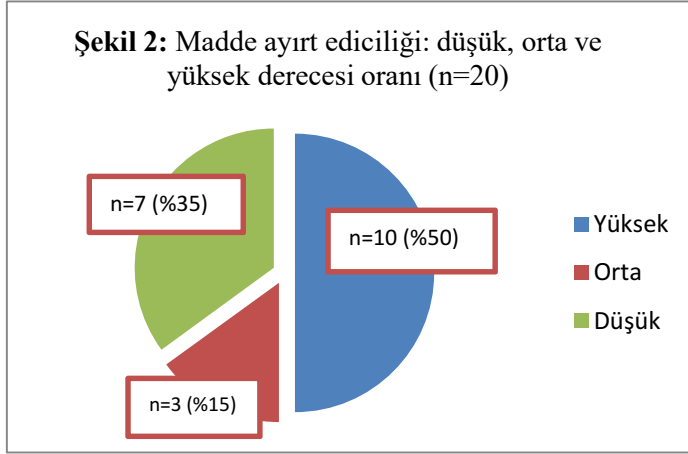
Bulgular

Yetmiş dokuz öğrencinin katıldığı çoktan seçmeli sınava yönelik yapılan madde analizi sonucunda 20 maddenin güçlük indeksi (p), ayırt ediciliği (d) ve güvenirlik (KR-20) değeri hesaplandı. En fazla 100 puanın alınabildiği çoktan seçmeli sınavdan; alınan en düşük puan 20, en yüksek puan 85'tir. Testten alınan ortalama puan 58,5'dir. Çoktan seçmeli sınavın güçlük indeksi, ayırt edicilik ve güvenirlik değerleri Tablo 4'de verildi.

Çoktan seçmeli sınavın madde güçlük değerinin, 0,11 ile 0,91 arasında değiştiği bulundu. Sınavın madde güçlüğü ortalamasının 0,59 olduğu ve maddelerin %35'inin güçlük indeksi açısından kolay ($p > 0,70$), maddelerin %25'inin güçlük indeksi açısından orta ($0,50 < p < 0,70$), maddelerin %40'ının ise güçlük indeksi açısından zor olduğu ($p < 0,30$) hesaplandı (Şekil 1).



Çoktan seçmeli sınavın madde ayırt ediciliği incelendiğinde, -0,02 ile 0,56 arasında değiştiği bulundu. Testin ortalama ayırt edebilme derecesinin 0,30 olduğu tespit edildi. Maddelerin %50'sinin ayırt ediciliğinin yüksek ($d > 0,35$), %15'inin ayırt ediciliğinin orta ($0,20 < d < 0,35$) ve %35'inin ayırt ediciliğinin düşük düzeyde olduğu ($d < 0,20$) saptandı (Şekil 2). Testin güvenilirliği incelendiğinde, KR-20 değerinin 0,56 ile sınırdaki test olduğu belirlendi.



Tablo 4. Testin özellikleri ve madde analiz sonuçları

Parametreler	Değer
Soru sayısı	20
Öğrenci sayısı	79
Testin ortalama puanı % (SD)	11,2 (2,8)
Testin puan aralığı	20-85
Madde güçlüğü % (P)	
Ortalama	58,5
Aralık	0,11 – 0,91
Madde ayırt ediciliği % (d)	
Ortalama	0,30
Aralık	-0,02 – 0,56
Test güvenilirliği % (KR) 20	0,56

Tartışma ve Sonuç

Çoktan seçmeli soruları içeren testler, bilişsel öğrenme alanını değerlendirmek için yaygın olarak kullanılan, etkili bir ölçme aracıdır (Singh, Anshu, 2012). Yirmi çoktan seçmeli sorudan oluşan sınavımız bütün olarak analiz edildiğinde, madde gücünün ortalama değeri 0,50 ila 0,70 arasında orta güçlükte olduğu bulundu. İncelediğimiz soruların %65'inin ayırt edicilik düzeyinin orta ve yüksek olması, sınav sorularımızın konuyu bilen ve bilmeyen öğrencileri ayırt etmekte başarılı olduğunu göstermektedir. İç tutarlık ve güvenilirlik açısından incelendiğinde sınavımızın KR-20 indeksinin 0,8-0,5 değerleri arasında oluşu, sınavı sınırdaki test olarak göstermektedir.

Madde gücü, öğrencilerin test edilen kavramı öğrenip öğrenmediğini ölçmektedir (Patel, 2017). Çalışmamızda toplam 9 madde zorluk indeksine göre kabul edilebilir (%30-%70) düzeye sahipti. Tüm test genelinde bu oran %45 olarak tespit edildi. 3 sorunun zor ($p < \%30$), 8 sorunun kolay ($p > \%70$) olduğu belirlendi. Patel'in yapmış olduğu bir çalışmada 50 soruluk testte kabul edilebilir güçlükte olan soruların sayısının 30 olduğu ve testin geneline göre bu oran %75 (p 30-70) olarak bulunmuştur. Sınavdaki kolay güçlükte olan soruların %17,5 ($p > 70$), zor güçlükte olan soruların ise %7,5 ($p < 30$) oranında olduğunu bulmuşlardır (Patel, 2017). Karelia ve arkadaşları değerlendirdikleri çoktan seçmeli sınavda; kabul edilebilir güçlükte olan soruların oranını % 61 (p 30-70), kolay güçlükte olan soruların oranını % 24 ($p > 70$) ve zor güçlükte olan soruların oranını % 15 ($p < 30$) olarak analiz etmişlerdir (Kareli, Pillai, Vegada 2013). Madde gücünün ortalama değerinin literatür ile benzerlik göstererek ideal aralıkta olması, öğrencilerin ortopedik rehabilitasyon bilgisini sınamak için uygun güçlükte bir sınav olduğunu göstermiştir.

Madde ayırt ediciliği, bir öğenin test edilen materyali ne kadar iyi bildiğini temel alarak öğrenciler arasında ayırım yapabilme yeteneğini ifade eder (Patel, 2017). Analiz ettiğimiz testte, toplam soruların %65'inin ayırt edicilik indeksinin 0,20'den büyük olduğu bulundu ($p > 0,20$). 10 sorunun (%50) mükemmel ayırtıcı güce, 3 sorunun (%15) kabul edilebilir ayırtıcı güce ve 7 sorunun (%35) zayıf ayırtıcı güce sahip olduğu belirlendi. Mahjabeen ve arkadaşlarının 2018'de yaptıkları çalışmada; soruların %52'sinin mükemmel ayırtıcı güce, %31'inin kabul edilebilir ayırtıcı güce ve %17'sinin de zayıf ayırtıcı güce sahip olduğu gösterilmiştir (Mahjabeen ve diğ., 2018). Kolte ve arkadaşlarının yaptıkları çalışmada; 24 (% 60) maddenin mükemmel ayırt edici güce, 7 (% 17,5) maddenin iyi ayırtıcı güce, 8 (% 20) maddenin kabul edilebilir ayırtıcı güce ve 15 (% 37,5) maddenin zayıf ayırtıcı güce sahip olduğunu

belirtmişlerdir (Kolte, 2015). Testimizin ortalama ayırt edicilik indeksine bakıldığında, sınav sorularının konuyu bilen ve bilmeyen öğrencileri ayırt etmede yapılan çalışmalarla benzer nitelikte olduğu görülmektedir.

Değerlendirdiğimiz ortopedik rehabilitasyon sınavının; iç tutarlık ve güvenilirlik (KR-20) indeksinin 0,8-0,5 değerleri arasında oluşu, sınavı sınırda test olarak tanımlamaktadır. Güvenirlik ve iç tutarlılığının düşük olması nedeniyle maddelerin tekrar gözden geçirilerek sınav sorularının güncellenmesi, uygun olmayan maddelerin testten çıkarılması gerekmektedir. Bu değişikliklerin yapılması ve daha kaliteli maddelerin eklenmesi ile, çoktan seçmeli sınavımızın güvenilirliği ve iç tutarlılığı yükseltilmiş olacaktır.

Bu analizlerin her sınavın sonunda yapılması ve öğrencilerin puanları ilan edilmeden önce ayıricılığı az olan maddelerin, KR-20 indeksi düşük olan soruların sınavlardan çıkarılarak sınav puanlamasının yeniden yapılması ve notların o şekilde ilan edilmesi önerilebilir. Ölçme ve değerlendirme yoluyla elde edilen bilgiler sadece öğrencinin kazanımlarını yansıtmamakla birlikte, öğretimin verimliliğine ve etkinliğine yönelik bilgiler de sağlar. Böylelikle eğitim sürecinin aksayan ve eksik yönlerini tespit etme olanağı elde edilir.

Çalışmadan elde edilen sonuçlar ışığında, madde analizi gibi ölçme değerlendirmeye ilişkin eğitimlerin alınması ve bölümler bazında bunların yaygınlaştırılmasının önemli olduğu kanısına varıldı. Özellikle ölçme değerlendirme eğitimlerinin bölümler çerçevesinde alınması ve bu konuda genç akademisyenlerin yetiştirilmesinin derslere yönelik verimi artırarak, öğrencilerin başarı düzeylerinin de objektif ölçülebilmesine yol açacağı görüşündeyiz.

Kaynakça

- Considine, J., Botti, M., & Thomas, S. (2005). Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian*, 12(1), 19-24.
- Çelik, D. (2000). Okullarda Ölçme Değerlendirme Nasıl Olmalı?(1. Baskı). *Ankara: MEB Yayınları*.
- Durak, H. İ. (2002). Eğitimde ölçme ve değerlendirmenin (sınama) öğretim üyeleri tarafından bilinmesi gereken temel ilkeleri. *Tıp Eğitimi Dünyası*, 7(7).
- Eskiocak S., Gökmen S. S., Erbaş H., Çakır E., Gülen Ş. (2004). Trakya üniversitesi tıp fakültesinde son 5 yılda yapılan biyokimya sınav sorularının analizi. *Türk Biyokimya Dergisi [Turkish Journal of Biochemistry-Turk J Biochem]*, 29(4), 273-276.
- Gajjar, S., Sharma, R., Kumar, P., & Rana, M. (2014). Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, 39(1), 17.
- Garg R., Kumar V., Maria J. (2019). Analysis of multiple choice questions from a formative assessment of medical students of a medical collage in Delhi India. *Int J Res Med Sci* 7(1):174-177.
- Karelia, B. N., Pillai, A., & Vegada, B. N. (2013). The levels of difficulty and discrimination indices and relationship between them in four-response type multiple choice questions of pharmacology summative tests of year II MBBS students. *IeJSME*, 7(2), 41-46.
- Kaur M., Singla S., Mahajan R. (2016). Item analysis of inuse multiple choice questions in pharmacology. *Int J Appl Basic Med Res*, 6(3):170-173.
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of educational psychology*, 30(1), 17.
- Kheyami D., Jaradat A., Al-Shibani T., Ali F. A. (2018). Item analysis of multiple choice questions at the department of paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos University Med J*, 18(1), 68-74.
- Kolte, V. (2015). Item analysis of multiple choice questions in physiology examination. *Indian J of Basic & Applied Medical Research*, 4(4), 320-326.
- Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S. ve diğerleri. (2018). Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Annals of PIMS-Shaheed Zulfiqar Ali Bhutto Medical University*, 13(4), 310-315.
- Mehta, G., & Mokhasi, V. (2014). Item analysis of multiple choice questions-an assessment of the assessment tool. *Int J Health Sci Res*, 4(7), 197-202.
- Obon, A. M., & Rey, K. A. M. (2019). Analysis of Multiple-Choice Questions (MCQs): Item and Test Statistics from the 2nd Year Nursing Qualifying Exam in a University in Cavite, Philippines. *In Abstract Proceedings International Scholars Conference*, 7(1), 499-511.
- Patel, R. M. (2017). Use of Item analysis to improve quality of Multiple Choice Questions in II MBBS. *Journal of Education Technology in Health Sciences*, 4(1), 22-29.
- Singh T. (2012). Item analysis and Question Banking In: Principles of Assessment in Medical Education. 1st ed. New Delhi: *Jaypee Brothers Medical Publishers Ltd*.
- Singh, T., Gupta, P., & Singh, D. (2009). Principles of medical education. *Jaypee Brothers Medical Publishers*.
- Tekin, H. (2000). Eğitimde Ölçme ve Değerlendirme. 14. Bas., *Ankara: Yargı Yayıncılık*.
- Topal, K., Aybek, H., Kara, C. O., Büke, A., & Aybek, Z. (2008). Paü tıp fakültesi dönem I öğrencilerine 2006-2007 eğitim ve öğretim Yılında uygulanan çoktan seçmeli sınavların madde ve test analizleri. *Pamukkale Tıp Dergisi*, (3), 120-126.
- Yılmaz H. (2004) Eğitimde ölçme ve değerlendirme, Konya, *Çizgi Kitabevi*.