

# TIP EĞİTİMİNDE SINAV DEĞERLENDİRME SİSTEMİNİN ÖNEMİ: KARABÜK ÜNİVERSİTESİ TIP FAKÜLTESİ TECRÜBESİ

Muhammed Kamil Turan<sup>(1)</sup>, Mehmet Kara<sup>(2)</sup>

(1) Karabük Üniversitesi, Tıp Fakültesi, Tıbbi Biyoloji Ana Bilim Dalı,  
[kamilturan@karabuk.edu.tr](mailto:kamilturan@karabuk.edu.tr) ; *Orcid:0000-0002-1086-9514*

(2) Karabük Üniversitesi, Tıp Fakültesi, Tıbbi Biyokimya Ana Bilim Dalı,  
[mehmetkara@karabuk.edu.tr](mailto:mehmetkara@karabuk.edu.tr); *Orcid:0000-0003-0511-1933*

## ÖZET:

Tıp eğitimi, karmaşık müfredat yapısı ve değerlendirme süreçlerinin çeşitliliğiyle, nitelikli hekim yetiştirme sürecinde kritik bir rol oynamaktadır. Bu çalışma, Karabük Üniversitesi Tıp Fakültesi'nde geliştirilen açık kaynaklı bir sınav değerlendirme sisteminin işlevselliğini incelemektedir. Sistem, sınavların tanımlanması, analiz edilmesi ve sonuçların raporlanması süreçlerini kolaylaştıran kullanıcı dostu bir arayüz sunmaktadır. Python programlama dili kullanılarak geliştirilen yazılım, sınav performans analizlerini detaylı bir şekilde gerçekleştirmek için madde güçlük, ayırt edicilik ve güvenilirlik indeksleri gibi metrikleri kullanmaktadır. Ayrıca, çeldirici etkinliği değerlendirilerek sınav maddelerinin ölçme işlevi optimize edilmektedir. Sistem, Bologna Bilgi Paketi kazanımlarıyla uyumlu çalışarak program hedeflerinin izlenmesini ve eğitim kalitesinin artırılmasını desteklemektedir. Geliştirilen yazılım, hızlı ve güvenilir analizlerin yapılmasını sağlayarak tıp fakültelerinde ölçme-değerlendirme süreçlerine sistematik bir yaklaşım sunmaktadır. Bu yaklaşım, akreditasyon süreçlerinde kalite standartlarını yükseltmek ve eğitim-öğretim faaliyetlerinin etkinliğini artırmak için önemli bir katkı sağlamaktadır.

## GİRİŞ

Tıp eğitimi, ülkemizde 2024 verilerine göre toplam 118 devlet, 106 vakıf üniversitesi olmak üzere toplam 224 üniversitede yapılmaktadır (1). Tıp eğitimi, yapısı gereği Sağlık Bakanlığı (SB), Yükseköğretim Kurulu (YÖK), Tıpta Uzmanlık Kurulu (TUK) ve Ulusal Tıp Eğitimi Akreditasyon Kurulu (UTEAK) gibi temel paydaşlarla yakın iş birliği içinde yürütülmektedir. Bu iş birliği, ulusal hedefler doğrultusunda nitelikli hekim yetiştirme misyonuna doğrudan katkı sağlamaktadır. Tıp fakülteleri, bu hedeflere ulaşmak için preklini, klini ve uygulamalı eğitim standartlarını karşılamakla yükümlüdür. Bu standartların yerine getirilmesi, tıp fakültelerinin eğitim müfredatını oldukça karmaşık ve zorlu bir yapıya dönüştürmektedir (2, 3).

Tıp fakültelerinde birinci, ikinci ve üçüncü sınıflar, komite sistemi ile organize edilen teorik derslerin yoğun olduğu müfredat yapısına sahiptir. Buna karşılık, pratik eğitimlerin ön planda olduğu dördüncü, beşinci ve altıncı sınıfların müfredat gereksinimleri farklılık göstermektedir. Bu farklılıklar, her seviyede uygulanacak ölçme ve değerlendirme yöntemlerinin de çeşitlendirilmesini zorunlu kılmaktadır. Eğitim, bireylerin davranış kazanmasını ve değişen, gelişen dünyaya uyum sağlamalarını sağlayan bir araç olarak tanımlandığında; eğitim-öğretim hedeflerine ulaşma durumunu sayısal verilerle ölçmek, değerlendirmek ve izlemek büyük önem taşımaktadır. Bu bağlamda, eğitim-öğretim hedeflerinin ne ölçüde karşılandığı, çoktan seçmeli sınavlar, yazılı sınavlar, sözlü sınavlar ve uygulama sınavları gibi farklı değerlendirme

yöntemleri ile somut olarak ölçülebilmektedir (4, 5). Tıp eğitimi, bireylerin yalnızca teorik bilgi birikimini değil, aynı zamanda hasta ile etkili iletişim kurma, klinik becerileri uygulama ve meslek etiğine uygun davranma düzeylerini de değerlendirmeyi amaçlamaktadır. Tıp eğitiminin bu çok yönlü yapısı, eğitim sürecinin hem teorik hem de uygulamalı değerlendirme yöntemleriyle desteklenmesini zorunlu kılmaktadır (6). Bu kapsamda Karabük Üniversitesi Tıp Fakültesi'nde tüm sınıflarda eğitim-öğretim amaç ve hedefleri açıkça tanımlanmıştır. Her bir ders konusu için belirlenen kazanımlar, Miller'in (7) tıp eğitiminde öğrenci yeterliliklerini değerlendirmek için geliştirdiği "yeterlilik piramidi" modeline uygun olarak yapılandırılmıştır.

Tıp fakültelerinin eğitim programlarında ölçme ve değerlendirme sistemi, kritik bir bileşen olarak öne çıkmaktadır. Sistemin etkin bir şekilde işleyebilmesi, neyin ve nasıl ölçüleceğine, ayrıca öğrenme hedefleri ile içeriğin doğru bir şekilde yapılandırılmasına bağlıdır. Tıp eğitiminde tıbbi bilgi, klinik beceriler, klinik karar verme yetisi, profesyonel tutumlar ve etik davranışlar gibi yetkinlik alanları değerlendirilmektedir.

Ölçme ve değerlendirme süreçlerinin öğrenci davranışları ve öğrenme sürecine olan etkisi çoğu zaman göz ardı edilmektedir. Öğrencilerin çalışma yöntemleri, sınavların türüne, içeriğine ve uygulanma sıklığına göre şekillenir. Örneğin, sınavlar büyük oranda sadece bilgi hatırlama düzeyini ölçmeye yönelikse, öğrencilerin ezberlemeye dayalı çalışması kaçınılmaz hale gelir. Bu nedenle, sınav türü ve niteliği, öğrenmenin ne şekilde ve hangi aralıklarla gerçekleşeceğini önemli ölçüde belirler. Bunun yanı sıra, ölçme ve değerlendirme araçlarının uygulanabilirliği ve maliyeti de göz önünde bulundurulması gereken faktörlerdir. Çok sayıda uzman gerektiren, uygulanması karmaşık ve maliyeti yüksek yöntemler, sağladıkları katkı maliyetlerini karşılamadığı sürece tercih edilmemektedir.

Öğretim programlarında hedeflenen kazanımlara ne ölçüde ulaşıldığının tespit edilmesi, programın güçlü ve geliştirilmesi gereken yönlerinin belirlenmesi, öğretim sürecinin etkin bir şekilde izlenip değerlendirilmesi ve gerekli iyileştirme adımlarının atılabilmesi için kapsamlı ve sistematik bir değerlendirme mekanizmasına ihtiyaç vardır.

Tıp fakültelerinde ölçme ve değerlendirme süreçlerinin standardizasyonu, eğitim kalitesinin artırılması için kritik öneme sahiptir ve akreditasyon kurumları tarafından da sıkı sıkıya kontrol altında tutulmaktadır. Özellikle modern teknolojiler ve yenilikçi yaklaşımlar, bu süreçlerin daha etkin, şeffaf ve güvenilir bir şekilde yürütülmesine katkı sağlamaktadır.

Bu çalışmanın amacı, Karabük Üniversitesi Tıp Fakültesi'nde geliştirilen ve tıp eğitiminde ölçme-değerlendirme araçlarının etkinliğini artırmayı hedefleyen tamamen açık kaynak kodlu sınav değerlendirme sistemini bir örnek olarak ele almak ve bu sistemin işlevselliğini incelemektir.

## **MATERYAL METOT**

Bu makalede kullanılan bilgisayar programı, AMD Ryzen 7 4800H işlemci, 16 MB RAM ve 250 GB SSD donanım özelliklerine sahip, Arch Linux 6.12.3-arch1-1 sürümünün yüklü olduğu bir bilgisayar üzerinde geliştirilmiştir. Geliştirme dili olarak, hızlı ve esnek yapısının yanı sıra geniş topluluk desteği, açık kaynaklı yazılımlara uygunluğu ve güçlü, standart hale gelmiş paketleri nedeniyle Python programlama dili (sürüm 3.12.7) tercih edilmiştir. Bu çalışmada,

matematiksel işlemler ve matris hesaplamaları için NumPy (sürüm 2.2.0), metin belgeleri ve geniş veri tabloları ile çalışmak için Pandas (sürüm 2.2.3), istatistiksel analizler için SciPy (sürüm 1.14.1) ve Statsmodels (sürüm 0.14.4 stable) kullanılmıştır. Arayüz tasarımı ise PyQt6 (sürüm 6.8.0) kütüphanesi kullanılarak gerçekleştirilmiştir. Karabük Üniversitesi Tıp Fakültesi Eğitim-Öğretim ve Sınav Yönergesi kapsamında, verilerin güvenli bir şekilde depolanması ve ihtiyaç duyulduğunda raporlanabilmesi amacıyla Structured Query Language (SQL) kullanılmıştır.

Geliştirilen arayüz, üç ana bölümden oluşmaktadır. Birinci bölümde, sınavların listelendiği ve veri tabanı üzerinden arama yapılmasına olanak tanıyan sınav listesi yer almaktadır. İkinci bölümde, yeni sınavların tanımlandığı yeni sınav sekmesi bulunmaktadır. Üçüncü bölümde ise sınav performans metriklerinin görüntülediği ve öğrencilerin aldığı notların listelendiği rapor kısmı yer almaktadır.

Birinci bölümde, tanımlanan tüm sınavların dönemi, sınav tipi ve sınav adına göre detaylı arama yapılmasına olanak sağlayan; arama kriterlerine uygun sınavların listelendiği ve seçilen bir sınavın metriklerine, meta verilerine veya analiz sonuçlarına kolayca erişim imkânı sunan sade ve kullanıcı dostu bir arayüz tasarlanmıştır (Şekil 1).

Sıra no	Dönemi	Sınav tipi	Sınav tarihi	Değerlendirme durumu	Sınav adı	Toplam soru sayısı	Toplam puan	Pratik soru puanı	Teorik soru puanı	Cevap anahtarı ön	Yanlış soru işareti	Ham veri dosyası
6	Dönem 3	Mazeret	2024-01-01	Baraj var	Dönem 3 Mazer...	100	100.0	0.0	1.0	Cevap	*	mazeret.txt
4	Dönem 1	Kurul	2024-02-08	Baraj var	Dönem 1 Kurul 3	100	100.0	0.0	1.0	Cevap	*	//home//mkt//P...
5	Dönem 1	Kurul	2024-03-29	Baraj var	Dönem 1 Kurul 4	100	100.0	0.0	1.0	Cevap	*	//home//mkt//P...

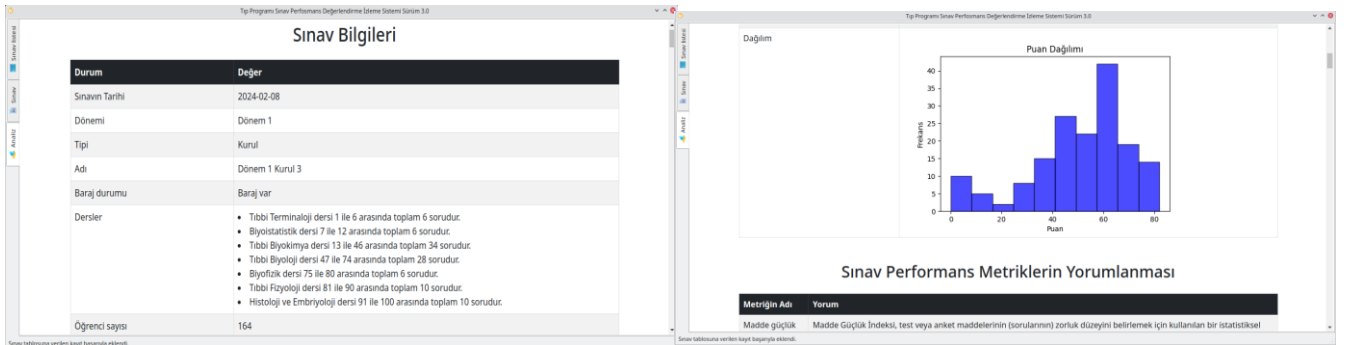
Şekil 1. Tıp Fakültesi Sınav Performans Değerlendirme ve İzleme Sistemi Sürüm 4.0'a ait sınav listesi bölümü. Arayüzün üst sağ kısmında, soldan sağa doğru sıralanmış dört buton yer almaktadır: Filtrele, Yenile, Meta Veriyi Görüntüle ve Seçili Sınavı Analiz Et.

İkinci bölümde, sınav meta verisi kullanılarak yeni sınav ya da sınavların tanımlanabileceği bir arayüz yer almaktadır. Bu bölümde ayrıca, sınav optik formlarının okunmasının ardından oluşan meta verilerin düzenlenmesi için form meta verisi bölümü ve bu bölüm üzerinde arama veya düzeltme yapılmasına olanak sağlayan bir editör kısmı bulunmaktadır. Sınav meta verisi; sınav dönemi, sınav tipi, tarihi, baraj durumu, sınav adı, toplam soru sayısı, toplam puan, pratik puan miktarı, teorik soru puanı ve form meta veri yolu gibi bilgileri içermektedir. Ayrıca, eğer tanımlanan sınav birden fazla dersi kapsıyorsa, bu derslerin listesi ve her birine ait soru sayıları da detaylandırılmıştır (Şekil 2).

Form verilerinin hemen altında, sınavda sorulan sorular ile bu soruların Bologna Bilgi Paketi (BBP) ders öğrenme çıktılarıyla eşleştirildiği ve hizmet ettikleri hedeflere yer verilmiştir. Ek olarak, BBP kapsamında her bir dersin hedeflerinin, program çıktıları üzerindeki etki düzeyleri de tanımlanmıştır.

**Şekil 2.** Tıp Fakültesi Sınav Performans Değerlendirme ve İzleme Sistemi Sürüm 4.0'a ait sınav tanımlama bölümü. Bu bölümde yer alan butonlar ve işlevleri, sol üstten sağ alta doğru sıralandığında şu şekildedir: 1) Sınav adını verilere göre otomatik olarak oluşturur. 2) Form meta verisi için veri yolu tanımlaması yapar. 3) Sınavda yer alan bir dersi listeye ekler. 4) Listede yer alan bir dersi listeden siler. 5) Sınavı kaydetmeden çıkış yapar ve meta veriyi temizler. 6) Sınavı kaydeder. 7) Arama sonuçları arasında gezinmeye olanak tanır.

Üçüncü bölümde, sınav analizleri sonucunda elde edilen verilere yer verilmiştir. Bu veriler, kolay paylaşım sağlamak ve web ortamında rahatlıkla ilan edilebilmek amacıyla HyperText Markup Language (HTML) formatında hazırlanarak kullanıcıya sunulmaktadır (Şekil 3).



**Şekil 3.** Tıp Fakültesi Sınav Performans Değerlendirme ve İzleme Sistemi Sürüm 4.0'a ait rapor bölümü. (a) Raporun başlık kısmına ait bir örnek, (b) Raporun içeriğinin ilerleyen bölümlerine ait bir örnek.

## Sınav performansına dönük metrikler

Bir testin puanlanabilen en küçük bileşeni, ölçme teorisinde madde olarak adlandırılır. Madde, her bir soru için kullanılan temel bir terimdir. Bir ölçme aracının kalitesi, büyük ölçüde onu oluşturan maddelerin niteliğine bağlıdır. Dolayısıyla, bir testin özellikleri ve güvenilirliği, testte yer alan maddelerin kalitesiyle doğrudan ilişkilidir (8).

Madde analizi, bir ölçme aracındaki maddelere verilen cevapları inceleyerek, maddelerin işlevselliği hakkında değerlendirme yapılmasını sağlayan bir süreçtir. Bu süreçte, elde edilen veriler doğrultusunda, her bir maddenin ölçülmek istenen niteliği ne derece yansıttığı belirlenir. Eğer bir madde beklenen performansı göstermiyorsa, bu durumun nedenleri araştırılır ve söz konusu madde, amacına uygun hale getirilmek üzere gerekli düzenlemelere tabi tutulur (8).

Madde güçlük indeksi (MGI), bir sorunun doğru yanıtlanma oranını belirleyen bir ölçüttür. Bu oran, doğru cevap verenlerin sayısının toplam katılımcı sayısına bölünmesiyle hesaplanır. İndeks, 0 ile 1 arasında bir değer alır ve sorunun zorluk derecesini gösterir. Değer 0'a yaklaştıkça sorunun daha zor olduğu, 1'e yaklaştıkça ise daha kolay olduğu söylenebilir. Madde Güçlük Analizine ait matematiksel ifade Denklem – 1 ile gösterilmiştir.

$$MGI = \frac{N_{Doğruyanıtlayansayısı}}{N} \quad (\text{Denklem -1})$$

Madde Güçlük İndeksi sınıflandırması için belirlenen kriterler, maddenin zorluk derecesini değerlendirmede kullanılır. Buna göre, MGI değeri %30'un altında ise madde çok zor, %30 ile %70 arasındaki değerler iyi, kabul edilebilir veya ortalama olarak değerlendirilir. MGI değeri %70'in üzerinde olduğunda madde çok kolay kabul edilir. Ayrıca, MGI değeri %50 ile %60 arasında ise madde, mükemmel ya da ideal olarak nitelendirilir. Bu sınıflandırma, maddelerin ölçme işlevini ne derece yerine getirdiğini analiz etmek için önemli bir ölçüt sağlar (9-11).

Ayırt edicilik indeksi, bir maddenin başarı düzeyi yüksek bireyler ile başarı düzeyi düşük bireyleri ne ölçüde ayırt edebildiğini gösteren bir ölçüttür. Bu indeks, sorunun konuyu iyi anlayan bireylerle anlamayanları ayırma gücünü ifade eder. Yüksek performans gösteren öğrencilerin, bir soruyu düşük performanslı öğrencilere kıyasla daha sık doğru cevaplama beklenir. Bu durumda, madde pozitif bir ayırt edicilik indeksine sahiptir ve bu değer 0.00 ile +1,00 arasında olur. Pozitif indeks, sorunun başarı düzeyi yüksek olan öğrenciler ile düşük olanları ayırt edebildiğini gösterir. Buna karşılık, eğer düşük performans gösteren öğrenciler bir soruyu yüksek performans gösterenlerden daha sık doğru cevaplıyorsa, madde negatif bir ayırt edicilik indeksine sahiptir ve bu değer -1.00 ile 0.00 arasında yer alır (11, 12). Ayırt edicilik indeksi aşağıdaki şekilde elde edilir (Denklem -2):

Cevap kağıtlarından (veya bilgisayardaki eşdeğer verilerden) en yüksek toplam puana sahip %27 ve en düşük toplam puana sahip %27 seçilir. Ara puanlara sahip kağıtlar göz ardı edilir. Belirli bir madde için indeks, üst gruptaki sınav katılımcılarının doğru cevabı verme oranından alt grup için eşdeğer oranın çıkarılmasıyla hesaplanabilir (13).

$$MAI = \frac{N_{Üstdilimdensoğruyapansayısı} - N_{Altdilimdenoğruyapansayısı}}{N} \quad (\text{Denklem -2})$$

MAI indeksi eğer: = Negatif, “kusurlu madde/yanlış cevap anahtarı”, = 0-0,19 ise “zayıf ayırım gücü”, = 0,2 - 0,29 ise “kabul edilebilir ayırım gücü”, = 0,3 - 0,39 ise “iyi ayırım yapabilme gücü”, > 0,4 ise “mükemmel ayırım yapabilme gücü” şeklinde yorumlanır (14, 15).

Madde Güvenirlilik İndeksi ( $IR_x$ ), bir maddenin testin toplam puanı ile ne derece tutarlı olduğunu belirlemek amacıyla kullanılan bir ölçüttür. Bu indeks, maddenin ölçmek istenen özelliği ne kadar doğru bir şekilde ölçtüğünü ve testin geneline ne ölçüde katkı sağladığını değerlendirmeye yarar. Madde güvenilirlik indeksi, Denklem 3'te tanımlanan matematiksel ifadeye göre hesaplanmıştır.

$$IR_x = \frac{(M^{Üstgrupcevapları} - M^{Altgrupcevapları}) \cdot s_x}{N} \quad (\text{Denklem -3})$$

IR ile madde güvenilirlik indeksi, x ile madde numarası, M ile üst ve alt gruplarda doğru cevap verebilenlerin oranı, s ile o maddeye cevap verenlere ait standart sapma değeri ifade edilmektedir. Yorumlama için genel olarak kullanılan değerleri Tablo – 3’de gösterilmiştir (16). Güvenirlilik indeksi değerlerine göre yorumlar şu şekildedir: 0.00 ile 0.20 arasındaki değerler oldukça düşük güvenilirlik seviyesini, 0.21 ile 0.40 arasındaki değerler düşük güvenilirlik düzeyini gösterir. 0.40 ile 0.60 arasındaki değerler güvenilebilir bir seviyeyi ifade ederken, 0.60 ile 0.80 arasındaki değerler ise yüksek güvenilirlik olarak kabul edilir.

Çeldirici etkinliği, bir soruda yer alan çeldiricilerin işlevselliğini değerlendiren bir metriktir. Bu metrik, çeldiricilerin işaretlenme oranının, toplam verilen yanıtlara oranı ile hesaplanır. Bir seçenek, işlevsel bir çeldirici olarak kabul edilebilmesi için sınava katılan öğrencilerin %5 ila %10’u tarafından işaretlenmiş olmalıdır. Eğer bir seçenek, %5’in altında bir oranla işaretlenmişse, bu seçenek işlevsiz çeldirici olarak nitelendirilir (17). Çeldirici etkinliği, Denklem 4’te tanımlanan matematiksel ifade kullanılarak hesaplanmaktadır.

$$DE = \frac{N_{İşlevselçeldirici}}{N_{İşlevsizçeldirici}} \times 100 \quad (\text{Denklem – 4})$$

Çeldirici etkinliği değerlerine göre yapılan yorumlar şu şekildedir: Çeldirici etkinliği %0 ise zayıf, %33,3 ise kabul edilebilir, %66.6 ise iyi ve %100 ise mükemmel olarak değerlendirilir (18).

## SONUÇ ve TARTIŞMA

Akreditasyon süreçlerinde eğitim-öğretim kalite standartlarını yükseltmek amacıyla, sınav sorularının (örneğin; not dağılımı, en yüksek ve en düşük notlar, madde analizi gibi) ve ders öğrenme kazanımlarının detaylı analizlerinin yapılması gerekmektedir. Bu analizler, eğitim programının etkinliğini değerlendirmek ve gerekli iyileştirme çalışmalarını gerçekleştirmek için önemli bir temel oluşturur.

Madde analizi, öğrencilerin verdiği yanıtların incelenmesi ve elde edilen verilerin yorumlanması yoluyla test maddelerinin kalitesi hakkında bilgi sağlayan, nispeten basit ancak oldukça değerli bir yöntemdir. Bu prosedür, maddelerin ölçme işlevini ne derece yerine getirdiğini anlamak ve testin genel geçerliliğini değerlendirmek için önemli bir araç olarak

kullanılır. Bu çalışmada, öğrencinin akademik alandaki öğrenme düzeyini değerlendirmede etkili bir araç olarak görülen tek en iyi yanıt tipi çoktan seçmeli soruların madde analizi yapılmıştır. Bu analizlerin gerçekleştirilmesi ve değerlendirme süreçlerinin daha sistematik bir şekilde yürütülebilmesi amacıyla, bilgisayar tabanlı bir yazılım geliştirilmiştir. Geliştirilen bu yazılım, madde analizinin hızlı, doğru ve güvenilir bir şekilde yapılmasına olanak sağlamaktadır.

Sorulara itirazların öğrenciler tarafından her zaman eksiksiz bir şekilde yapılmaması, soru okuma toplantılarının düzenli olarak iş yükü nedeni ile icra edilememesi gibi çeşitli nedenler; madde analizi sonuçlarına ihtiyacı daha da arttırmıştır. Bu yazılım ile performans metrikleri daha detaylı incelenmiş ve çeşitli yanlışlıklar düzeltilmiştir.

Bu sistem, akreditasyon süreçlerinde eğitim standartlarının yükseltilmesine ve ders kazanımlarının etkin bir şekilde değerlendirilmesine önemli bir katkı sunarak, tıp fakültelerinde kalite odaklı bir değerlendirme kültürünün oluşturulmasına olanak tanımaktadır.

#### **KAYNAKLAR:**

1. YÖK. Tıp Fakültesi Lisans Programı [Available from: <https://yokatlas.yok.gov.tr/lisans-bolum.php?b=10206>].
2. Tokuç B, Varol G. COVID-19 PANDEMİSİ ve TÜRKİYE'DE TIP EĞİTİMİ. Namık Kemal Tıp Dergisi. 2020;8(3):595-9.
3. Odabaşı O. Türkiye Tıp Fakülteleri 2023. Sürekli Tıp Eğitimi Dergisi. 2023;32(1):37-61.
4. Anıl D. Tıp Eğitiminde Ölçme ve Değerlendirme. SD (Sağlık Düşüncesi ve Tıp Kültürü) Dergisi. 2015;33:82-7.
5. Mıdık Ö, Karabilgin ÖS. TIP EĞİTİMİNDE ÖĞRENCİ DEĞERLENDİRMELERİNDE STANDART BELİRLEME: SANAL BİR SINAV KURGUSUNDA GEÇME KALMA SINIRI. Tıp Eğitimi Dünyası. 2011;29(29):21-33.
6. Buluta İ, Sapancı A, Kara İH. Tıp eğitiminde kullanılan ölçme ve değerlendirme araçlarının geleneksel ve alternatif ölçme ve değerlendirme araçları olarak sınıflandırılması. Journal of Medical Education And Informatics. 2015;1(1):2-11.
7. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9 Suppl):S63-7.
8. Hasaıçebi B, Terzi Y, Küçük Z. Madde güçlük indeksi ve madde ayırt edicilik indeksine dayalı çeldirici analizi. Gümüşhane Üniversitesi Fen Bilimleri Dergisi. 2020;10(1):224-40.
9. Christian DS, Prajapati AC, Rana BM, Dave VR. Evaluation of multiple choice questions using item analysis tool: a study from a medical institute of Ahmedabad, Gujarat. Int J Community Med Public Health. 2017;4(6):1876-81.
10. Date AP, Borkar AS, Badwaik RT, Siddiqui RA, Shende TR, Dashputra AV. Item analysis as tool to validate multiple choice question bank in pharmacology. International Journal of Basic & Clinical Pharmacology. 2019;8(9):1999-2003.
11. Rao C, Kishan Prasad H, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. Int J Educ Psychol Res. 2016;2(4):201-4.

12. Taib F, Yusoff MSB. Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. *Journal of Taibah University Medical Sciences*. 2014;9(2):110-4.
13. Burton RF. Do item-discrimination indices really help us to improve our tests? *Assessment & Evaluation in Higher Education*. 2001;26(3):213-20.
14. Marsevani M. Item analysis of multiple-choice questions: an assessment of young learners. *English Review: Journal of English Education*. 2022;10(2):401-8.
15. Mitra N, Nagaraja H, Ponnudurai G, Judson J. The levels of difficulty and discrimination indices in type A multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. *IeJSME*. 2009;3(1):2-7.
16. Pradana PW, Febriani F, Ibnusaputra M, Jumadi J. Development of Physics Test Instrument to Measure Verbal Representation of High School Student on Optical Instrument Topic. *Jurnal Penelitian Pendidikan IPA*. 2023;9(10):7963-8.
17. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Med Educ*. 2009;9:40.
18. Chauhan GR, Chauhan BR, Vaza JV, Chauhan PR. Relations of the Number of Functioning Distractors With the Item Difficulty Index and the Item Discrimination Power in the Multiple Choice Questions. *Cureus*. 2023;15(7):e42492.