

Ege Üniversitesi Tıp Fakültesi Öğrenci ve Öğretim Üyesi Değerlendirme Formları ile Yapılan Ölçümlere İlişkin Geçerlilik - Güvenilirlik

Reliability and Validity of the Measurements Done by Using Students' And Teachers' Evaluation Forms in Ege University School of Medicine

¹Kevser Vatanserver (Orcid: 0000-0002-8943-9874) ²Şöhret Aydemir (Orcid: 0000-0001-8354-9100)

³A. Hilal Batı (Orcid: 0000-0002-8781-6816) ⁴Cenk Can (Orcid: 0000-0002-1992-5367)

⁵Mahmut Çoker (Orcid: 0000-0001-6494-9539) ⁶Selda Erensoy (Orcid: 0000-0002-7052-8359)

⁷Figen Gövsa (Orcid: 0000-0001-9635-6308) ⁸Özen Kaçmaz Başoğlu (Orcid: 0000-0001-8168-6611)

⁹Lütfiye Kanıt (Orcid: 0000-0001-5160-411X) ¹⁰Nilgün Kültürsay (Orcid: 0000-0003-0867-1514)

¹¹Oktay Nazlı (Orcid: 0000-0002-3748-9941) ¹²Eser Sözmen (Orcid: 0000-0002-6383-6724)

¹³Sıla Elif Türün (Orcid: 0000-0001-9022-6457) ¹⁴Meltem Çiçeklioğlu (Orcid: 0000-0002-7059-7573)

¹Dr. Emekli, Ege Üniversitesi Tıp Fakültesi Program Değerlendirme Kurulu Üyesi (2004-2017)

²Prof. Dr. Ege Üniversitesi Tıp Fakültesi Tıbbi Mikrobiyoloji Anabilim Dalı, Program Değerlendirme Kurulu Üyesi (2017-2018), Eğitim Koordinatörü (2017- 2018)

³Doç. Dr. Ege Üniversitesi Tıp Fakültesi Tıp Eğitimi Anabilim Dalı, Program Değerlendirme Kurulu Üyesi

⁴Prof. Dr. Ege Üniversitesi Tıp Fakültesi Tıbbi Farmakoloji Anabilim Dalı, Program Değerlendirme Kurulu Üyesi

⁵Prof. Dr. Ege Üniversitesi Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı, Program Değerlendirme Kurulu Başkanı (2015-2019)

⁶Prof. Dr. Ege Üniversitesi Tıp Fakültesi Tıbbi Mikrobiyoloji Anabilim Dalı, Program Değerlendirme Kurulu Üyesi, Eğitimden Sorumlu Dekan Yardımcısı (2015-2017)

⁷Prof. Dr. Ege Üniversitesi Tıp Fakültesi Anatomi Anabilim Dalı, Program Değerlendirme Kurulu Üyesi

⁸Prof. Dr. Ege Üniversitesi Tıp Fakültesi Göğüs Hastalıkları Anabilim Dalı, Program Değerlendirme Kurulu Üyesi (2017-2018), Eğitimden Sorumlu Dekan Yardımcısı (2017-2018)

⁹Prof. Dr. Ege Üniversitesi Tıp Fakültesi Fizyoloji Anabilim Dalı, Program Değerlendirme Kurulu Üyesi

¹⁰Prof. Dr. Ege Üniversitesi Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı, Program Değerlendirme Kurulu Üyesi

¹¹Prof. Dr. Ege Üniversitesi Tıp Fakültesi Üroloji Anabilim Dalı, Program Değerlendirme Kurulu Üyesi

¹²Prof. Dr. Ege Üniversitesi Tıp Fakültesi Tıbbi Biyokimya Anabilim Dalı, Program Değerlendirme Kurulu Üyesi

¹³Dr. Ege Üniversitesi Tıp Fakültesi Tıp Eğitimi Anabilim Dalı Program Değerlendirme Kurulu Üyesi

¹⁴Prof. Dr. Ege Üniversitesi Tıp Fakültesi Halk Sağlığı Anabilim Dalı, Program Değerlendirme Kurulu Başkanı (2019-devam ediyor)

Sorumlu Yazar: Kevser Vatanserver

Künye: Vatanserver K, Aydemir Ş, Batı H, Can C, Çoker M, Erensoy S ve ark. Ege Üniversitesi Tıp Fakültesi Öğrenci ve Öğretim Üyesi Değerlendirme Formları ile Yapılan Ölçümlere İlişkin Geçerlilik - Güvenilirlik. Tıp Eğitimi Dünyası. 2020;19(57):37-54

Anahtar Sözcükler:

program değerlendirme, iç tutarlılık, yapı geçerliliği, öğrenci değerlendirme formu, öğretim üyesi değerlendirme formu,

Keywords:

program evaluation, internal consistency, construct validity, student evaluation form, teacher evaluation form

Gönderilme Tarihi *Submitted:* 26.08.2019

Kabul Tarihi

Accepted: 27.12.2019

ÖZET:

Giriş: Program değerlendirme etkinlikleri, Ege Üniversitesi Tıp Fakültesi'nde 2001 yılında başlayan program geliştirme çalışmalarının önemli bir bileşeni olmuştur. Bu çalışmanın amacı Ege Üniversitesi Tıp Fakültesi'nde eğitim programını değerlendirmek için kullanılan güncellenmiş öğrenci ve öğretim üyesi blok ve staj değerlendirme formları ile yapılan ölçümlerin geçerlilik ve güvenilirliklerinin belirlenmesidir.

Gereç ve Yöntem: Eğitim programını değerlendirmek için kullanılan ve en son 2017 Ocak-Şubat aylarında güncellendikten sonra 2017 Nisan-Haziran arasında pilot olarak uygulanan öğrenci ve öğretim üyesi blok ve staj değerlendirme formları ile yapılan ölçümlerin sonuçlarının güvenilirlik ve geçerliliği, metodolojik tasarım tipinde bir çalışmayla değerlendirilmiştir. İç tutarlılık güvenirliliğinin belirlenmesinde Cronbach'ın alpha katsayısı hesaplanmıştır. Puanlayıcılar arası uyum ve güvenilirlik için sınıf içi korelasyon katsayısı kullanılarak tutarlılık araştırılmıştır.

Kapsam geçerliliği için madde kapsam geçerlilik oranları, yapı geçerliliği için açılımlayıcı faktör analizi kullanılmıştır.

Bulgular: Cronbach'ın alpha değerlerinin İkinci ve Üçüncü Sınıf Blok Değerlendirme formunun iki faktörü dışında 0,7 üzerinde bulunması,

güncellenen öğrenci ve öğretim üyesi değerlendirme formlarının sonuçlarıyla ilgili iç tutarlılık güvenirliliğini desteklemektedir. Yapı geçerliliği analizinde öğrenci formlarının, İkinci ve Üçüncü Sınıf Blok Değerlendirme Formu dışında tek boyutlu olduğu, Öğretim Üyesi Değerlendirme Formu ise üç faktörlü yapıya sahip olduğu bulunmuştur.

İkinci ve Üçüncü Sınıf Blok Değerlendirme formunda Faktör II, Öğretim Üyesi Blok ve Staj Değerlendirme Formunda ise Faktör III için puanlayıcılar arası uyum ve güvenilirlik katsayılarının kabul edilebilir düzeylerin altında olması bu boyutlar için yapılan ölçüme ilişkin güvenirliliği desteklememektedir.

Sonuç: Bu çalışma EÜTF'nde kullanılan güncel değerlendirme formlarıyla yapılan ölçümlerin geçerlilik ve güvenirliliğini destekleyen kanıtların yanı sıra, sonuçların yorum ve kullanımında dikkatli olunması gereken noktaları da ortaya koymuştur.

Çalışma ayrıca, program değerlendirme formlarının geçerlilik ve güvenirliliğinin çok yönlü değerlendirilmesi açısından ülkemizdeki tıp fakültelerine örnek teşkil edilebilir.

ABSTRACT:

Introduction: Program evaluation has been a crucial component of program development activities in Ege University School of Medicine since 2001. Aim of this study is to determine validity and reliability of the measurements done by using updated student and teachers block and clerkship evaluation forms in Ege University School of Medicine.

Materials and Methods: In this methodological design study, reliability and validity of the results obtained by students' and teachers' evaluation forms were assessed, which were updated in 2016-2017 academic term between January-

February 2017 and implemented on April-June 2017 as a pilot. For internal consistency Cronbach's alpha coefficient was calculated. For interrater agreement and reliability, consistency was examined by calculating intraclass correlation coefficient. For content validity; item content validity ratios and for construct validity exploratory factor analysis were used.

Results: Cronbach's alpha coefficients that were over 0,7, except for Second and Third Year Student Block Evaluation Form, corroborated the internal consistency reliability for updated student and teacher evaluation forms. In construct analysis, student forms, except the Second and Third Year Block Evaluation Form were found to have a one-dimension construct, while teacher forms were determined to have a three-dimensions construct.

Reliability of measurement of the Factor II of Second and Third Year Block Student Forms and Factor III of Teachers' Evaluation Forms was not corroborated as the interrater agreement and reliability coefficients were below acceptable levels.

This study revealed evidences that support the validity and reliability of measurements done by current evaluation forms in Ege University School of Medicine, and also detected the points to be paid attention while interpreting and using the results.

Also it can be a reference for medical schools in our country, for versatile evaluation of validity and reliability of program evaluation forms.

GİRİŞ

Diğer tüm okullarda olduğu gibi tıp fakülteleri için de program değerlendirilmesinin temel amacı programların geliştirilmesinde rehberlik edecek verileri sağlamaktır. İdeal bir değerlendirme yöntemi geçerli, güvenilir, kabul edilebilir

ve düşük maliyetli olmalıdır. Bu yaklaşım benimsendiğinde değerlendirme sürecinin kendisi tıp fakültesi eğitim programı üzerinde olumlu bir eğitsel etki gösterir (1). Program değerlendirmede yan tutmayı azaltmak önemlidir; bu amaçla öğrenciler, mezunlar, akranlar, eğitim almış gözlemciler, eğitimciler ya da hastalar gibi farklı grupların görüşleri alınabilir (1, 2).

Değerlendirmede kullanılacak araçları geliştirmek için en iyi çözüm, var olanların, fakültenin kendi katkılarıyla yeni değerlendirme gereksinimlerine göre uyarlanmasıdır (3).

Öğrenci ve öğretim üyesi geri bildirimlerinden elde edilecek verilerin gerçek durumu yansıtabilmesi için, kullanılan formların psikometri çalışmaları yapılarak, geçerlilik ve güvenilirliklerinin kanıtlanmış olması gereklidir (4, 5). Öğrenci, öğretim üyesi geri bildirimleri alındıktan sonra sıra verileri yorumlamaya geldiğinde kullanılan araçların orijini ne olursa olsun sonuçlara güvenilip güvenilmeyeceğini bilmek gerekir. Bunun yolu güvenilirlik ve geçerliliğinin desteklenmesidir (3).

Güvenilirlik test veya ölçek sonuçlarının kavramsal yapıya ilişkin olguyu doğru bir şekilde ortaya çıkarması; ölçüm aracının farklı yer ve zamanlarda aynı popülasyondan seçilen farklı örnek gruplarına uygulandığında benzer sonuçları vermesi olarak tanımlanmaktadır (6). Geçerlilik ise kullanılan ölçüm aracının ölçülmek istenen özelliğe uygun olması, verilerin ölçülecek özelliğin niteliğini yansıtması ve verilerin amaca yönelik olarak yararlı olması şeklinde tanımlanmaktadır (6).

Ülkemizde tıp fakültesi eğitim programlarını değerlendirmek için Ulusal Tıp Eğitimi Akreditasyon Kurulu tarafından oluşturulan mezuniyet öncesi tıp eğitimi ulusal standartları içinde de program değerlendirilmesinin temel ve gelişim standartları ayrıntılı olarak

tanımlanmıştır. Bu metinde yer alan program değerlendirme standartları içinde tıp fakültelerinin program değerlendirme süreci ve sonuçlarının geçerlilik ve güvenilirlik kanıtlarını toplaması ve göstermesinin gerekliliği belirtilmiştir. Böylece kullanılan değerlendirme yönteminin ölçülmek istenen değişkeni ölçüp ölçmediği ve ne derece tutarlılık ve duyarlılıkla ölçebildiği belirlenir (5).

Ege Üniversitesi Tıp Fakültesinde (EÜTF) sistematik program değerlendirme ve izlem yolu ile sürekli gelişim sağlamak üzere, 2002-2003 eğitim öğretim yılından beri ilk üç yıl blokları ile 4, 5 ve 6. sınıf staj blokları/stajları için öğrencilerin değerlendirmeleri alınmaktadır.

Fakülte Eğitim Komisyonu, 2000 yılında öncelikle staj programlarındaki olumlu ve geliştirilmesi gereken yönleri “işaret” edebilecek bir program değerlendirme projesi (İşaret Projesi) sürdürmek üzere bir komisyon görevlendirmiştir (7). Komisyon önce bir ölçek geliştirme çalışması sürdürmüş ve her staj rotasyonu sonrası veri toplanması, analizi, raporlama ve sonuçların kullanımı yönüne uygun bir çalışma başlatmıştır (7). İzleyen 2002-2003 akademik yılında, program değerlendirme etkinlikleri, ilk üç yıl bloklarını da kapsayacak şekilde genişletilmiştir. Öğretim üyelerinin de görüşlerini almak amacıyla 2004 - 2010 yılları arasında iki kez güncellenen öğretim elemanı ilk üç yıl program değerlendirme formları kullanılmış ancak sistematik veri toplama amacına ulaşamamıştır. Bu nedenle 2016-2017 akademik yılı sonunda öğretim üyelerinin ilk üç yıl ve staj eğitim programlarına yönelik iki ayrı formla yıllık değerlendirmelerinin alınacağı bir uygulamaya geçilmiştir.

Öğrenci ilk üç yıl blokları ile 4. ve 5. sınıf staj değerlendirme formları, 2016-2017 akademik yılına dek üçer kez güncellenmiştir.

Bu çalışmada Ege Üniversitesi Tıp Fakültesi’nde

ilk üç yıl blokları ile dördüncü - beşinci sınıf staj bloklarında yürütülmekte olan eğitim programlarını değerlendirmek için güncellenen öğrenci ve öğretim üyesi blok ve staj değerlendirme formları ile yapılan ölçümlerin geçerlilik ve güvenilirliklerinin belirlenmesi amaçlanmıştır. Araştırma soruları şu şekildedir:

1. EÜTF’nde eğitim programını değerlendirmek için kullanılan öğrenci blok/staj değerlendirme formları aracılığıyla elde edilen sonuçlar geçerli ve güvenilir midir?
2. EÜTF’nde eğitim programını değerlendirmek için kullanılan öğretim üyesi blok/staj değerlendirme formları aracılığıyla elde edilen sonuçlar geçerli ve güvenilir midir?

GEREÇ ve YÖNTEM

Metodolojik tasarım tipinde planlanan bu çalışmada öğrenci ve öğretim üyesi değerlendirme formlarının geliştirilmesi ve psikometrik değerlendirme süreçleri dört bölüm halinde sunulmuştur (8, 9).

1. Aşama: Madde havuzunun oluşturulması
2. Aşama: Kapsam geçerliliğinin değerlendirilmesi
3. Aşama: Formların uygulanması
4. Aşama: Yapılan ölçümlere ilişkin geçerlilik ve güvenilirlik analizi

İlk üç aşama öğrenci ve öğretim üyesi değerlendirmesi için iki ayrı bölüm halinde verilmiş, son aşama gereç yöntem açısından tüm değerlendirme formlarında benzer olduğundan aynı bölümde ele alınmıştır.

I. Öğrenci Blok ve Staj Değerlendirme Formları için ilk üç aşama

Madde havuzunun oluşturulması

Bu aşamada EÜTF program değerlendirme sisteminde Bushnell’in (10) önerdiği girdi-

süreç-çıkıtı özelliklerinin değerlendirilmesi yaklaşımı temelinde belirlenmiş olan değerlendirme parametreleri yeniden gözden geçirilmiştir. Literatürden elde edilen eğitim değerlendirmesine yönelik başlıklar temelinde hazırlanmış olan EÜTF Program Değerlendirme Matrisindeki değerlendirme başlıkları şu şekildedir (11, 12):

- Girdi - Süre, öğrenciye sunulan olanaklar, konu başlıkları ve beceriler
- Süreç - Öğrenciye yapılan açıklamalar, ilan edilen programa uyum, öğrenciye sunulan öğrenme fırsatları, eğitim etkinliklerinin niteliği, öğretim üyeleri ile iletişim, sınav yöntemleri.
- Çıktı - Eğitimin mesleki açıdan yararlılığı, eğitimin zayıf ve güçlü yönlerine ilişkin öğrenci önerileri.

Bu başlıklarda, önceki yıllarda kullanılan öğrenci blok ve staj değerlendirme formlarında yer alan maddeler yeniden düzenlenmiş ve ilk üç yıl için 18'i, stajlar için 23'ü seçilerek madde havuzları oluşturulmuştur.

Kapsam geçerliliğinin değerlendirilmesi

Kapsam geçerliliğinin değerlendirilmesi, Program Değerlendirme Kurulu üyelerinden oluşan uzman panelinde 5'li Likert ölçeği (1: kesinlikle olumsuz. 2: olumsuz. 3: ne olumlu ne olumsuz. 4: olumlu. 5: kesinlikle olumlu FY: fikrim yok) kullanılan ve maddelerin gereklilik ve anlaşılabilirliğini sorgulayan iki soru ile gerçekleştirilmiştir. Ocak/Şubat 2017 aylarında yapılan dört toplantıda birinci sınıf öğrenci blok değerlendirme formu için 12, ikinci ve üçüncü sınıf öğrenci blok değerlendirme formu için 13 maddenin yeterli olduğuna karar verilmiştir. İkinci ve üçüncü sınıf öğrenci blok değerlendirme formuna mesleki yabancı dil eğitimini sorgulayan bir madde eklenmiştir.

Staj değerlendirme formu için de 16 madde seçilmiştir. Üç form için de seçilen maddeler, Eğitim Komisyonunun bir toplantısında görüşülüp geçerliliği (9) açısından değerlendirilmiş, önerilere uygun değişiklikler yapılmıştır. İlk üç yıl formlarında "Blok kapsamındaki teorik ders sayısı yetersizdi", staj formunda "Teorik dersler daha fazla olmalıydı" soruları kontrol sorusu olarak belirlenmiştir. Bu maddelerin her birinin "(0- fikrim yok, 1- kesinlikle katılmıyorum 9- kesinlikle katılıyorum)" seçenekleri üzerinden yanıtlanması benimsenmiştir.

Formun uygulanması

Tüm öğrenci değerlendirme formları internet ortamına yüklenerek öğrenci.med.ege.edu.tr sayfasında erişime açılmıştır. İlk üç yıl öğrenci blok değerlendirme formları 2016-2017 akademik yılının son blokları boyunca, dördüncü ve beşinci sınıf staj blokları için son rotasyonları süresince, 2017 Nisan-Haziran arasında öğrenci.med.ege.edu.tr adresinde erişilebilir halde sunulmuştur. Öğrenciler blok/staj boyunca değerlendirmelerini yapabilmektedirler. Blok/staj sonunda sınav sonuçlarını öğrenebilmek için değerlendirme formlarını tamamlamış olmaları gerekmektedir. Bu nedenle tüm öğrenci değerlendirme formları için doldurma oranları % 100'dür.

II. Öğretim Üyesi Blok ve Staj Değerlendirme Formları için ilk üç aşama

Madde havuzunun oluşturulması EÜTF program değerlendirme sisteminde temel alınan parametreler gözden geçirilmiş, yeni öğretim üyesi değerlendirme formlarında girdi ve süreç özelliklerine ait 14 maddeye yer verilmiştir:

- Girdi - Öğrenci sayısı, öğrencilerin bilgi ve becerisi, programın yapısı ve organizasyonu, sınav sistemi ve işleyişi, derslikler ve

laboratuvarlar, öğrenci işleri bürosu, sınıf ve laboratuvar görevlileri, anabilim dalı sekreterleri, diğer anabilim dalları ve öğretim üyeleri, fakülte yönetimi

• Süreç - Program işleyişi, program değişiklikleri hakkında bilgilendirme, blok/staj bloğu yürütme kurulunun öğretim üyeleri ile iletişimi, öğrenci geri bildirimleri

Kapsam geçerliliğinin değerlendirilmesi

Formlar Program Değerlendirme Kurulu üyeleri ve bu üyelerle aynı anabilim dallarında çalışan toplam 12 öğretim üyesinin erişimine açılarak kapsam geçerliliği değerlendirilmiştir. Maddelerin gerekliliği ve anlaşılabilirliği konusunda katılımcılar tüm maddeleri anlaşılır bulmuş, bir üye bir maddenin gereksiz olduğunu belirtmiştir (Kapsam Geçerlilik Oranları: % 91.7-100). İnternet ortamındaki form, görünüş geçerliliğinin değerlendirilmesi amacıyla Eğitim Komisyonu üyelerine gönderilmiş, blok ve staj değerlendirme formları sırasıyla sekiz ve dokuz öğretim üyesi tarafından değerlendirilmiştir. Katılımcıların önerileri üzerine ölçek değerlendirme şekli dikotom ölçekten (olumlu/olumsuz) beşli Likert ölçeğine (1: kesinlikle olumsuz. 2: olumsuz. 3: ne olumlu ne olumsuz. 4: olumlu. 5: kesinlikle olumlu FY: fikrim yok) dönüştürülmüştür. Madde havuzundan seçilen 14 maddenin tümü formda yer almıştır.

Formun uygulanması

Öğretim üyelerinin 2016-2017 akademik yılı blok ve staj değerlendirmeleri, 2017 Haziran ayı boyunca EÜTF Öğrenme Kaynakları Merkezi'nde internet üzerinden Moodle programı aracılığıyla alınmıştır. Öğretim üyesi blok değerlendirme formları 96, staj değerlendirme formları 88 öğretim üyesi tarafından yanıtlanmıştır.

III. Öğretim Üyesi ve Öğrenci Değerlendirme Formları için dördüncü aşama

Ölçüme ilişkin geçerlilik ve güvenilirlik analizi

Yapılan ölçüme ait iç tutarlılık güvenilirliğinin belirlenmesinde Cronbach'ın alpha güvenilirlik katsayısı hesaplanmış, her bir madde ile toplam skor arasındaki ilişki madde-toplam korelasyon katsayıları kullanılarak değerlendirilmiştir. Güvenilirlik için, Cronbach'ın alpha güvenilirlik katsayısının 0,7 veya üzerinde, madde toplam korelasyonun 0,3'ün üzerinde olması gerekmektedir (6).

Güvenilirlik için Cronbach'ın alpha değeri ile incelenen iç tutarlılığın yanı sıra, farklı gözlemciler tarafından verilen kararların yorum ve kullanımına ilişkin olarak gözlemciler ya da puanlayıcılar arası uyum (interrater agreement) ve güvenilirlik (interrater reliability) ayrıca araştırılmıştır.

IV. Tüm formlar için puanlayıcılar arası uyum ve güvenilirlik analizleri

Genel bilgi

Bu çalışmada formlar aracılığıyla bireylerden (öğrenci ve öğretim üyelerinden) veri toplamadaki amaç, bu bireylere ait belirli bir özelliğin ölçülmesi değil, bu bireylerin yaptıkları gözlem/değerlendirmeye dayalı olarak her bir maddeye verdikleri puanlar üzerinden yürütülen eğitimin belirli özelliklere sahip olup olmadığına ilişkin niceliksel geri bildirim elde edilmesidir. Bu bağlamda puanlayıcılar, kişisel tutum ya da davranışlarından çok, belirli bir özellikle ilişkili kendi algı/gözlemlerini bildirdikleri için birey olarak kendileri hakkında bilgi veren cevaplayıcılardan ayrı olarak ele alınmalıdır (13). Diğer bir deyişle gözlem birimi (unit of observation) bireyler olsa da; analiz

birimi (unit of analysis) ‘birey’lerin kendisi değil, yürütülen eğitimin belirli özelliklerine ilişkin öğrenci ya da öğretim üyesi ‘grup’larının yapmış oldukları gözlem/değerlendirmedir (13-17). Bu nedenle, ölçüm sonuçlarının yorum ve kullanımına temel oluşturan veri analizinde, her bir öğrenci/öğretim üyesi için form aracılığıyla elde edilen “birey düzeyindeki” toplam puan yerine, tüm öğrenci/öğretim üyelerinden elde edilen ‘grup düzeyindeki’ puanlar dikkate alınmaktadır (14). Öğrenci/öğretim üyelerinin değerlendirme/gözlemlerini yansıtan bireysel puanlar birleştirilerek (aggregated) grup düzeyinde hesaplanan ortalama ve ortanca değerleri, değerlendirme/gözlem yapanların tamamı üzerinden gözlenen özelliğe atfedilmektedir. Birçok kişiden elde edilen cevapların, veri analizinde kullanılmak üzere, tek bir cevap gibi ele alınabilmesi için puanlayıcılar arasında belirli düzeyde uyum ya da fikir birliğinin bulunmasının bir ön koşul olduğu belirtilmektedir (13).

Puanlayıcılar arası uyum ve güvenilirlik ölçütleri

Geri bildirim formları aracılığı ile yapılan ölçümlerin yorum ve kullanımında temel hata kaynağının, maddeler arası uyumun yokluğundan ziyade puanlayıcılar arasındaki uyum ya da tutarlılığın yokluğu ile ilişkili olduğu belirtilmektedir (17). Öğrenci geri bildirim formları aracılığıyla elde edilen ölçümlere ilişkin güvenilirliğin araştırılmasında puanlayıcılar arası uyum (interrater agreement-IRA), puanlayıcılar arası güvenilirlik (interrater reliability-IRR), puanlayıcılar arası güvenilirlik ve uyuma (IRR+IRA) ilişkin göstergelerin incelenmesinin gerekliliği özellikle vurgulanmaktadır (13, 16-19).

Çalışmamızda farklı gözlemciler tarafından verilen kararların birey düzeyinde değil de, grup

düzeyinde birleştirilmiş (aggregated) olarak dikkate alınması nedeniyle bu birleştirmenin doğrulanması ya da dayanağının gösterilmesinde ayrıca IRA için rWG(J) ve ortalama sapma (average deviation-AD) değerleri araştırılmıştır. Bu çalışmada, tutarlılık ya da güvenilirlik için önerilen diğer yöntemlerin (ortalama madde puanları için sınıf içi korelasyon katsayısı / intraclass correlation coefficient-ICC, ortalama uyum oranları ya da puanlayıcı profilleri arası ortalama korelasyon vb.) ya da tek başına IRR’nin kullanılmamasının nedeni, yapılan başka çalışmalarda veriler aşık uyum ya da tutarlılık gösterdiğinde bile bu yöntemlerle düşük, negatif ya da yorumlamaya elverişsiz sonuçların elde edildiğinin gösterilmiş olmasıdır (20).

Hesaplanan rWG(J) değerleri için beklenen varyansın belirlenmesinde seçilen dağılım, orta seviyede çarpık madde puanı dağılımıdır (16). Puanlayıcılar arası güvenilirliğe ilişkin olarak ICC için iki yönlü karma etki modeli kullanılarak % 95 güven aralığında tutarlılık araştırılmıştır.

Formların faktör yapısına bağlı olarak bu değerler, formun tamamı ve her bir faktör için ayrı ayrı hesaplanmıştır.

Grup içi uyum (within group agreement) açısından gözlemciler tarafından verilmiş olan puanların uyumuna ilişkin bir gösterge olan ve 0 ile 1 arasında bir değeri bulunan rWG(J) için 0,00-0,30 arası değerler uyumun yokluğunu; 0,31-0,50 arası değerler zayıf uyum, 0,51 ve 0,70 arası değerler orta uyum, 0,71-0,90 güçlü uyum, 0,91-1,00 arasındaki değerler ise çok güçlü uyum varlığını belirten değerler olarak yorumlanmaktadır (21).

Puanlayıcılar arası uyum açısından araştırılmış olan AD değerlerinin yorumlanmasında maddelere ait cevap kategorilerinin düzeyi dikkate alınmaktadır (22, 23). Dokuzlu kategori

için 1,5; 5'li kategori için ise 0,8'in altındaki AD değerleri yeterli uyum göstergesi olarak kabul edilmektedir (22). Puanlayıcı sayısının 100'den fazla olduğu durumlarda ise 2.01'in altındaki AD değerlerinin de yeterli uyuma işaret edebileceği gösterilmiştir (22, 23). ICC değerlerinin 0,5'ten az olmasının zayıf, 0,5-0,75 arasında olmasının ise orta derecede güvenilirlik olarak yorumlanabileceği belirtilmektedir (24).

Tüm formlar için yapı geçerliliği analizleri

Formların yapı geçerliliğinin belirlenmesi için açımlayıcı faktör analizi kullanılmış, temel bileşenler analizi seçilmiş ve eğik döndürme yöntemi olan direkt oblimin döndürme uygulanmıştır. Direkt oblimin döndürme yönteminin seçilmesinin nedeni, form maddelerinin birden fazla faktöre dağılacakları konusunda bir öngörü olmaması ve faktörler arası korelasyon katsayısının 0,32'den büyük olmasıdır (6). Maddelerin faktör analizi yapmaya yetecek düzeyde korelasyonu olup olmadığını değerlendirmek üzere Kaiser-Meyer-Olkin (KMO) ve Bartlett küresellik testi analizleri yapılmış, KMO değeri >0,5 ve Bartlett küresellik testi p değeri <0,05 ise anlamlı olarak kabul edilmiştir.

Toplam açıklanan varyans ve her bir faktörün açıkladığı varyans hesaplanmış, faktör sayısının belirlenmesinde öz değerleri (Eigen values) birden büyük olan faktör sayısı kriteri kullanılmıştır. Faktör yüklerinin 0,4'ten büyük olması istenmiş ve birden fazla faktörün altında faktör yükleri 0,1 veya daha az farklı olan maddelerin analizden çıkarılması benimsenmiştir.

Birden fazla faktör yapısına sahip olduğu saptanan formlar aracılığıyla yapılan ölçüme ilişkin güvenilirlik katsayıları her bir faktör yapısı için ayrı ayrı hesaplanmıştır.

Elde edilen verilerin çözümlenmesinde, tanımlayıcı istatistikler, Cronbach'ın alpha katsayısı, Faktör Analizi ve ICC için IBM SPSS Versiyon 25.0 paket programı kullanılmıştır. AD değerleri, Microsoft Excel programı kullanılarak verilen formüle göre (24) elde hesaplanmış; rWG(J) değerlerinin hesaplanmasında Biemann ve Cole tarafından (22) geliştirilen "An Excel 2007 Tool for Computing Interrater Agreement (IRA) & Interrater Reliability (IRR) Estimates for Consensus Composition Constructs Version 1.5" adlı Excel programı kullanılmıştır (25).

BULGULAR

I. Birinci Sınıf Blok Değerlendirme Formları Geçerlilik ve Güvenilirlik Çalışması

Bu form için Kaiser-Meyer-Olkin (KMO) değeri 0,954 ve Bartlett küresellik testi analizinde ki-kare:3844,08, p:0,000 olarak hesaplanmıştır. Buna göre maddeler arasında faktör analizi yapmaya yeterli düzeyde bir korelasyon vardır. Faktör analizinde, formun tek faktörlü bir yapıya sahip olduğu, Eigen değerinin 8,876 olduğu ve açıkladığı toplam varyans yüzdesinin 73,9 olduğu saptanmıştır.

Maddelerin ortalama, standart sapma, faktör yükleri ve madde-toplam korelasyon katsayıları Tablo 1'de sunulmuştur. Kontrol sorusu (11. soru) çıkarılarak faktör analizi tekrarlandığında formun tek faktörlü yapısı değişmemiş, açıkladığı toplam varyans yüzdesinin 76,0 olduğu saptanmıştır.

Birinci sınıf formundaki 12 maddenin Cronbach'ın Alpha katsayısı 0,965 olarak belirlenmiş ve yapılan ölçümün iç tutarlılığına ilişkin güvenilirlik düzeyinin yüksek olduğu saptanmıştır. Ayrıca her bir madde çıkarılarak güvenilirlik katsayısı tekrar hesaplanmış, çıkarıldığında Cronbach'ın alpha katsayısını arttıran bir madde olmadığı için formun faktör yapısı hiçbir madde çıkarılmadan incelenmiştir.

Tek faktör yapısına sahip formun tümüne ait rWG(J) , AD ve ICC değerleri sırasıyla 0,91, 1,57 ve 0,875 olarak bulunmuştur.

II. İkinci ve Üçüncü Sınıf Blok Değerlendirme Formu Geçerlilik ve Güvenilirlik Çalışması

Kaiser-Meyer-Olkin (KMO) değerinin 0,926 ve Bartlett küresellik testi analizinde ki-kare:4819,92, p:0,000 olması maddeler arasında faktör analizi yapmaya yeterli düzeyde korelasyon olduğunu göstermektedir. Faktör analizine göre form iki faktörlü bir yapıya sahiptir. Eigen değerleri ilk faktör için 6.658. ikinci faktör için 1.236'dır. İlk faktör için açıklanan varyans 51.21, ikinci faktör için 9.51'dir. Toplam varyans yüzdesi 60,72'dir. Maddelerin ortalama, standart sapma, faktör yükleri ve madde-toplam korelasyon katsayıları Tablo 2'de sunulmuştur.

Formda yer alan 13 maddenin tamamı için hesaplanan Cronbach'ın alpha katsayısı 0,914; birinci faktörde yer alan on madde için 0,916; ikinci faktörde yer alan üç madde için 0,671 olarak bulunmuştur. Formdan çıkarıldığında Cronbach'ın alpha katsayısını arttıran bir madde saptanmamıştır.

Formun tamamına ait rWG(J), AD ve ICC değerleri sırasıyla 0,74, 1,66, 0,991 olarak bulunmuştur. Formun iki faktörlü yapısına göre ise aynı değerlerin sırasıyla birinci faktör için 0,75, 1,62 ve 0,98; ikinci faktör için 0,50, 2,01 ve 0,98 olduğu görülmüştür.

III. Staj Dönemi Eğitim Programı Değerlendirme Formu Geçerlilik ve Güvenilirlik Çalışması

Bu form için KMO değeri 0,979 ve Bartlett küresellik testi analizinde ki-kare:19460,65, p:0,000 olarak saptandığından maddeler arasında yeterli korelasyon olduğu düşünülmüştür. Faktör analizinde, formun tek faktörlü bir

yapıya sahip olduğu, Eigen değerinin 13.262 olduğu ve açıkladığı toplam varyans yüzdesinin 82,9 olduğu saptanmıştır. Maddelerin ortalama, standart sapma, faktör yükleri ve madde-toplam korelasyon katsayıları Tablo 3'te sunulmuştur.

Formun negatif kontrol değişkeni olan "Teorik dersler daha fazla olmalıydı" maddesinin korelasyon katsayısı 0,788 olarak saptanmıştır. Bu madde işlevini tam olarak görmese de öğrenci görüşlerinin önemli bir boyutunu yansıtması nedeniyle formdan çıkarılmamıştır. Formdaki maddeler için hesaplanan Cronbach'ın Alpha katsayısı 0,986 olarak bulunmuştur; Çıkarıldığında Cronbach'ın alpha katsayısında artış görülen bir madde yoktur.

Tek faktör yapısına sahip formun tamamı için hesaplanan rWG(J) , AD ve ICC değerleri sırasıyla 0,88, 1,63 ve 0,97'dir.

IV. Öğretim Üyesi Blok ve Staj Değerlendirme Formları Geçerlilik ve Güvenilirlik Çalışması

Öğretim üyesi blok ve staj değerlendirme formları aynı başlıkları içerdiğinden geçerlilik ve güvenilirlik analizleri form verileri birleştirilerek yapılmıştır.

KMO değeri 0,845 ve Bartlett küresellik testi analizinde ki-kare:845,020, p:0,000 olarak belirlenmiştir. Maddeler arasında faktör analizi yapmak için yeterli korelasyon vardır.

Faktör analizinde, formun üç faktörlü bir yapıya sahip olduğu belirlenmiştir. Eigen değerleri birinci faktör için 5,746, ikinci için 1,348, üçüncü faktör için 1,183'tür. Faktörler için açıklanan varyans sırasıyla % 41,044, % 9,627 ve %8,451 dir. Üç faktör birlikte varyansın % 59,121'ini açıklamaktadır. Maddelerin ortalama, standart sapma, faktör yükleri ve madde-toplam korelasyon katsayıları Tablo 4'te sunulmuştur.

Maddelerin faktör yükleri incelendiğinde, blok/staj bloğu Yürütme Kurulunun öğretim üyeleri ile iletişimine ve fakülte yönetimi ile iletişime

dair öğretim üyelerinin görüşünü soran sırasıyla 7 ve 13. maddelerin faktör 1 ve 2'deki yükleri arasında 0,1'den az fark (sırasıyla 0,019 ve 0,024) olduğu görülmüştür. Benzer şekilde öğrencilerin bilgi ve becerisi ile derslikler/ laboratuvarlara ilişkin öğretim üyesi görüşlerini araştıran 2 ve 8. maddelerin faktör 1 ve faktör 3'deki yükleri arasındaki fark sırasıyla 0,57 ve 0,061 olarak saptanmıştır. Bu maddeler çıkarıldığında maddelerin faktörlere dağılımının değişmediği ve faktör yüklerinde minimum değişiklikler olduğu görülmüştür.

Faktör yüklerinde minimum değişiklikler olması, maddelerin faktörlere dağılımının değişmemesi ve öğretim üyelerinin bu maddelerle ilgili görüşlerinin fakülte eğitim programı ve derslerin işleyişi üzerindeki etkilerin değerlendirilmesi açısından önemli olması nedeniyle maddelerin formdan çıkarılmamasına karar verilmiştir.

İlk faktör, programın yürütülmesine ilişkin temel süreç özelliklerini (yapı, organizasyon, işleyiş, program değişiklikleri, fakülte yönetimi ile iletişim, öğrenci işleri bürosu ve sınavlar); ikinci faktör insan kaynakları ile ilgili girdi özelliklerini (sınıf-laboratuvar görevlileri, diğer anabilim dalları/öğretim üyeleri, anabilim dalı sekreterleri, BYK-SBYK ile iletişim, AD sekreterleri); üçüncü faktör eğitim programının yürütülmesi üzerinde etkili olabilecek temel girdi özelliklerini (öğrenci sayısı, öğrencilerin ilgisi-becerisi, derslikler-laboratuvarlar) kapsamaktadır.

Tüm formlar için güvenilirlik (iç tutarlılık) analizleri

Formda yer alan tüm maddelerin tamamı için hesaplanan Cronbach'ın alpha katsayısı, öğretim üyesi blok değerlendirme formunda 0,850, staj değerlendirme formunda 0,815, ortak veriler üzerinden hesaplandığında ise 0,838 olarak belirlenmiştir. Yapılan faktör analizinde

üç faktörlü yapı ortaya çıkmış ve ilk faktörde yedi, ikinci faktörde dört, üçüncü faktörde üç maddenin yer aldığı görülmüştür; bu üç faktörde yer alan maddeler için hesaplanan Cronbach'ın alpha katsayısı sırasıyla 0,858, 0,770 ve 0,491 olarak bulunmuştur. Formdan çıkarıldığında Cronbach'ın alpha katsayısını arttıran bir madde yoktur.

Öğretim Üyesi Blok ve Staj Değerlendirme Formunun tamamına ait rWG(J), AD ve ICC değerleri sırasıyla 0,72, 0,57 ve 0,98 olarak hesaplanmıştır. Üç faktörlü yapıya göre hesaplandığında aynı değerlerin sırasıyla birinci faktör için 0,86, 0,52, 0,90; ikinci faktör için 0,82, 0,49, 0,94; üçüncü faktör için 0,42, 0,79, 0,98 olduğu görülmüştür.

TARTIŞMA

Program değerlendirme etkinlikleri, Ege Üniversitesi Tıp Fakültesi'nde 2001 yılında başlayan program geliştirme çalışmalarının önemli bir bileşeni olmuştur (26). Bu çalışmada, en son 2016-2017 eğitim döneminde güncellenen öğrenci ve öğretim üyesi blok ve staj değerlendirme formları kullanılarak yapılan ölçümlerin sonuçlarının geçerlilik ve güvenilirliği değerlendirilmiştir.

Bu çalışmanın, ülkemizde tıp fakültelerinde yaygın kullanılan program değerlendirme/geri bildirim formlarının kendisinin değil formlarla yapılan ölçümlerin geçerlilik ve güvenilirliğinin değerlendirilmesi ve puanlayıcılar arası güvenilirlik ve uyumun araştırılması ile ilgili ilk yayın olma ve program değerlendirme/geri bildirim formlarının çok yönlü ve şeffaf bir şekilde tartışılabilmesine örnek teşkil edebilme niteliğini taşıdığı düşünülmektedir.

Diğer yandan, çalışmamızda formların uygulamadan uygulamaya tutarlı sonuçlar verebilme ve zamana göre değişmezlik gösterebilme gücünü (kararlılık/değişmezlik)

sınamaya yarayan test – tekrar test yöntemi (27) kullanılmamıştır. Bunun yanı sıra, bu çalışmada güvenilirlik ölçümleri, ölçme hatalarını tüm grup için hesaplayabilen Klasik Test Kuramı (28) çerçevesinde yapılmıştır. Özellikle aynı bireyler (öğrenciler/öğretim üyeleri) aynı formlarla, farklı zamanlarda farklı alanları/ birimleri (bu çalışmada bloklar ve stajlar) değerlendirdiğinde, hata kaynaklarının ve etkileşiminin incelenmesinde Genellenabilirlik Kuramına dayanan hesaplamalar önemli veriler sağlasa da (29) çalışmamızda G ve/veya phi katsayıları hesaplanmamıştır.

Çalışmada 2016-2017 eğitim döneminde güncellenen öğrenci ve öğretim üyesi blok ve staj değerlendirme formları kullanılarak yapılan ölçümler için hesaplanan Cronbach'ın alpha değerlerinin kabul edilebilir sınırnın üzerinde bulunması iç tutarlılık güvenilirliğini destekleyen bir kanıttır. İkinci ve Üçüncü Sınıf Blok Değerlendirme formunda Faktör II, Öğretim Üyesi Blok ve Staj Değerlendirme Formunda ise Faktör III için hesaplanan Cronbach'ın alfa değerleri ise, yapılan ölçüme ilişkin iç tutarlılığı desteklememektedir.

Yukarıda söz edilen her iki faktör için rWG(J), ve AD değerlerinin de kabul edilebilir düzeylerin altında olması nedeniyle, bu faktörlerde temsil edilen boyutlar hakkında yapılacak yorumlarda güvenilirliğin kısıtlı olacağı unutulmaması gerekir.

Formlarda yer alan maddelerden hiçbirinin faktör yükü beklenen değer altında değildir. Ayrıca çok faktörlü öğretim üyesi ve öğrenci formlarının faktör yapısı ve maddelerin dağılımı, EÜTF program değerlendirme sisteminde benimsenen yaklaşımın (girdi, süreç ve çıktı) temel özelliklerini yansıtmaktadır. Bu faktörlerin yükleri arasındaki farkın yeterli düzeyde olması, formda yer alan maddelerin faktörlere birbirinden bağımsız dağılım

gösterdiği sonucuna işaret etmektedir.

Öğrenci formlarının madde-toplam korelasyon değerleri incelendiğinde, sadece negatif kontrol değişkeni olarak eklenen “Blok kapsamındaki teorik ders sayısı yetersizdi” maddesinin toplam puanla ilişkisinin düşük korelasyon göstermesi, negatif kontrol değişkeni için istenen bir sonuçtur ve bu anlamda işlevini görmektedir. Ancak, staj formunda negatif kontrol değişkeni olan “Teorik dersler daha fazla olmalıydı” maddesinin korelasyon katsayısının oldukça yüksek olması, öğrencilerin klinik yıllarda daha fazla ders anlatılmasını gerçekten istediklerini göstermektedir. Bu eğilimin, öğrencilerin yaklaşan TUS'a yönelik kaygılarından kaynaklandığı düşünülebilir.

Morley (17) makalesinde öğrencilerin kullandığı eğitim değerlendirmesi formlarının güvenilirliğinin incelendiği çalışmalarda genellikle hatalı güvenilirlik yöntem seçildiğine işaret etmektedir. Cronbach'ın alpha katsayısının formun iç tutarlılığını değerlendirdiğini ancak puanlayıcıların kendilerini değil dışsal bir ögeyi (örn. eğitimi /eğiticiyi) değerlendirdiği durumlarda puanlayıcılar arası uyum ve güvenilirliğin de mutlaka hesaplanması gerektiğini savunmaktadır (17).

Diğer yandan literatürde eğitimin değerlendirildiği ve puanlayıcılar arası güvenilirliğin hesaplandığı çok az sayıda çalışmada, güvenilirlik katsayılarının oldukça farklılık gösterdiği görülmektedir (17, 30, 31). Cohen ve arkadaşlarının (32) 8 haftalık cerrahi stajında eğiticilerin değerlendirilmesi amacıyla kullandıkları form için hesapladıkları sınıf içi korelasyon (ICC) orta düzeyde (0,65) bulmuştur. Türkiye’de yapılan bir çalışma (33) dahil yüksek öğrenimde ve tıp fakültelerinde öğrencilerin eğitimi değerlendirdiği çeşitli çalışmalarda iç tutarlılık ölçütleri ile puanlayıcılar arası güvenilirlik ve uyum (IRR,

IRA, IRR+IRA) ölçütlerinin ilişkisiz olduğu, IRR ve IRA'nın daha düşük hesaplandığı (0,22-0,29 arası) ve güvenilirlik konusuna tek boyutlu yaklaşmamak gerektiği bildirilmiştir (33-36).

SONUÇ

EÜTF program değerlendirme sisteminde yararlanılan ve değerlendirme sistemi yapısına uygun olarak tasarlanmış olan güncel değerlendirme formlarıyla yapılan ölçümlerin geçerlilik ve güvenilirliğine dair bu çalışmanın sonuçları, ölçümlere ilişkin geçerlilik ve güvenilirliği destekleyici kanıtlarla birlikte, ölçüm sonuçlarının yorum ve kullanımında dikkatli olunması gereken noktaları da ortaya koymuştur.

EÜTF eğitim programıyla ilgili olarak bu formların ortaya koyduğu durumun sürekli izlenmesi açısından, izleyen eğitim dönemlerinde biriken verilerle doğrulayıcı faktör analizi ve/veya yapısal eşitlik modeli ile yapı geçerliliğinin, temel bileşenler analizi ile puanlayıcılar arası güvenilirliğin doğrulanması ve genellenebilirlik (G) ve phi katsayıları ile ölçüm hata kaynaklarının ve aralarındaki etkileşimin değerlendirilmesinde yarar vardır.

Diğer yandan bu çalışma tıp fakültelerinde ve genel olarak yükseköğretimde öğrenci ve/veya öğretim üyesi değerlendirme formlarıyla eğitim programlarının değerlendirildiği çalışmalarında güvenilirlik ölçümleriyle ilgili sık yapılan hatalara dikkat çekmektedir.

İlk dikkat çekilmek istenen nokta eğitimin değerlendirilmesinde kullanılan formlarla puanlayıcının özelliklerini değil, programın özelliklerini değerlendirmenin amaçlandığının unutulmaması gerektiğidir. Bu nedenle yaygın kullanılan iç tutarlılık ölçütlerini tek başına kullanmak hatalıdır ve puanlayıcılar arası uyum ve güvenilirlik ile grup içi korelasyon değerleri hesaplanarak puanlama yapan öğrenci ve/veya

öğretim üyelerinden elde edilen sonuçların güvenilirliğinin değerlendirilmesi gerekir. Dikkat çekilmek istenen diğer bir konu da; güvenilirlik ve geçerlilik analizlerinin formların kendisini değil formla yapılan "o" ölçümü değerlendirildiğinin unutulmamasıdır. Bu anlamda tek bir ölçüm sonrası yapılan geçerlilik ve güvenilirlik analizleri o formun ilelebet geçerli ve güvenilir olduğunu göstermede yeterli değildir ve tekrarlayan ölçümlerde farklı geçerlilik ve güvenilirlik yöntemlerinin analizlere dâhil edilmesinde yarar vardır.

Kaynaklar:

1. Morrison, J. ABC of learning and teaching in medicine: Evaluation. BMJ: British Medical Journal. 2003; 326 (7385):385.
2. Kogan JR, & Shea JA. Course evaluation in medical education. Teaching and Teacher Education. 2007; 23 (3):251-64.
3. Cook DA. Twelve tips for evaluating educational programs. Medical Teacher. 2010; 32 (4):296-301.
4. Dulski L, Kelly M, Carroll VS. Program outcome data: What do we measure? What does it mean? How does it lead to improvement? Quality Management in Health Care. 2006; 15(4):296-9.
5. UTEAK. MÖTE standartları 2015. http://tepdad.org.tr/uploads/files/Belgeler%20ve%20formlar/MOTE_STANDARTLAR2015.pdf adresinden 24.05.2018 tarihinde erişilmiştir.
6. Şencan H. Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik. Ankara: Seçkin

Yayıncılık; 2006.

7. Durak HI, Vatansver K, van Dalen J, & van der Vleuten C. Factors determining students' global satisfaction with clerkships: an analysis of a two year students' ratings database. *Advances in health sciences education*. 2008; 13(4):495-502.

8. Ercan İ, & Kan İ. Ölçeklerde güvenilirlik ve geçerlik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi* 2004; 30(3):211-216.

9. Yurdugül H. Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması. XIV. Ulusal Eğitim Bilimleri Kongresi DENİZLİ, Pamukkale Üniversitesi Eğitim Fakültesi, 28-30 Eylül 2005:771-774.

10. Bushnell DS. Input, process, output: A model for evaluating training. *Training and Development Journal*. 1990; 44 (3):41-3.

11. EÜTF Program Değerlendirme Kurulu. İşaret Listesi Projesi. İzmir; Ege Üniversitesi Tıp Fakültesi; 2003.

12. EÜTF Program Değerlendirme Komisyonu. Program Değerlendirme Çalışmaları 2003-2004. İzmir; Ege Üniversitesi Tıp Fakültesi: 2004.

13. Wagner SM, Rau C, Lindemann E. Multiple informant methodology: a critical review and recommendations *Sociological Methods & Research*. 2010; 38 (4):582-618.

14. Cranton P, & Smith RA. Reconsidering the unit of analysis: A model of student ratings of instruction. *Journal of Educational Psychology*. 1990; 82:207-212.

15. Ludtke O, Robitzsch A, Trautwein U, & Kunter M. Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*. 2009; 34:120-131.

16. Marsh HW. Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness. In Perry RP & Smart JC. (Eds.), *The scholarship of teaching and learning in higher education*. Dordrecht: Springer; 2007:319-383.

17. Morley D. (2014) Assessing the reliability of student evaluations of teaching: choosing the right coefficient, *Assessment & Evaluation in Higher Education*. 2014; 39 (2):127-139.

18. Clayton DE. Student evaluation of teaching and matters of reliability, *Assessment & Evaluation in Higher Education* 2018; 43:4:666-681.

19. Nelson PM, Christ TJ. Reliability and agreement in student ratings of the class environment. *School Psychology Quarterly*. 2016; 31 (3):419-430.

20. James L, Demaree R, and Wolf G. "Estimating within-group interrater reliability with and without response bias." *Journal of Applied Psychology*. 1984; 69(1):85-98.

21. Biemann T, Cole MS, Voelpel S. Within-group agreement: On the use (and misuse) of rWG and rWG(J) in leadership research and some best practice guidelines. *Leadership Quarterly*. 2012;23:66-80.

22. Burke MJ, Dunlap WP. Estimating interrater

agreement with the average deviation index: a user's guide. *Organizational Research Methods*. 2002; 5 (2):159-172.

23. LeBreton JM, Senter JL. Answers to twenty questions about interrater reliability and interrater agreement. *Organizational Research Methods*. 2008;11:815–52.

24. Koo TK, Li MYA. Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*. 2016; 15(2):155–63.

25. Biemann T, Cole MS. An Excel 2007 Tool for Computing Interrater Agreement (IRA) & Interrater Reliability (IRR) Estimates Version 1.5. 2014. 05.07.2018 tarihinde http://www.sbuweb.tcu.edu/mscole/docs/Tool%20for%20Computing%20IRA%20and%20IRR%20Estimates_v1.5.zip adresinden erişildi.

26. EÜTF Plan Grubu. Klinik öncesi dönem eğitimi programını yatay ve dikey entegrasyon zemininde yeniden yapılandırma çalışmaları rehberi ve 2002-2003 yılı çalışma takvimi İzmir; Ege Üniversitesi Tıp Fakültesi: 2002.

27. Aker S, Dündar C, Pekşen Y. Ölçme Araçlarında iki Yaşamsal Kavram: Geçerlik ve Güvenirlik: Derleme. *Deneyel ve Klinik Tıp Dergisi*. 2005; 22(1):50-60.

28. Çelen Ü. Klasik test kuramı ve madde tepki kuramı yöntemleriyle geliştirilen iki testin geçerlilik ve güvenilirliğinin karşılaştırılması. *İlköğretim Online* 2008; 7(3):758-68.

29. Güler N. Rasgele veriler üzerinde Genellenebilirlik Kuramı ve Klasik Test Kuramı'na göre güvenilirliğin karşılaştırılması.

Eğitim ve Bilim. 2011; 36(162):225-34.

30. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, & Mandrekar JN. How reliable are assessments of clinical teaching?. *Journal of General Internal Medicine*. 2004; 19(9): 971-7.

31. Kogan JR, and Shea JA. Course evaluation in medical education. *Teaching and Teacher Education*. 2007; 23(3):251-64.

32. Cohen R, MacRae H & Jamieson C. Teaching effectiveness of surgeons. *The American Journal of Surgery*. 1996; 171(6):612-4.

33. Kalender İ. Reliability-Related Issues in the Context of Student Evaluations of Teaching in Higher Education. *International Journal of Higher Education*. 2015; 4(3):44-56.

34. Solomon DJ, Speer AJ, Rosebraugh CJ, & DiPette DJ. The reliability of medical student ratings of clinical teaching. *Evaluation & the health professions*. 1997; 2(3):343-52.

35. Cook DA & Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine* 2006; 119(2): 166-e7-e16.

36. Rantanen P. The number of feedbacks needed for reliable evaluation: A multilevel analysis of the reliability, stability and generalisability of students' evaluation of teaching. *Assessment & Evaluation in Higher Education*. 2013; 38(2): 224-39.

Tablo 1. Birinci Sınıf Blok Değerlendirme Formu maddeleri ve maddelerle ilgili analiz sonuçları

| <i>Maddeler</i> | <i>Ort.</i> | <i>SS</i> | <i>Faktör yükü</i> | <i>Madde-toplam korelasyon katsayısı</i> |
|---|-------------|-----------|--------------------|--|
| 1. Blok başında, bloğun amacı, içeriği, uygulanması, geçme - kalma kriterleri ve sınavlar anlaşılır bir şekilde açıklandı | 6,16 | 2,62 | 0,898 | 0,868 |
| 2. Teorik dersler anlayarak öğrenmemi sağlayacak biçimde anlatıldı | 5,83 | 2,52 | 0,891 | 0,869 |
| 3. Uygulamalı eğitimler anlayarak öğrenmemi sağlayacak şekilde gerçekleştirildi | 6,35 | 2,45 | 0,907 | 0,878 |
| 4. Bloкта teorik dersler ve uygulamalar tematik bütünlük içinde birbirini tamamladı | 6,25 | 2,48 | 0,920 | 0,893 |
| 5. Blok kapsamındaki teorik dersler ve uygulamalı eğitimlerle ilgili değişiklikler için zamanında haber verildi | 6,30 | 2,62 | 0,878 | 0,846 |
| 6. Blok Ara Sınavları zayıf ve güçlü yönlerimi değerlendirme, bildiklerimi pekiştirme fırsatı verdi | 5,62 | 2,51 | 0,881 | 0,854 |
| 7. Uygulamalı eğitim sınavları, blok amaç ve hedefleriyle uyumlu olarak planlanmıştı | 6,11 | 2,55 | 0,915 | 0,890 |
| 8. Blok Yazılı Sınavı (BYS) blok amaç ve hedefleriyle uyumlu olarak planlanmıştı | 5,91 | 2,51 | 0,903 | 0,877 |
| 9. Bu bloкта ilk üç yıl için hazırlanan blok kılavuzundan etkin biçimde yararlandım | 5,35 | 2,69 | 0,817 | 0,786 |
| 10. Blokla ilgili önerilen kaynaklardan etkin olarak yararlandım | 5,34 | 2,54 | 0,862 | 0,842 |
| 11. Blok kapsamındaki teorik ders sayısı yetersizdi | 3,74 | 2,49 | 0,424 | 0,386 |
| 12. Dersler ve uygulamalar, öğrenme ve öğrendiklerimi kullanma konusundaki motivasyonumu olumlu etkiledi | 5,77 | 2,51 | 0,902 | 0,877 |

*Faktör yapısı matrisi kullanılmıştır

Tablo 2. İkinci ve Üçüncü Sınıf Blok Değerlendirme Formu maddeleri ve maddelerle ilgili analiz sonuçları

| Maddeler | Ort. | SS | Faktör yükleri* | | Madde-toplam korelasyon katsayısı |
|--|------|------|-----------------|----------|-----------------------------------|
| | | | Faktör 1 | Faktör 2 | |
| 4. Bloкта teorik dersler ve uygulamalar tematik bütünlük içinde birbirini tamamladı | 5,89 | 2,19 | 0,848 | | 0,734 |
| 2. Teorik dersler anlayarak öğrenmemi sağlayacak biçimde anlatıldı | 5,77 | 2,12 | 0,841 | | 0,758 |
| 3. Uygulamalı eğitimler anlayarak öğrenmemi sağlayacak şekilde gerçekleştirildi | 5,92 | 2,24 | 0,810 | | 0,690 |
| 6. Blok Ara Sınavları zayıf ve güçlü yönlerimi değerlendirme, bildiklerimi pekiştirme fırsatı verdi | 5,66 | 2,31 | 0,770 | | 0,712 |
| 1. Blok başında, bloğun amacı, içeriği, Uygulanması, geçme-kalma kriterleri ve sınavlar anlaşılır bir şekilde açıklandı | 6,41 | 2,43 | 0,712 | | 0,598 |
| 7. Uygulamalı eğitim sınavları, blok amaç ve hedefleriyle uyumlu olarak planlanmıştı | 6,02 | 2,25 | 0,795 | | 0,700 |
| 12. Blok kapsamındaki dersler ve uygulamalar, öğrenme ve öğrendiklerimi kullanma konusundaki motivasyonumu olumlu etkiledi | 5,35 | 2,21 | 0,788 | | 0,752 |
| 10. Blokla ilgili önerilen kaynaklardan etkin olarak yararlandım | 4,88 | 2,40 | 0,723 | | 0,735 |
| 5. Blok kapsamındaki teorik dersler ve uygulamalı eğitimlerle ilgili değişiklikler için zamanında haber verildi | 5,37 | 2,55 | 0,704 | | 0,643 |
| 8. Blok Yazılı Sınavı (BYS) blok amaç ve hedefleriyle uyumlu olarak planlanmıştı | 5,39 | 2,56 | 0,559 | | 0,542 |
| 9. Bu bloкта ilk üç yıl için hazırlanan blok kılavuzundan etkin biçimde yararlandım | 4,51 | 2,59 | | 0,751 | 0,683 |
| 11. Blok kapsamındaki teorik ders sayısı yetersizdi | 3,09 | 2,32 | | 0,745 | 0,292 |
| 13. Mesleksel yabancı dil eğitimi yeterliydi | 4,00 | 2,63 | | 0,737 | 0,540 |

Tablo 3. Staj Blokları Geribildirim Formu maddeleri ve maddelerle ilgili analiz sonuçları

| <i>Maddeler</i> | <i>Ort.</i> | <i>SS</i> | <i>Faktör yükü</i> | <i>Madde-toplam korelasyon katsayısı</i> |
|--|-------------|-----------|--------------------|--|
| 1. Staj bloğunun başında bloğun amacı, içeriği, uygulanması, geçme - kalma ölçütleri ve sınavlar anlaşılır bir şekilde açıklandı | 5,28 | 2,42 | 0,873 | 0,855 |
| 2. Teorik dersler öğrenmemi kolaylaştıracak şekilde anlatıldı | 5,03 | 2,30 | 0,932 | 0,921 |
| 3. Staj konuları klinikte yeterli ve anlaşılır şekilde aktarıldı | 4,87 | 2,33 | 0,936 | 0,926 |
| 4. Uygulamalı eğitimler öğrenmemizi kolaylaştıracak şekilde yürütüldü | 4,88 | 2,31 | 0,938 | 0,929 |
| 5. Entegre oturumlar öğrenmemi kolaylaştıracak şekilde aktarıldı | 4,92 | 2,29 | 0,934 | 0,923 |
| 6. Temel bilim konuları. daha önce öğrendiklerimi hatırlamam ve klinik bilimler konularıyla ilişkilendirmem açısından faydalıydı | 4,80 | 2,33 | 0,906 | 0,892 |
| 7. Staj boyunca teorik dersler ve uygulamalar birbirini tamamladı | 4,77 | 2,31 | 0,933 | 0,922 |
| 8. İlan edilen teorik ders ve uygulamalı eğitim programına uyuldu | 4,65 | 2,38 | 0,906 | 0,892 |
| 9. Klinik rotasyonlara ayrılan süreleri genel olarak yeterli buldum | 4,90 | 2,36 | 0,901 | 0,886 |
| 10. Hastalarla uygulama yapma olanakları yeterliydi | 4,81 | 2,36 | 0,910 | 0,897 |
| 11. Staj bloğu sırasında ne zaman nerede olmam gerektiğini daima biliyordum | 4,81 | 2,39 | 0,915 | 0,902 |
| 12. Teorik sınavlar staj bloğunun öğrenme amaç ve hedeflerini değerlendirecek şekilde yapıldı | 4,88 | 2,32 | 0,934 | 0,924 |
| 13. Uygulamalı/sözlü sınavlar staj bloğunun öğrenme amaç ve hedeflerini değerlendirecek şekilde yapıldı | 4,96 | 2,33 | 0,926 | 0,915 |
| 14. Teorik dersler daha fazla olmalıydı | 4,42 | 2,38 | 0,811 | 0,788 |
| 15. Blokla ilgili önerilen kaynaklardan etkin olarak yararlandım | 4,81 | 2,31 | 0,911 | 0,899 |
| 16. Staj bloğunda öğrendiklerimi ileride meslek yaşamımda kullanacağımı düşünüyorum | 5,31 | 2,39 | 0,892 | 0,877 |

Tablo 4. Öğretim üyesi blok-staj değerlendirme formu maddeleri ve maddelerle ilgili analiz sonuçları

*Faktör yapısı matrisi kullanılmıştır

| <i>Maddeler</i> | <i>Ort.</i> | <i>SS</i> | <i>Faktör yükleri*</i> | | | <i>Madde-Toplam Korelasyon Katsayısı</i> |
|---|-------------|-----------|------------------------|-----------------|-----------------|--|
| | | | <i>Faktör 1</i> | <i>Faktör 2</i> | <i>Faktör 3</i> | |
| 3.Programın yapısı / organizasyonu | 3,77 | ,872 | ,872 | | | 0,685 |
| 4. Program işleyişi | 3,82 | ,809 | ,819 | | | 0,602 |
| 14.Öğrenci geri bildirimleri | 3,66 | ,969 | ,761 | | | 0,541 |
| 5.Program değişiklikleri hakkında bilgilendirme | 3,85 | ,909 | ,722 | | | 0,499 |
| 13.Fakülte yönetimi iletişim | 4,11 | ,825 | ,686 | | | 0,559 |
| 9. Öğrenci işleri bürosu | 4,00 | ,755 | ,643 | | | 0,462 |
| 6.Sınav sistemi ve işleyişi | 3,80 | ,846 | ,626 | | | 0,470 |
| 10. Sınıf/Laboratuvar görevliler | 4,04 | ,803 | | ,786 | | 0,531 |
| 12.Diğer anabilim dalları / öğretim üyeleri | 3,88 | ,732 | | ,778 | | 0,494 |
| 11.Anabilim dalı sekreterleri | 4,20 | ,752 | | ,752 | | 0,408 |
| 7.Blok/Staj Bloğu yürütme kurulunun öğretim üyeleri ile iletişimi | 3,78 | ,939 | | ,641 | | 0,530 |
| 1. Öğrenci sayısı | 2,36 | 1,161 | | | ,890 | 0,253 |
| 2. Öğrencilerin ilgisi-becerisi | 3,48 | ,928 | | | ,583 | 0,416 |
| 8.Derslikler /Laboratuvarlar | 3,48 | 1,036 | | | ,489 | 0,332 |