

## Tıp Eğitiminde Genellenebilirlik Kuramı Uygulaması Generalizability Theory Application in Medical Education

Giray Kolcu<sup>1,2</sup>, Mukadder İnci Başer Kolcu<sup>1</sup>

<sup>1</sup> Dr.Öğr.Üyesi, Süleyman Demirel Üniversitesi Tıp Eğitimi ve Bilişimi Ana Bilim Dalı, Isparta, Türkiye

<sup>2</sup> Dr.Öğr.Üyesi, Süleyman Demirel Üniversitesi Sağlık Bilimleri Enstitüsü, Isparta, Türkiye

### Özet:

Tıp eğitimi programlarında öğrenmenin değerlendirilebilmesi için güvenilir ölçüm araçlarının geliştirilmesi önerilmektedir. Genellenebilirlik kuramı (G-kuramı) bilgi, beceri ve tutum ölçülmesinde güvenilirliğin değerlendirilmesini, güvenilir gözlemlerin araştırılmasını ve kavramsallaştırılmasını sağlayan istatistiksel bir kuramdır. Bu çalışmada Süleyman Demirel Üniversitesi Tıp Fakültesi mezuniyet öncesi tıp eğitimi programında (SDÜ-MÖTEP) uygulanmış olan bir adet çoktan seçmeli yazılı ölçüm aracının güvenilirlik analizinin genellenebilirlik kuramı ile değerlendirilmesi amaçlanmıştır. Çalışmada SDÜTF-MÖTEP içerisinde 2018-2019 eğitim öğretim yılında uygulanmış olan dönem.1 kurul.1 sınavı genellenebilirlik kuramı bağlamında bir yüzeyle çaprazlanmış desen olarak değerlendirilmiştir. Tek sınavı irdeleyen bu makale, tıp eğitiminde olgu sunumu olarak değerlendirilmektedir. Çalışma, 76 öğrenciye uygulanmış 82 çoktan seçmeli sorudan oluşan ölçüm aracı ile gerçekleştirilmiştir. Ölçüm aracının ortalama madde zorluğu 0,66 , ortalama ayırıcılık 0,24 , olarak hesaplandı. Ölçüm aracının G değeri 0,85 olarak bulundu. Bu deneyim doğrultusunda tıp eğitiminde ölçme/değerlendirme verilerinin analizinde genellenebilirlik kuramının tercih edilebileceği düşünülmektedir.

**Anahtar Kelimeler:** Değerlendirme, genellenebilirlik, güvenilirlik, kuram, ölçme, tıp eğitimi

### Summary:

It is recommended to develop reliable measurement tools to evaluate learning in medical education programs. Generalizability theory (G-Theory) is a statistical theory that enables the evaluation of reliability in measuring knowledge, skills and attitudes, research and conceptualization of reliable observations. In this study, it was aimed to evaluate the reliability analysis of a multiple-choice written measurement tool applied in Süleyman Demirel University Faculty of Medicine undergraduate medical education program (SDUFM-UMEP) with generalizability theory. In the study, the first semester first exam applied in the (SDUFM-UMEP) in the 2018-2019 academic year was evaluated as a faceted crossed pattern in the context of generalizability theory. This article, which examines a single exam, is considered as a case report in medical education. A measurement tool consisting of 82 multiple-choice questions applied to 76 students was evaluated. The mean item difficulty of the measuring instrument was calculated as 0.66 and the average discrimination as 0.24. The G value of the measuring tool was calculated as 0.85. In line with this experience, it was believed that generalizability theory can be preferred in the analysis of measurement / evaluation data in medical education.

**Key Words:** Evaluation, generalizability, reliability, theory, measurement, medical education.

*Kabul Tarihi: 01.10.2019*

### Giriş

Program değerlendirme bir programın sürdürülmesi, geliştirilmesi veya sonlandırılması ile ilgili karar verme süreci için bilimsel yöntemler ile bilgi üreten bir değerlendirme sürecidir. Öğrenmenin değerlendirilmesi, eğitim programlarının değerlendirilmesinin önemli bölümlerinden biridir (1,2,3). Öğrenmenin değerlendirilmesinde ölçme değerlendirmenin monitorizasyonu eğitim programı ile ilgili birçok veri sunar (4,5,6). Ölçme değerlendirmenin monitorizasyonu için ölçüm araçlarının bir

kuram bağlamında düzenli ve sürekli istatistiksel analizlerinin yapılması gerekmektedir (7,8,9). Bu verilerin değerlendirilebilmesi için öğrencilerin değerlendirilmesinde güvenilir geçerli ölçümlerin yapılması önemli kriterler arasındadır (10,11,12,13,14).

Genellenebilirlik kuramı (G-kuramı) bilgi, beceri ve tutum ölçülmesinde güvenilirliğin değerlendirilmesini, güvenilir gözlemlerin araştırılmasını ve kavramsallaştırılmasını sağlayan istatistiksel bir kuramdır (15). Varyans analizine dayalı G-kuramı bir ölçüm aracının

farklı hata kaynaklarının değerlendirilmesine tek bir analiz ile imkân veren bir istatistiksel kuramdır (15,16,17,18).

Bu çalışmada Süleyman Demirel Üniversitesi Tıp Fakültesi mezuniyet öncesi tıp eğitimi programında (SDÜ-MÖTEP) uygulanmış olan bir adet çoktan seçmeli yazılı ölçüm aracının güvenilirlik analizlerinin genellenebilirlik kuramı ile değerlendirilmesi amaçlanmıştır.

Çalışmada SDÜTF-MÖTEP içinde 2018-2019 eğitim öğretim yılında uygulanmış olan dönem.1 kurul.1 sınavı değerlendirilmiştir. Tek sınavı irdeleyen bu makale, tıp eğitiminde olgu sunumu olarak değerlendirilmektedir. Çalışma için SDÜTF etik kurulundan onay ve fakülte yönetiminden yasal izin alınmıştır. Çoktan seçmeli yazılı sınav aynı öğrenci grubuna, bir kez, aynı form ile uygulandığı ve maddeler dışında başka potansiyel hata kaynağı olmaması nedeniyle “bir yüzeyle çaprazlanmış desen” olarak değerlendirilmiş; çalışma verileri MS-Excel, TAP, EduG ve SPSS ile analiz edilmiştir (19). Mevcut çalışma sırasında kullanılan ve / veya analiz edilen veri kümeleri, makul talep üzerine ilgili yazardan temin edilebilir.

Çalışmada, 82 çoktan seçmeli sorudan oluşan ölçüm aracı 76 öğrenciye uygulandı. (Tablo.1). Ortalama madde zorluğu 0,66, ortalama ayırıcılık 0,24, ortalama nokta çift seri korelasyon katsayısı 0,22, KR-20 0,85 olarak hesaplandı. 19 madde potansiyel problemler olarak değerlendirildi. Ölçüm aracının tanımlayıcı analizleri aşağıdaki tabloda belirtilmiştir (Tablo.1)

Ölçüm aracı puanlarının G-kuramı ile analizinde bir yüzeyle çaprazlanmış desende bireyler için kestirilen varyans bileşeni yüzdesinin bağıl değeri 5%, maddeler için kestirilen varyans bileşeni yüzdesi 18,1% ve birey-madde için kestirilen varyans bileşeni yüzdesi 76% olarak hesaplanmıştır (Tablo.2).

Uygulanan 82 maddelik uygulama için G değeri 0,85 olarak hesaplanmıştır. Ölçüm aracı madde sayısı ile yapılan karar (K) çalışmasında hesaplanan G değerleri tabloda gösterilmiştir (Tablo.3).

Çalışmada dönem.1 kurul.1 sınavı örnek olarak tercih edilmiştir. Ölçüm aracının madde analizlerinde ortalama madde zorluğu 0,785,

ortalama ayırıcılık indeksi 0,262 olarak hesaplanmıştır. Bu analizler doğrultusunda 14 maddenin gözden geçirilmesi önerilmiştir. Ayrıca ölçüm aracı puanlarının G-kuramı ile analizinde bir yüzeyle çaprazlanmış desende bireyler için kestirilen varyans bileşeni yüzdesinin bağıl değerinin düşük olması ayırıcılığın yetersiz olduğunu, maddeler için kestirilen varyans bileşeni yüzdesi güçlük düzeylerindeki dengesiz dağılım nedeni ile genellemenin zayıfladığını ifade etmektedir, birey-madde için kestirilen varyans bileşeninin büyüklüğü sistematik veya sistematik olmayan hata kaynaklarının kontrol edilemediğini ifade etmektedir. Ölçüm aracı madde sayısı ile yapılan karar (K) çalışmasında kapsam geçerliliği göz önüne alınarak soru sayısında değişiklikler yapılabileceği önerilmektedir. Ölçüm puanlarının güvenilirlik analizlerinde uygulama genellenebilirlik kuramına göre 0,90 olarak hesaplanmış ve güvenilir olarak değerlendirilmiştir. Bu analizlerle birlikte ölçüm araçlarının madde düzeyinde kalitesinin geliştirilmesi için bir geribildirim elde edilmiştir.

Çalışmada tek bir ölçüm aracının değerlendirilmesi bir kısıtlılık olarak değerlendirilmeye birlikte analizlerin gözden geçirilmesi ve durumun değerlendirilmesi yönünden oldukça değerli bilgiler üretilmesini sağlamıştır. Bu deneyim doğrultusunda tıp eğitiminde ölçme/değerlendirme verilerinin analizinde genellenebilirlik kuramının tercih edilebileceği düşünülmektedir.

Tablo 1. Ölçüm aracının tanımlayıcı analizi

Madde	Ayırıcılık İndeksi	Zorluk İndeksi	Nokta Çif Serili Korelasyon Katsayısı	Madde	Ayırıcılık İndeksi	Zorluk İndeksi	Nokta Çif Serili Korelasyon Katsayısı
Madde 01	0.63	0.4	0.37	Madde 42	0.39	0.25	0.22
Madde 02	0.95	0.09	0.21	Madde 43	0.62	0.31	0.19
Madde 03	0.93	0.23	0.43	Madde 44	0.8	0.36	0.47
Madde 04	0.95	0.14	0.21	Madde 45	0.49	0.31	0.32
Madde 05	0.83	0.36	0.39	Madde 46	0.26	0.02	0.14
Madde 06	0.88	0.04	0.08	Madde 47	0.72	0.54	0.48
Madde 07	0.92	0.18	0.35	Madde 48	0.82	0.27	0.34
Madde 08	0.83	0.18	0.25	Madde 49	0.82	0.36	0.41
Madde 09	0.84	0.18	0.11	Madde 50	0.54	0.54	0.49
Madde 10	0.43	0.3	0.18	Madde 51	0.71	0.54	0.43
Madde 11	0.66	0.22	0.2	Madde 52	0.79	0.31	0.28
Madde 12	0.64	0.59	0.51	Madde 53	0.76	0.5	0.47
Madde 13	0.88	0.05	0.05	Madde 54	0.7	0.4	0.31
Madde 14	0.82	0.41	0.38	Madde 55	0.66	0.5	0.38
Madde 15	0.63	0.36	0.26	Madde 56	0.76	0.45	0.45
Madde 16	0.72	0.31	0.34	Madde 57	0.61	0.4	0.35
Madde 17	0.25	0.39	0.35	Madde 58	0.67	0.4	0.35
Madde 18	0.62	0.54	0.46	Madde 59	0.79	0.41	0.47
Madde 19	0.63	0.26	0.24	Madde 60	0.45	0.44	0.34
Madde 20	0.33	0.43	0.37	Madde 61	0.74	0.27	0.26
Madde 21	0.62	0.5	0.36	Madde 62	0.97	0.09	0.19
Madde 22	0.29	0.11	0.11	Madde 63	1	0	****
Madde 23	0.03	0	0.05	Madde 64	0.86	0.13	0.17
Madde 24	0.09	0.04	0.07	Madde 65	1	0	****
Madde 25	0.59	0.12	0.03	Madde 66	0.93	0.09	0.25
Madde 26	0.21	0.15	0.1	Madde 67	0.93	0.04	0.17
Madde 27	0.51	0.21	0.18	Madde 68	0.5	0.26	0.18
Madde 28	0.5	0.45	0.34	Madde 69	0.79	0.06	0.05
Madde 29	0.71	0.54	0.45	Madde 70	0.41	0.44	0.36
Madde 30	0.2	0.36	0.38	Madde 71	1	0	****
Madde 31	0.11	0.1	0.14	Madde 72	0.91	0.23	0.33
Madde 32	0.76	0.36	0.44	Madde 73	0.79	0.5	0.54
Madde 33	0.97	0.09	0.35	Madde 74	0.99	0	0.04
Madde 34	0.49	0.26	0.24	Madde 75	1	0	****
Madde 35	0.29	0.62	0.5	Madde 76	0.75	0.5	0.48
Madde 36	0.91	0.27	0.41	Madde 77	0.95	0.14	0.34
Madde 37	0.74	0.4	0.38	Madde 78	0.99	0.05	0.31
Madde 38	0.09	0.15	0.22	Madde 79	0.89	0.18	0.36
Madde 39	0.32	0.12	0.06	Madde 80	1	0	****
Madde 40	0.38	0.25	0.15	Madde 81	0.96	0.05	0.04
Madde 41	0.33	0.2	0.18	Madde 82	0.64	0.22	0.2

**Tablo 2.** Ölçüm aracının genellenebilirlik kuramı ile varyans analizi

Varyans kaynağı*	Kareler Toplamı	Serbestlik derecesi	Kareler ortalaması	Varyans	Varyans yüzdesi
Birey	319.286	278.000	1.149	0.012	0.05
Madde	944.127	81.000	11.656	0.041	0.18
BireyXMadde	0.175	22518.000	0.175	0.175	0.77

**Tablo 3.** Ölçüm aracı için Karar çalışması analizi

	60 Madde	70 Madde	80 Madde	82 Madde	90 Madde
G-coefficients	0.80	0.82	0.84	0.85	0.86

## Kaynaklar

1. TEPDAD. Tıp Eğitimi Programlarını Değerlendirme Ve Akreditasyon Derneği Mezuniyet Öncesi Tıp Eğitimi Programı Öz Değerlendirme Raporu Hazırlama Kılavuzu Sürüm 4. 2018.
2. Tepdad. Mezuniyet Öncesi Tıp Eğitimi Ulusal Standartları-2020. 2020.
3. Norcini JJ, McKinley DW. Assessment methods in medical education. Teach Teach Educ 2007;23(3):239-50.
4. Blouin D, Tekian A, Kamin C, Harris IB. The impact of accreditation on medical schools' processes. Med Educ 2018;52(2):182-91.
5. Stufflebeam D. Evaluation models. New Dir Eval 2001;2001(89):7-98.
6. Behçet P, Yrd O, Taha D. Eğitimde program geliştirme ve değerlendirme. Eğitimde program geliştirme ve değerlendirme. 2017.
7. Özcan KV, Aydoğan YİB. Gaziosmanpaşa Üniversitesi Tıp Fakültesi'nde Uygulanan Çoktan Seçmeli Sınavların Betimsel Analizi. Gaziosmanpaşa Üniversitesi Tıp Fakültesi Derg. 2015;6(4):281-94.
8. Topal K, Aybek H, Kara CO, Akile BZA. Item and test analysis of multiple-choice exams which applied to the first term students of Pamukkale University Medical Faculty in 2006-2007 academic year. Pamukkale Tıp Derg 2008;1(3):120-6.
9. Onan A, Abay E, Odabaşı O. Öğrenci geribildirimlerinin mesleksi beceri eğitimi programı değerlendirmesinde kullanılması. Kocaeli Tıp Derg 2016;5(1):25-32.
10. Cho E, Kim S. Cronbach's coefficient alpha: Well known but poorly understood. Organ Res Methods 2015;18(2):207-30.
11. Taşdelen TG, Odabaşı O. Reliability of scores obtained from standardized patient and instructor assessments. Eur J Dent Educ 2019;23(2):88-94.
12. Güler N, Eroğlu Y, Akbaba S. Reliability of

critereye bağımlı ölçüm araçları genellenebilirlik kuramına göre değerlendirilmiştir. Uygulama: Yemeğin Yemeğin İçerdiği Besinlerin Kaynakları Üzerine Bir Çalışma. Tıp Dergisi 2014;14:217-32.

13. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.
14. Kassab SE, Du X, Toft E, Cyprian F, Al-Moslih A, Schmidt H, et al. Measuring medical students' professional competencies in a problem-based curriculum: a reliability study. BMC Med Educ 2019;19(1):155.
15. Atılğan. Genellenebilirlik Kuramı ve Uygulaması. 1. Baskı. Ankara; 2019;1.
16. Güler. Genellenebilirlik kuramı ve SPSS ile GENOVA programlarıyla hesaplanan G ve K çalışmalarına ilişkin sonuçların karşılaştırılması. Eğitim ve Bilim. 2009;34(154):93-104.
17. Atılğan H. Genellenebilirlik kuramı ve puanlayıcılar arası güvenilirlik için örnek bir uygulama. Eğit ve Bili. 2005;(7):95-108.
18. Baykul Y. Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması. Ankara: ÖSYM Yayınları; 2000.
19. Mushquash C, O'Connor BP. SPSS, SAS, and MATLAB programs for generalizability theory analyses. Behav Res Methods 2006;38(3):542-7.

## İletişim:

Dr. Öğr. Üyesi Giray Kolcu  
SDÜ Tıp Eğitimi ve Bilişimi Anabilim Dalı,  
Isparta, Türkiye  
Tel: +90.505.5306179  
E-posta: giraykolcu@gmail.com