



Evaluation of the reliability of the online multiple choice assessment tool by the generalizability theory

Çevrimiçi çoktan seçmeli değerlendirme aracının güvenilirliğinin genellenebilirlik kuramı ile değerlendirilmesi

Giray Kolcu¹, Sadettin Demir², Mukadder İnci Başer Kolcu³

¹⁾ Süleyman Demirel University, Department of Medical Education and Informatics, Assist. Prof. Dr., Isparta, Türkiye

²⁾ Süleyman Demirel University, Institute of Health Sciences, Assist. Prof. Dr., Isparta, Türkiye

³⁾ Süleyman Demirel University, Department of Medical Education and Informatics, Dr., Isparta, Türkiye

İletişim adresi:

Giray Kolcu

E-mail: giraykolcu@gmail.com

Geliş tarihi: 27.11.2020

Kabul tarihi: 31.01.2021

Yayın tarihi: 25.03.2021

Alıntı Kodu: Kolcu G. ve ark. Evaluation of the reliability of the online multiple choice assessment tool by the generalizability theory. Jour Turk Fam Phy 2021; 12 (1): 3-11. Doi: 10.15511/tjtfp.21.00003.

Özet

Giriş: Ülkemizdeki COVID-19 pandemisi nedeniyle Yükseköğretim Kurulu tarafından sadece pandemi dönemi için geçerli olmak üzere yükseköğretimde ölçme ve değerlendirmenin çevrimiçi olarak uygulanmasına karar verilmiştir.

Amaç: Çalışmamızda Süleyman Demirel Üniversitesi'nde 1. yıl 3. kurul öğrencileri için çevrimiçi olarak uygulanan çoktan seçmeli değerlendirme aracının temel analizlerinin klasik test ve genellebilirlik kuramı ile değerlendirilmesi amaçlanmaktadır.

Yöntem: Çalışmamız nicel araştırma deseninde tasarlanmıştır. Çalışma evreni Süleyman Demirel Üniversitesi Tıp Fakültesi'nde aktif olarak öğrenim gören 1. sınıf öğrencileri olarak belirlendi (n: 271). Çalışmada 1. yıl öğrencilerine çevrimiçi olarak uygulanan 3. kurul çoktan seçmeli değerlendirme aracı SPSS ve EduG ile analiz edilmiştir.

Bulgular: Sınav 100 puan üzerinden değerlendirildiğinde ortalama 78.5 ± 11.05 (min: 27.4; max: 98.0), varyans 122.229, basıklık -1.196 ve çarpıklık 1.683 olarak hesaplandı. Ortalama madde gücü 0.785, ortalama ayırt etme indeksi 0.262, güvenilirlik katsayısı (KR-20) ise 0.902 olarak hesaplandı. Doksan beş madde için G değeri 0.91, Phi değeri 0.90 olarak hesaplandı.

Tartışma: Fakültemiz YÖK ve Tıp Eğitimi Programlarını Değerlendirme ve Akreditasyon Derneği'nin önerileri doğrultusunda uzaktan eğitimde ölçme-değerlendirme uygulamalarının izlenmelerini de sağlamıştır. Tek bir değerlendirme aracının değerlendirilmesi çalışmamız için bir kısıtlılık olarak değerlendirilmekle birlikte çalışmamız çoktan seçmeli ölçme araçlarının temel analizlerinin gözden geçirilmesi ve durum değerlendirmesi ile ilgili değerli bilgiler sağlamıştır. Bu deneyime dayanarak çevrimiçi ölçme ve değerlendirme uygulamalarının temel analizlerinin pandemi sonrasında da ölçme araçlarının analizi için tercih edilebileceği kanaatindeyiz.

Anahtar Kelimeler: Tıp eğitimi, Ölçme, Güvenilirlik, Genellebilirlik Kuramı

Summary

Introduction: Due to the COVID-19 pandemic in our country, it has been decided by the Higher Education Council (CoHE) to apply online measurement and evaluation in higher education, valid only for the pandemic period.

Aim: In our study, it is aimed to evaluate the basic analyzes of the multiple-choice assessment tool, which is applied online for 1st year and 3rd committee students at Süleyman Demirel University, with classical test and generalizability theory.

Method: Our study is designed in quantitative research design. The population of the study was determined as first year students actively studying at Süleyman Demirel University Faculty of Medicine (n: 271). In the study, the 3rd committee applied online to the 1st year students was analyzed with the multiple choice assessment tool SPSS and EduG.

Results: When the exam was evaluated over 100 points, the average was 78.5 ± 11.05 (min: 27.4; max: 98.0), variance 122.229, kurtosis -1.196 and skewness 1.683. The average item difficulty was 0.785, the mean discrimination index was 0.262, and the reliability coefficient (KR-20) was 0.902. For the 95-item exam, the G value was calculated as 0.91 and the Phi value as 0.90.

Discussion: Our faculty has also ensured the monitoring of measurement-evaluation practices in distance education in line with the recommendations of CoHE and the Association for Evaluation and Accreditation of Medical Education Programs. Although the evaluation of a single assessment tool was considered as a limitation for our study, our study provided valuable information about the review of the basic analysis and assessment of the multiple-choice measurement tools. Based on this experience, we believe that the basic analysis of online measurement and evaluation applications can be preferred for the analysis of measurement tools after the pandemic.

Keywords: Medical education, Assessment, Reliability, Generalizability Theory

Alıntı Kodu: Kolcu G. ve ark. Evaluation of the reliability of the online multiple choice assessment tool by the generalizability theory. *Jour Turk Fam Phy* 2021; 12 (1): 3-11. Doi: 10.15511/tjfp.21.00003.

Introduction

Due to the COVID-19 pandemic in our country, it was decided by the Council of Higher Education (CoHE) that assessment and evaluation would be completed online in higher education, to be valid only for the pandemic period.⁽¹⁾ Main principles such as transparency, fairness, and controllability were identified for assessment and evaluation practices in higher education during the pandemic.⁽²⁾

In line with these principles, the Association for Development of Medical Education (TEPDAD) published a document titled “Suggestions of Association for Evaluation and Accreditation of Medical Education Programs on Assessment and Evaluation in Medical Education during the COVID-19 Pandemic” on May 31, 2020.⁽³⁾ Also, in the National Standards for Pre-Graduate Medical Education-2020, it is recommended that the medical faculty continuously improve its system with assessment and evaluation practices as well as in development standards and evaluate the effectiveness of practices.⁽⁴⁾

These recommendations have provided in-depth explanations on the assessment and evaluation process.⁽⁵⁾ An “Assessment / Evaluation monitorization” in which the validity, reliability, and practicality of assessment tools are evaluated, is recommended for this evaluation.^(4,7) One of the essential parameters of this monitoring is reliability.⁽⁸⁾ Reliability tests the acquisition of reliable assessment results, and therefore, it is recommended that “the reliability of scores obtained through an assessment tool be measured”.⁽⁹⁾

The generalizability theory focuses on the generalization of assessment results to the universe.⁽¹⁰⁾ It

calculates a single reliability value by evaluating multiple error sources simultaneously. The G-theory enables the following: evaluating multiple sources of variance in a single analysis, identifying the size of each source of variance, calculating two different coefficients related to relative decision making based on individual performance and to absolute decision making based on individual performance (respectively; G coefficient and phi coefficient), and performing assessments in which the measuring error could be minimized (Decision “D” studies).^(11,12) In the present study, the G-theory, in which multiple error sources could be evaluated, involving reliability analysis and decision study was preferred.

The study aims to evaluate the multiple-choice assessment tool offered to students at Süleyman Demirel University online for 1st year, Committee 3 using the classical test and generalizability theory.

Method

Our study had a quantitative research design. The study universe was identified as 1st-year students who were actively studying at the Medical Faculty of Süleyman Demirel University (n:271). In the study, the multiple-choice assessment tool for Phase 1, Committee 3 offered to students online was analyzed.

The assessment tool was prepared by the faculty members of the relevant department, and it was comprised of 95 items in accordance with the learning objectives of Phase 1, Committee 3. This assessment tool was conducted through the faculty’s Learning Management System (MOODLE). In the assessment tool that had been prepared using the learning management system, exam security was ensured by student ID num-

bers, student passwords, IP addresses, and analysis of exam speed.

Students were able to reconnect after connection problems. Questions and answers were mixed for each implementation. The exam duration [the number of questions multiplied by 1.50] and exam implementation duration [the number of questions multiplied by 1.25] were calculated. 271 students at the phase participated in the online exam. There were no students who could not participate in the exam. After the implementation, the assessment tool was shared with students for further questions and score objection. Then, feedback on online assessment and evaluation was provided.

MS-Excel, SPSS, and EduG software programs were used in data analysis.^(13,14) Applications of generalizability were compared with SPSS and EduG.

Results

In our study, we evaluated the exam for Phase 1, Committee 3, held on April 24, 2020. The exam was accessible for 2 hours and 30 minutes. Two hundred seventy-one students participated in the process. When the exam is evaluated on a 100-point scale, it was found that the mean was 78.5±11.05 (min: 27.4; max: 98.0), the variance was 122.229, the kurtosis was -1.196 and skewness was 1.683. The general characteristics of this assessment are given in **Table 1**.

The mean item difficulty was found to be 0.785, and the mean discrimination index was 0.262, while the reliability coefficient (KR-20) was 0.902. High group min score (n=75) was 82.000 and low group max score (n=80) was 71.00. The mean score of questions, difficulty indices, and discrimination indices for all items on the assessment tool is provided in **Table 2**.

In the analysis of scores on the assessment tool using the G-theory, in a crossed design with a single surface, the relative percentage of the estimated variance component for individuals was 7.2%, percentage of the estimated variance component for items was 17.8%, and percentage of the estimated variance component for individual-item was 75% (**Table 3**).

G value was found to be 0.91, and the Phi value was 0.90 for the 95-item exam. G and Phi values calculated in the decision (D) study conducted with the

Table 1: General features of the assessment

Exam name	1st Year Board 3 Exam
Start the exam	Friday, April 24, 2020, 14:00
Finish the exam	Friday, April 24, 2020, 16:30
Open	150 min
Application time	120 min
Number of first applications	271
Overall average	74,58
Total scored applications	271
Median grade (for the highest rated application)	% 81,05
Standard deviation (for the highest rated application)	% 11,66
Grade distribution skew (for the highest score application)	-1,2024
Grade distribution kurtosis (for highest rated application)	1,7366
Coefficient of internal consistencies (for highest-scored implementation)	90,18%
Error rate (for highest rated execution)	31,34 %
Standard error (for highest rated execution)	% 3,65

Table 2: Descriptive features of the assessment tool

Item No	Difficulty Index	Standard deviation	Random prediction score	Effective weight	Discrimination Index	Point biserial correlation coefficient
1	% 65,68	% 47,56	% 1,05	% 1,21	0,26	0,26
2	% 85,24	% 35,54	% 1,05	% 1,24	0,32	0,42
3	% 89,30	% 30,97	% 1,05	% 1,08	0,22	0,37
4	% 81,18	% 39,16	% 1,05	% 1,33	0,37	0,44
5	% 78,60	% 41,09	% 1,05	% 1,16	0,3	0,32
6*	% 95,94	% 19,77	% 1,05	% 0,81	0,11	0,33
7	% 91,88	% 27,36	% 1,05	% 1,06	0,22	0,4
8*	% 97,05	% 16,96	% 1,05	% 0,73	0,05	0,31
9	% 49,45	% 50,09	% 1,05	% 1,06	0,26	0,22
10	% 43,17	% 49,62	% 1,05	% 0,87	0,17	0,15
11	% 78,60	% 41,09	% 1,05	% 1,02	0,2	0,25
12	% 77,12	% 42,08	% 1,05	% 1,21	0,32	0,34
13*	% 95,57	% 20,61	% 1,05	% 0,87	0,13	0,36
14	% 90,77	% 28,99	% 1,05	% 0,86	0,17	0,25
15*	% 96,68	% 17,95	% 1,05	% 0,67	0,07	0,24
16	% 83,03	% 37,61	% 1,05	% 1,41	0,43	0,52
17	% 94,46	% 22,91	% 1,05	% 0,99	0,16	0,41
18	% 87,08	% 33,60	% 1,05	% 1,14	0,26	0,38
19*	% 96,31	% 18,89	% 1,05	% 0,78	0,11	0,31
20	% 77,49	% 41,84	% 1,05	% 1,38	0,4	0,45
21	% 90,41	% 29,51	% 1,05	% 0,74	0,1	0,18
22	% 75,28	% 43,22	% 1,05	% 1,39	0,45	0,44
23	% 92,62	% 26,19	% 1,05	% 0,81	0,12	0,25
24	% 92,25	% 26,79	% 1,05	% 1,12	0,22	0,46
25	% 83,03	% 37,61	% 1,05	% 1,26	0,32	0,41
26	% 90,41	% 29,51	% 1,05	% 1,09	0,24	0,4
27	% 77,86	% 41,60	% 1,05	% 1,35	0,42	0,43
28	% 88,19	% 32,33	% 1,05	% 1,24	0,32	0,46
29*	% 95,57	% 20,61	% 1,05	% 0,68	0,09	0,22
30	% 56,83	% 49,62	% 1,05	% 1,39	0,43	0,38
31	% 81,92	% 38,56	% 1,05	% 0,96	0,28	0,23
32*	% 96,68	% 17,95	% 1,05	% 0,81	0,11	0,36
33*	% 96,31	% 18,89	% 1,05	% 0,68	0,07	0,24
34	% 93,36	% 24,95	% 1,05	% 0,60	0,07	0,14
35	% 87,45	% 33,19	% 1,05	% 0,86	0,18	0,22
36	% 83,39	% 37,28	% 1,05	% 0,80	0,18	0,17
37	% 38,01	% 48,63	% 1,05	% 1,19	0,39	0,28
38	% 49,45	% 50,09	% 1,05	% 1,28	0,38	0,32
39	% 48,34	% 50,06	% 1,05	% 1,40	0,47	0,38
40	% 77,12	% 42,08	% 1,05	% 1,08	0,27	0,27
41	% 62,73	% 48,44	% 1,05	% 0,97	0,22	0,19
42	% 80,81	% 39,45	% 1,05	% 1,08	0,3	0,29
43	% 76,75	% 42,32	% 1,05	% 1,50	0,57	0,52
44	% 80,44	% 39,74	% 1,05	% 1,24	0,34	0,38
45	% 64,94	% 47,80	% 1,05	% 1,53	0,61	0,48
46	% 84,50	% 36,26	% 1,05	% 1,17	0,26	0,37
47	% 74,91	% 43,43	% 1,05	% 1,18	0,36	0,31
48	% 74,91	% 43,43	% 1,05	% 1,43	0,46	0,46
49	% 70,48	% 45,70	% 1,05	% 1,10	0,24	0,26
50	% 81,92	% 38,56	% 1,05	% 0,96	0,22	0,23
51	% 91,14	% 28,46	% 1,05	% 0,92	0,18	0,29
52	% 88,93	% 31,43	% 1,05	% 0,96	0,15	0,29
53	% 80,44	% 39,74	% 1,05	% 1,20	0,35	0,35
54	% 91,14	% 28,46	% 1,05	% 1,06	0,22	0,39
55	% 93,73	% 24,29	% 1,05	% 0,84	0,14	0,28
56	% 93,73	% 24,29	% 1,05	% 0,86	0,11	0,29
57	% 84,50	% 36,26	% 1,05	% 1,14	0,22	0,35
58	% 47,23	% 50,02	% 1,05	% 1,03	0,25	0,21
59	% 84,13	% 36,60	% 1,05	% 1,33	0,39	0,47
60	% 60,15	% 49,05	% 1,05	% 1,16	0,34	0,27
61	% 63,84	% 48,14	% 1,05	% 1,13	0,29	0,26
62	% 67,53	% 46,91	% 1,05	% 1,47	0,51	0,45
63	% 53,87	% 49,94	% 1,05	% 1,61	0,68	0,51
64*	% 96,31	% 18,89	% 1,05	% 0,80	0,1	0,33
65	% 90,41	% 29,51	% 1,05	% 1,18	0,25	0,46
66	% 35,06	% 47,80	% 1,05	% 1,16	0,27	0,27
67	% 47,23	% 50,02	% 1,05	% 1,30	0,47	0,33
68	% 73,06	% 44,45	% 1,05	% 1,34	0,39	0,4
69	% 94,10	% 23,61	% 1,05	% 0,86	0,1	0,3
70	% 76,38	% 42,55	% 1,05	% 1,17	0,35	0,32
71	% 88,19	% 32,33	% 1,05	% 1,22	0,25	0,45
72	% 69,74	% 46,02	% 1,05	% 1,31	0,38	0,36
73	% 67,16	% 47,05	% 1,05	% 1,28	0,31	0,34
74	% 72,32	% 44,82	% 1,05	% 1,35	0,43	0,4
75	% 86,35	% 34,40	% 1,05	% 1,34	0,36	0,51
76	% 68,63	% 46,48	% 1,05	% 1,50	0,52	0,47
77	% 60,52	% 48,97	% 1,05	% 1,39	0,48	0,39
78	% 83,39	% 37,28	% 1,05	% 1,15	0,31	0,34
79	% 91,88	% 27,36	% 1,05	% 1,23	0,26	0,54
80	% 91,88	% 27,36	% 1,05	% 1,04	0,22	0,38
81	% 92,99	% 25,58	% 1,05	% 1,09	0,2	0,45
82	% 91,14	% 28,46	% 1,05	% 1,20	0,26	0,5
83*	% 95,57	% 20,61	% 1,05	% 0,86	0,14	0,35
84	% 73,80	% 44,05	% 1,05	% 0,88	0,26	0,17
85	% 43,54	% 49,67	% 1,05	% 1,18	0,38	0,28
86	% 35,42	% 47,92	% 1,05	% 1,02	0,31	0,21
87	% 85,98	% 34,79	% 1,05	% 1,04	0,23	0,3
88	% 42,80	% 49,57	% 1,05	% 0,92	0,17	0,17
89	% 28,78	% 45,36	% 1,05	% 0,71	0,14	0,11
90*	% 99,26	% 8,57	% 1,05	% 0,35	0,03	0,14
91*	% 98,89	% 10,48	% 1,05	% 0,21	0,03	0,04
92	% 92,99	% 25,58	% 1,05	% 0,57	0,09	0,12
93*	% 96,68	% 17,95	% 1,05		0,02	0
94*	% 99,26	% 8,57	% 1,05	% 0,38	0,03	0,17
95	% 86,35	% 34,40	% 1,05	% 0,54	0,06	0,08

* marks potential problems (p<0,2 or p>0,95, D<0, pbis<0, adjpbis<0) These results have been sorted by item number

number of items on the assessment tool are given in **Table 4.**

Discussion

The decision of CoHE on the fact that final exams and other exams during the spring semester of 2019-2020 academic year could not be held face-to-face due to the COVID-19 pandemic and on implementing alternative methods via digital means or assignments and projects were notified to universities on May 11.⁽¹⁾ In this context, in the evaluation process in distance learning, online supervised or unsupervised open-ended/multiple-choice exams, assignments, online quizzes, projects, activities on Learning Management System (LMS), the use of LMS analytics and similar practices could be used.

Based on “transparency and controllability” on online exams, it was recommended that exam security measures be implemented according to the LMS or digital means. Among these measures were a ran-

dom selection of questions, functioning of full screen or browser lock.

First of all, in online assessment and evaluation practices, the approach, system, and main principles for the assessment and evaluation system were identified at our faculty with the help of literature.⁽¹⁵⁻¹⁶⁾ The online service infrastructure of the faculty was evaluated in this context. Before assessment and evaluation practices, trial practices and analyses, comprised of a few questions, were conducted for the adaptation of students into the system. In line with these evaluations, it was decided that the assessment and evaluation practice in the learning management system would meet the requirements of this service.

It had been recommended that information and training activities be provided for faculty members and students on the use of online exams at universities. Faculty members and students were provided with information and technical support regularly on our faculty’s website. Practice principles including the cour-

Table 3. Estimated variance components and percentages (ANOVA table)

Variance source *	Sum of Squares	Degree of Freedom	Squares Average	Variance	Percentage of Variance	SE
Student	348.67415	270	1.29139	0.01226	7.2	0.00117
Item	777.25438	94	8.26866	0.03004	17.8	0.00440
SXI	3218.51404	25380	0.12681	0.12681	75.0	0.00113
Total	4344.44257	25744	1.29139	0.01226	100%	

Table 4: D-Study on the Number of Assessment Tool Items

	80 Item	90 Item	95 Item (current application)	100 Item	110 Item	120 Item
G coefficient	0.89	0.90	0.90	0.91	0.91	0.92
Phi coefficient	0.86	0.88	0.88	0.89	0.90	0.90

se of the exam, exam duration, grading principles, student responsibilities, exam objections and ethical rules on online assessment/evaluation were prepared and published on the faculty's website.⁽¹⁷⁾ Then, an assessment tool, comprised of items in accordance with learning objectives, was formed.

A trial assessment/evaluation practice was conducted for the adaptation of students into the system. After the improvements in line with the feedback provided during this preparation phase, the implementation period started. After the implementation, the assessment tool was shared with students for further questions and score objection. Then, feedback on online assessment and evaluation was provided.

It is recommended that an assessment tool be valid, reliable, and practical.^(7,18,19) Since subject-matter expert faculty members in accordance with the learning objectives of the faculty's accredited education program prepared the online Phase 1, Committee 3 exam held on April 24, 2020, the scope validity of the study was met. When evaluated in terms of practicality, the exam demonstrates that it is practical in terms of item entry, the course of exams, and the satisfaction of student feedback.

When evaluated in terms of reliability, the exam implementation was found to be 0.90, which was reliable according to the classical test and generalizability theories. The mean item difficulty was 0.785, and the mean discrimination index was 0.262 in item analyses of the assessment tool. Based on analyses 14 potential problematic questions were identified. 13 of these items were considered problematic because the difficulty level was greater than 0.95, and Item 93 had both

the difficulty level was greater than and negative point biserial correlation coefficient. Feedback was provided in order to improve the item-level quality of assessment tools.

Moreover, in the analysis of scores on the assessment tool using the G-theory, in a crossed design with a single surface, the fact that the relative value of the estimated variance component percentage for individuals is low suggests that it is insufficient. The fact that the estimated variance component percentage for items has an unbalanced distribution at difficulty levels weakens the generalizability.

In contrast, the size of the estimated variance component for individual-item suggests that systematic or non-systematic error sources cannot be controlled. Given the scope validity in the decision (D) study conducted with the number of items on the assessment tool, it is recommended that the number of questions is changed. The unique value of this study is that it presents the analysis, which has not been frequently seen in medical education, to the use of the field.

Although the evaluation of a single assessment tool is a limitation in our study, the study has provided valuable information regarding the review and status evaluation of analyses and also it is precious for the sustainability of future analysis. In line with recommendations of CoHE and Association for Evaluation and Accreditation of Medical Education Programs, our faculty has performed assessment-evaluation practices in distance education as well as ensuring their monitorization.

Along with the pandemic experience, new approaches will be developed for the practices proposed in the measurement and evaluation area.^(20,22) The most

important practical implication of this study is to enable the faculty's assessment and evaluation system to be monitored with sustainable analysis and the social implication of this study is its contribution to the training of physicians with the quality promised by the faculty goals in relation to increasing the quality in higher education. Based on the experience, we are of the opinion that online assessment and evaluation practices could be reliably used for summative assess-

ment during the pandemic and for formative assessment after the pandemic.

Acknowledgement

We thank Süleyman Demirel University Rector's Office, which supports the distance education approach of our faculty materially and spiritually during the distance education process.

Kaynaklar:

1. Yüksek Öğretim Kurulu. YÖK'ten Üniversitelerdeki Uzaktan Eğitime Yönelik Değerlendirme. <https://www.yok.gov.tr/Sayfalar/Haberler/2020/uzaktan-egitime-yonelik-degerlendirme.aspx> adresinden 08.05.2020 tarihinde erişilmiştir.
2. Yüksek Öğretim Kurulu. YÖK'ten Sınavlara İlişkin Karar. <https://www.yok.gov.tr/Sayfalar/Haberler/2020/yok-ten-sinavlara-iliskin-karar.aspx> adresinden 08.05.2020 tarihinde erişilmiştir.
3. TEPDAD. Mezuniyet Öncesi Tıp Eğitimi Ulusal Standartları-2020. <http://tepdad.org.tr/announcement/9> adresinden 08.05.2020 tarihinde erişilmiştir.
4. TEPDAD. COVID-19 Nedeniyle TEPDAD tarafından yapılan önerilerin tümü. <http://tepdad.org.tr/announcement/9> adresinden 04.06.2020 tarihinde erişilmiştir.
5. Van der Vleuten CPM. and Schuwirth LWT. Assessing professional competence: from methods to programmes. *Medical education*. 2005; 39(3): 309–17. doi:10.1111/j.1365-2929.2005.02094.x.
6. Ercan İ. and Kan İ. Ölçeklerde Güvenirlik ve Geçerlik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi* 2004; 30(3): 211–6.
7. Norcini JJ. and McKinley DW. Assessment methods in medical education. *Teaching and Teacher Education* 2007; 23(3): 239–50. doi:<https://doi.org/10.1016/j.tate.2006.12.021>.
8. Cronbach, LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334. doi: 10.1007/BF02310555.
9. Güler N. Genellenabilirlik kuramı ve SPSS ile GENOVA programlarıyla hesaplanan G ve K çalışmalarına ilişkin sonuçların karşılaştırılması. *Eğitim ve Bilim* 2009; 34(154): 93–104.
10. Brennan RL. *Generalizability theory*. New York, US: Springer-Verlag Publishing (Statistics for social science and public policy), 2001. doi: 10.1007/978-1-4757-3456-0.
11. Shavelson RJ. and Webb NM. *Generalizability theory: A primer. Vol.1*. Thousand Oaks, CA, US: Sage Publications, Inc (Meas-

Alıntı Kodu: Kolcu G. ve ark. Evaluation of the reliability of the online multiple choice assessment tool by the generalizability theory. *Jour Turk Fam Phy* 2021; 12 (1): 3-11. Doi: 10.15511/tjtfp.21.00003.

- urement methods for the social sciences series), 1991.
12. Atılğan H. Genellenebilirlik Kuramı ve Uygulaması. 1. Baskı. Ankara, 2019.
 13. Mushquash C. and O'Connor, BP. SPSS, SAS, and MATLAB programs for generalizability theory analyses. *Behavior Research Methods* 2006; 38(3): 542–7.
 14. EduG. English program, IRDP. Institut de recherche et de documentation pédagogique. <https://www.irdp.ch/institut/english-program-1968.html> adresinden 05.06.2020 tarihinde erişilmiştir.
 15. Al-Wardy NM. Assessment methods in undergraduate medical education. *Sultan Qaboos University medical journal* 2010;10(2): 203–9. doi: 10.4103/0331-8540.108463.
 16. Epstein RM. Assessment in Medical Education. *The new england journal of medicine* 2011; 364(January): 7–64. doi: 10.1056/NEJMr054784.
 17. Tıp Fakültesi Dekanlığı. Uzaktan Eğitim İle İlgili Duyurular. Süleyman Demirel Üniversitesi Tıp Fakültesi Dekanlığı. tip.sdu.edu.tr/tr/sayfalar/uzaktan-egitim-ile-ilgili-duyurular-11163s.html adresinden 10.06.2020 tarihinde erişilmiştir.
 18. Hays R. Assessment in medical education: Roles for clinical teachers. *Clinical Teacher* 2008; 5(1): 23–7. doi: 10.1111/j.1743-498X.2007.00165.x.
 19. Speyer R. et al. Reliability and validity of student peer assessment in medical education: A systematic review. *Medical Teacher* 2011;33(11): 572–85. doi: 10.3109/0142159X.2011.610835.
 20. Challis M. AMEE Medical Education Guide No.11 (revised): Portfolio-based learning and assessment in medical education. *Medical Teacher* 1999; 21(4): 370–86. doi: 10.1080/01421599979310.
 21. Pololi LH. et al. A needs assessment of medical school faculty: Caring for the caretakers. *Journal of Continuing Education in the Health Professions* 2003; 23(1): 21–9. doi: 10.1002/chp.1340230105.
 22. Durak HI. et al. Use of case-based exams as an instructional teaching tool to teach clinical reasoning. *Medical Teacher* 2007; 29(6): 170–4. doi: 10.1080/01421590701506866.

Alıntı Kodu: Kolcu G. ve ark. Evaluation of the reliability of the online multiple choice assessment tool by the generalizability theory. *Jour Turk Fam Phy* 2021; 12 (1): 3-11. Doi: 10.15511/tjtfp.21.00003.