



# Evaluating the Reliability and Validity of the Maastricht Clinical Teaching Questionnaire in Bahrain

Ahmed Al Ansari<sup>1,2\*</sup> and Khaled Saeed Tabbara<sup>3</sup>

<sup>1</sup>Training and Education Department, Bahrain Defense Force Hospital, Riffa, Bahrain

<sup>2</sup>Medical Education Department, College of Medicine and Medical Sciences, Arabian Gulf University, Manama, Bahrain

<sup>3</sup>Department of Microbiology, Immunology and Infection Diseases, College of Medicine and Medical Sciences, Arabian Gulf University, Manama, Bahrain

## ARTICLE INFO

### Article history:

Received: 1 August 2018

Accepted: 24 February 2019

### Online:

DOI 10.5001/omj.2019.78

### Keywords:

Reliability and Validity;  
Teaching Rounds.

## ABSTRACT

**Objectives:** The Maastricht Clinical Teaching Questionnaire (MCTQ) was developed to evaluate clinical teachers' supervisory skills during undergraduate clinical rotations. Evidence exists supporting the reliability and validity of this questionnaire. Our study sought to examine the reliability and validity of the MCTQ in a Middle Eastern context. **Methods:** Between 2016 and 2017, we evaluated a total of 549 medical students in their final year who were undergoing clinical rotations using the MCTQ. The construct validity was assessed using exploratory and confirmatory factor analysis. Internal consistency reliability was measured using Cronbach's alpha. Factor scores were compared with other outcomes to explore the relationship with other relevant variables. **Results:** A four-factor model demonstrated an adequate fit with the data. The findings showed good internal consistency reliability. The following results were obtained for the four-factor model: chi-square divided by degrees of freedom was 5.026, and the comparative index, goodness of fit index, normalized fit index, and non-normalized fit index were all above 0.800 (0.955, 0.858, 0.950, and 0.952, respectively). The standardized root mean square residual was 0.016, and the root mean square error of approximation score was 0.086. Acceptable reliability was achieved with 10 evaluations per teacher. We observed a strong correlation between factors and overall judgment. **Conclusions:** Our study suggests that the MCTQ is a valid and reliable instrument to evaluate teachers' performance during clinical rotations in Bahrain.

Clinical rotations in hospitals are an integral part of the medical school curricula and are a highly influential component of student learning. Clinical rotations also provide students with practical skills such as history taking, physical examination, and clinical reasoning.<sup>1</sup> Clinical rotations have been shown to enhance student development by improving their knowledge and skills as well as leading to other areas of development, which are arguably unachievable in a classroom context such as developing relationships, forming a sense of professional identity, and modifying attitudes.<sup>1-4</sup>

The quality of supervision provided by clinical tutors at these rotations is essential for the effectiveness of the teaching.<sup>2,4</sup> However, clinical staff nowadays experience increased demands, and this has been found to diminish the quality of clinical

education.<sup>5</sup> Therefore, providing an instrument to measure clinical tutor effectiveness at clinical rotations could be a beneficial tool to understand and enhance clinical supervision. The Maastricht Clinical Teaching Questionnaire (MCTQ) was developed by Stalmeijer et al,<sup>6</sup> to evaluate clinical teachers' supervisory skills during undergraduate clinical rotations in the medical curriculum.

The MCTQ is based on the cognitive apprenticeship model, which emphasizes teachers' cognitive processes while performing a complex task. It is made up of six teaching strategies: modeling, coaching, scaffolding, articulation, reflection, and exploration. Researchers have demonstrated effectiveness in implementing a cognitive apprenticeship model of instruction in educational settings.<sup>7,8</sup> The first teaching strategy 'modeling' occurs in two parts. Behavioral modeling is when

\*Corresponding author: ✉drahmedalansari@gmail.com

a person observes an experienced individual giving instructions. The second part is cognitive modeling, and that is when the instructor expresses their thought process. Coaching is when an instructor observes a student while he/she is performing a task and provides feedback. In scaffolding, the student demonstrates some form of mastery over the concept. In this stage, the instructor provides selective feedback. The fourth teaching method 'articulation' is when the students articulate their understanding of a specific task using an assessment of content mastery. Reflection occurs when the students are allowed to reflect on their understanding of the concept and strategies in problem-solving and compare them to other students or experts. Lastly, exploration is when the students discuss what they have learned and can understand how the information they learned can be applied.

The MCTQ has been validated and shown to be beneficial for evaluating clinical teachers during rotations.<sup>5,9</sup> This instrument could help improve the quality of education by identifying the areas of improvement where more training is needed and can also help faculty developments measure the return of their investment in training these identified areas.<sup>10</sup> Moreover, measuring teaching effectiveness is also beneficial in guiding, supporting, and motivating the clinical tutors to improve their teaching.<sup>11-13</sup>

The validity of this questionnaire has been established in different types of medical contexts (general and veterinary).<sup>5,9</sup> Additionally, the validity of this instrument has been demonstrated in different countries including the Netherlands, Australia, Canada, Columbia, Ireland, and the UK.<sup>5,14-18</sup> However, this study has not been validated in a Middle Eastern context such as Bahrain. Educational practices are context-specific; therefore, culture is a key component that can influence validity in instruments that evaluate the quality of teaching.<sup>19</sup> Demonstrating validity of this questionnaire across many different cultures will strengthen its use as well as demonstrate the generalizability of the questionnaire across different cultural contexts. This will also allow more people to use and benefit from this instrument.

Our study sought to evaluate the psychometric properties of the MCTQ in medical students in Bahrain. This will be achieved by assessing the construct validity, internal reliability, performing factor analysis, conducting a generalizability study,

and measuring the relationship between instrument scores and other variables with relevance to the construct being measured.

## METHODS

This study took place at the Arabian Gulf University (AGU) in Bahrain between 2016 and 2017. Questionnaires were distributed to all medical students who were asked to evaluate 98 clinical tutors at the university. A total of 549 questionnaire responses were collected.

Data was collected using the MCTQ.<sup>6</sup> The questionnaire is composed of 24 items, which involve a set of statements scored on a five-point Likert scale from fully disagree to fully agree. The second section of the questionnaire asks participants to provide an overall assessment of their tutor by rating their skills on a scale ranging from 1 to 10. A higher score indicates a better overall assessment.

Each item in the questionnaire belongs to one of six domains. The first domain is the general learning climate. This domain involves statements that intend to measure the extent to which the tutor creates a safe learning environment for the students and treats them with respect. The second domain is modeling, which assesses whether tutors carry out a task by acting as a role model and create an opportunity for students to observe and build a conceptual model of the process necessary to complete the task. The coaching domain measures whether the tutor observes students while performing different tasks and gives feedback. The articulation domain involves statements that measure whether the tutors ask students to explain their actions for them to be aware of gaps in their knowledge and skills to increase their understanding and motivate them to ask questions. The fifth domain is reflection, and it assesses whether the tutor encourages students to be aware of their strengths and weaknesses and consider what they can do to improve things. The final domain is exploration, which assesses whether the tutors encourage the students to formulate learning objectives based on identifying strengths and weaknesses and challenging the students to learn new things.

Exploratory factor analysis (EFA) was used to assess validity, and the analysis was conducted using SPSS Statistics (SPSS Inc. Released 2007. SPSS for

Windows, Version 16.0. Chicago, SPSS Inc). A Kaiser-Meyer-Olkin (KMO) analysis and Bartlett's test of sphericity were conducted to determine if the data was suitable to continue with principal component analysis.

Confirmatory factor analysis (CFA) was used to determine the construct validity of the MCTQ. IBM SPSS AMOS V.22 for Windows was used for this analysis. First, the normality of the distribution was assessed by calculating the skewness and kurtosis values of all the data. The data was normally distributed. The estimation method for CFA was maximum likelihood estimation. Several fit indices were used to evaluate the model fit. The fit indices taken into account were relative chi-square divided by degrees of freedom (CMIN/df), goodness of fit index (GFI), comparative fit index (CFI), root mean square error of approximation (RMSEA), non-normalized fit index (NNFI), normalized fit index (NFI), and standardized root mean square residual (SRMR).<sup>20,21</sup> A CMIN/df value < 2 indicated a good fit.<sup>17</sup> The GFI, CFI, NFI, and NNFI values range from 0–1, values  $\geq 0.80$  indicate an acceptable model fit.<sup>22,23</sup> An RMSEA value between 0.08 and 0.10 suggests an average model of fit and a value < 0.08 demonstrates a good fit. SRMR values < 0.08 demonstrate a good fit.<sup>24</sup>

Inter-rater reliability was assessed using a generalizability study to determine the number of student ratings required to provide reliable feedback for teachers using the factors derived from the CFA. For this analysis, uRGeneva (G-String-IV) version 6.3.8 was used. A generalizability (G) coefficient of at least 0.70 was required to demonstrate good reliability.

Moreover, Cronbach's alphas were computed for each scale to determine internal consistency reliability. Coefficients > 0.70 were considered acceptable. This analysis was conducted using SPSS.

The AGU research ethical committee approved the study, which was in accordance with the Declaration of Helsinki and comparable ethical standards. All participants were assured that participation was voluntary and that their data would remain anonymous and may be published.

## RESULTS

Table 1 includes the average score responses for each item of the questionnaire including the overall tutor

**Table 1:** The Maastricht Clinical Teaching Questionnaire items and average response for each item. Total responses = 549.

Questionnaire items	Mean	SD
Q1 Consistently demonstrated how different tasks should be performed.	4.4	0.8
Q2 Clearly explained the important elements for the execution of a given task.	4.4	0.8
Q3 Created sufficient opportunities for me to observe them.	4.5	0.8
Q4 Was a role model as to the kind of health professional I wish to become.	4.4	0.8
Q5 Observed me multiple times during patient encounters.	4.4	0.9
Q6 Provided me with useful feedback during or following direct observation of patient encounters.	4.4	0.8
Q7 Helped me understand which aspects I needed to improve.	4.4	0.8
Q8 Adjusted teaching activities to my level of experience.	4.4	0.8
Q9 Offered me sufficient opportunities to perform activities independently.	4.4	0.9
Q10 Supported me in activities I find difficult to perform.	4.3	0.9
Q11 Gradually reduced the support given to allow me to perform certain activities more independently.	4.4	0.8
Q12 Asked me to provide a rationale for my actions.	4.4	0.8
Q13 Helped me to become aware of gaps in my knowledge and skills.	4.4	0.8
Q14 Asked me questions aimed at increasing my understanding.	4.5	0.8
Q15 Encouraged me to ask questions to increase my understanding.	4.4	0.8
Q16 Stimulated me to explore my strengths and weaknesses.	4.4	0.8
Q17 Stimulated me to consider how I might improve my strengths and weaknesses.	4.4	0.8
Q18 Encouraged me to formulate learning goals.	4.4	0.8
Q19 Encouraged me to pursue my learning goals.	4.4	0.8
Q20 Encouraged me to learn new things.	4.4	0.8
Q21 Created a safe learning environment.	4.4	0.8
Q22 Took sufficient time to supervise me.	4.4	0.9
Q23 Was genuinely interested in me as a student.	4.4	0.9
Q24 Showed me respect.	4.4	0.9
<b>Overall tutor assessment out of 10</b>	<b>8.8</b>	<b>1.8</b>

SD: standard deviation.

assessment. To assess the validity of MCTQ, an EFA was conducted. The EFA revealed a KMO value of

**Table 2:** Exploratory factor analysis output.

Questionnaire items		Components			
		1	2	3	4
Q1	Consistently demonstrated how different tasks should be performed.	-	-	-	0.663
Q2	Clearly explained the important elements for the execution of a given task.	-	-	-	0.671
Q3	Created sufficient opportunities for me to Observe them.	-	-	-	0.646
Q4	Was a role model as to the kind of health professional I wish to become.	-	-	-	0.581
Q5	Observed me multiple times during patient encounters.	-	0.716	-	-
Q6	Provided me with useful feedback during or following direct observation of patient encounters.	-	0.626	-	-
Q7	Helped me understand which aspects I needed to improve.	-	0.543	-	-
Q8	Adjusted teaching activities to my level of experience.	-	0.581	-	-
Q9	Offered me sufficient opportunities to perform activities independently.	-	0.708	-	-
Q10	Supported me in activities I find difficult to perform.	-	0.636	-	-
Q11	Gradually reduced the support given to allow me to perform certain activities more independently.	-	0.573	-	-
Q12	Asked me to provide a rationale for my actions.	0.604	-	-	-
Q13	Helped me to become aware of gaps in my knowledge and skills.	0.580	-	-	-
Q14	Asked me questions aimed at increasing my understanding.	0.687	-	-	-
Q15	Encouraged me to ask questions to increase my understanding.	0.666	-	-	-
Q16	Stimulated me to explore my strengths and weaknesses.	0.675	-	-	-
Q17	Stimulated me to consider how I might improve my strengths and weaknesses.	0.658	-	-	-
Q18	Encouraged me to formulate learning goals.	0.657	-	-	-
Q19	Encouraged me to pursue my learning goals.	0.596	-	-	-
Q20	Encouraged me to learn new things.	-	-	0.539	-
Q21	Created a safe learning environment.	-	-	0.614	-
Q22	Took sufficient time to supervise me.	-	-	0.684	-
Q23	Was genuinely interested in me as a student.	-	-	0.735	-
Q24	Showed me respect.	-	-	0.690	-

0.976, and Bartlett's test of sphericity was significant ( $p < 0.001$ ). The results were sufficient enough to allow the analysis to continue with a principal component approach.

The item-loadings were first observed in the sixth-factor model. However, in the sixth-factor model, only two items loaded on the fifth component (minimum items per component is three). The next step was to examine the factor loadings in a five-factor model. This yielded unsatisfactory results as well as no items loaded on the fifth factor. In the four-factor alternative, each factor had more than three items, and all values had a high loading ( $> 0.45$ ). Thus, a four-factor model with 24 items was more appropriate for this study. The total four factors explained an overall 87.1% of the total variance [Table 2].

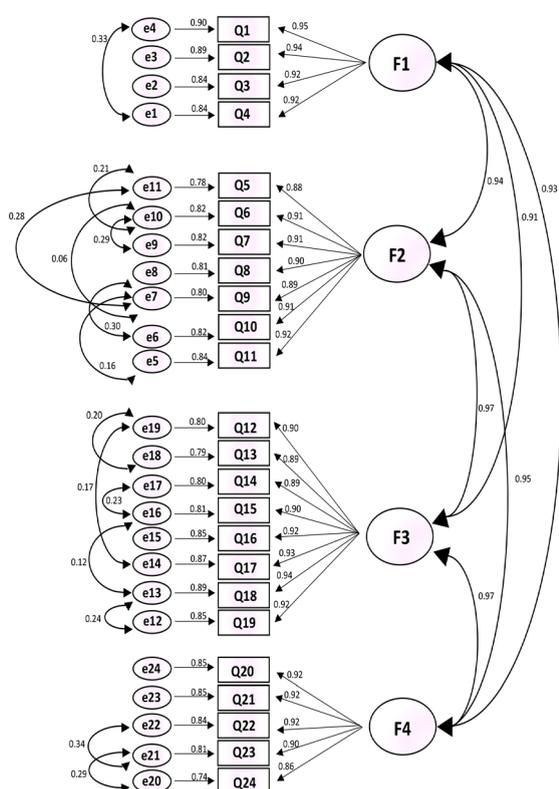
The CFA revealed that the original 24-item model with six factors did not fit the data [Table 3]. After reducing the number of factors and reorganizing items according to the modification indices, an acceptable model of fit was found for a 24-item questionnaire with four factors. The following results were obtained for the four-factor model: CMIN/df was 5.026, CFI, GFI, NFI, and NNFI were all  $> 0.80$  (0.955, 0.858, 0.950, and 0.952, respectively), SRMR was 0.016 and the RMSEA score was 0.086. The results show that all indices met the criteria for a reasonable model of fit, except for CMIN/df. Moreover, the correlations between the factors and with overall tutor assessment varied between 0.874 and 0.930.

The results from the CFA are presented in Figure 1. The standardized solution is given, and

**Table 3:** The model fit indices.

Numbers of factors	CMIN/df	CFI	GFI	SRMR	RMSEA	NFI	NNFI
4	5.026	0.955	0.858	0.016	0.086	0.950	0.952
6	5.428	0.951	0.856	0.017	0.090	0.940	0.941

CMIN/df: chi-square divided by degrees of freedom; CFI: comparative fit index; GFI: goodness of fit index; SRMR: standardized root mean square residual; RMSEA: root mean square error of approximation; NFI: normalized fit index; NNFI: non-normalized fit index.



**Figure 1:** The four-factor path diagram.

**Table 4:** Summary of the generalizability (G) coefficients ( $p^2$ ) for decision study.

Raters, n	G-coefficient
4	0.504
5	0.560
6	0.604
7	0.640
8	0.670
9	0.696
10	0.718
11	0.734
12	0.753

parameters estimates for latent factors are shown. The lambda-ksi estimates are analogous of factors

loadings in EFA, and values of 0.79 or higher for these parameters indicate well-defined latent constructs. The lambda-ksi estimates range from 0.74–0.95. Learning environment and modeling demonstrate the greatest variability in the magnitude of their estimates with values ranging from 0.74–0.92 and 0.79–0.92, respectively. This variability corresponds to their lower internal reliability coefficients relative to the other teaching scales coefficient alphas are 0.962 and 0.960, respectively.

Cronbach’s alpha coefficient for all factors ranged between 0.960–0.976; this indicated high internal consistency. High internal consistency reliability was found in this study ( $\alpha = 0.980$ ). All scales had high reliability ( $> 0.700$ ). Articulation and reflection scales had the highest reliability coefficient ( $\alpha = 0.976$ ) followed by coaching ( $\alpha = 0.970$ ), learning environment ( $\alpha = 0.962$ ), and modeling ( $\alpha = 0.960$ ).

The results of the generalizability study demonstrated that the variance associated with tutors for the overall judgment was 0.132, and the variance associated with students nested within tutors was 0.520. The G-coefficient was 0.504.

Results from a decision study are found in Table 4, which provides the G-coefficients per factor as a function of the number of student’s responses. At least 10 student ratings are required to give teachers reliable feedback.

## DISCUSSION

This study aimed to evaluate the reliability and validity of the MCTQ in evaluating clinical teachers during clinical rotations in the Middle East. This study assessed the construct validity of the MCTQ as an instrument to draw out student’s feedback on the teaching quality of individual teachers based on the Collins et al, (1989) teaching methods in the cognitive apprenticeship model.<sup>25</sup> The CFA produced a four-factor model with 24 items. The result of CFA demonstrated that the four-factor

model fits the data reasonably well with all but one of the statistical criteria met. Only chi-square divided by degrees of freedom did not meet the criteria for a reasonable fit ( $CMIN/df = 5.026 > 2$ ).

The four-dimensional scale found in this study demonstrates similarity with a previous study by Stalmeijer et al.<sup>9</sup> In this study, articulation and reflection were combined as one factor; this is not surprising because articulation and reflection both tend to stimulate self-regulated learning. Additionally, the high correlation between factors and with factors and overall judgment seems to support the validity of this questionnaire.

Moreover, this study demonstrated good internal consistency reliability. The generalizability study findings showed that 10 student ratings are required to provide reliable data. This number is easily obtainable in most clinical settings. Thus, the results demonstrate that the MCTQ is a reliable instrument to be used in clinical educational settings in the Middle East. This expands the use of this questionnaire to another cultural context.

The finding that the MCTQ is valid and reliable in the Middle East adds to the literature suggesting that the MCTQ is valid and reliable in other contexts such as general and veterinary medical contexts in the Netherlands.<sup>5,9</sup>

These findings may not be generalizable to medical students in other years. This study was conducted only in final year medical students. However, the applicability of the cognitive apprenticeship model could vary across different stages of education.<sup>5</sup> For example, students in earlier years may require more supervision and guidance, and so their responses will differ from those in their final year. Future research should incorporate findings from students in other levels of education in Bahrain to demonstrate that this tool can be used in Bahrain for medical students at different stages.

Another possible limitation is that the study does not address the MCTQ's effectiveness at improving teaching in Bahrain. This has been suggested by the American Psychological Education as another source of validity evidence.<sup>26</sup> Whether or not the tutors will respond to the feedback and improve their skills is important to identify the effectiveness of this questionnaire. Future studies should investigate the effect the MCTQ can have on changing the teaching behaviors of the tutors in Bahrain, and therefore enhancing the students' education.

## CONCLUSION

Student ratings are easy to obtain,<sup>27</sup> and can be used as a tool to understand clinical education during rotations and provide an appropriate baseline to make changes and enhance student education. This study demonstrated that a four-factor model of the MCTQ is a valid and reliable tool to evaluate clinical teachers' supervisory skills during medical rotations in Bahrain, and by extension, in other Middle Eastern countries.

### Disclosure

The authors declared no conflict of interest. No funding was received for this study.

### REFERENCES

1. Smith BP, Walsh D. Teaching the art of clinical practice: the veterinary medical teaching hospital, private practice, and other externships. *J Vet Med Educ* 2003;30(3):203-206.
2. Dolmans DH, Wolfhagen HA, Essed GG, Scherpbier AJ, Van Der Vleuten CP. Students' perceptions of relationships between some educational variables in the out-patient setting. *Med Educ* 2002 Aug;36(8):735-741.
3. Smith SE, Tallentire VR, Cameron HS, Wood SM. The effects of contributing to patient care on medical students' workplace learning. *Med Educ* 2013 Dec;47(12):1184-1196.
4. Spencer J. Learning and teaching in the clinical environment. *BMJ* 2003 Mar;326(7389):591-594.
5. Boerboom TB, Dolmans DH, Jaarsma AD, Muijtjens AM, Van Beukelen P, Scherpbier AJ. Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations. *Med Teach* 2011;33(2):e84-e91.
6. Stalmeijer RE, Dolmans DH, Wolfhagen IH, Muijtjens AM, Scherpbier AJ. The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity. *Med Teach* 2008;30(8):e272-e277.
7. Collins A, Brown JS, Holum A. Cognitive apprenticeship: Making thinking visible. *Am Educ* 1991;15(3):6-11.
8. Kilminster SM, Jolly BC. Effective supervision in clinical practice settings: a literature review. *Med Educ* 2000 Oct;34(10):827-840.
9. Stalmeijer RE, Dolmans DH, Wolfhagen IH, Muijtjens AM, Scherpbier AJ. The Maastricht clinical teaching questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med* 2010 Nov;85(11):1732-1738.
10. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med* 1998 Jun;73(6):688-695.
11. Snell L, Tallett S, Haist S, Hays R, Norcini J, Prince K, et al. A review of the evaluation of clinical teaching: new perspectives and challenges. *Med Educ* 2000 Oct;34(10):862-870.
12. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med* 2000 Feb;75(2):161-166.
13. Ramani S, Leinster S. AMEE Guide no. 34: Teaching in the clinical environment. *Med Teach* 2008;30(4):347-364.
14. Vaughan B. Developing a clinical teaching quality questionnaire for use in a university osteopathic pre-

- registration teaching program. Research Gate. BMC Med Educ 2015 Apr 8;15:70.
15. Stalmeijer RE, Dolmans DH, Wolfhagen IH, Peters WG, van Coppenolle L, Scherpbier AJ. Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Adv Health Sci Educ Theory Pract* 2010 Aug;15(3):315-328.
  16. Olmos-Vega F, Dolmans D, Donkers J, Stalmeijer RE. Understanding how residents' preferences for supervisory methods change throughout residency training: a mixed-methods study. *BMC Med Educ* 2015 Oct;15(1):177.
  17. Kelly M, Bennett D, McDonald P. Evaluation of clinical teaching in general practice using the Maastricht clinical teaching questionnaire. *Med Teach* 2012;34(12):1089-1089.
  18. Boerboom TB, Mainhard T, Dolmans DH, Scherpbier AJ, Van Beukelen P, Jaarsma AD. Evaluating clinical teachers with the Maastricht clinical teaching questionnaire: how much 'teacher' is in student ratings? *Med Teach* 2012;34(4):320-326.
  19. Kikukawa M, Stalmeijer RE, Emura S, Roff S, Scherpbier AJ. An instrument for evaluating clinical teaching in Japan: content validity and cultural sensitivity. *BMC Med Educ* 2014 Aug;14(1):179.
  20. Gable R, Wolf M. Instrument development in the affective domain. Boston: Kluwer Academic Publishers; 1993.
  21. Harrington D. Confirmatory factor analysis. Oxford: Oxford University Press; 2009.
  22. Kline RB. Principles and practice of structural equation modeling. 3rd ed. Guilford publications; 2015.
  23. Byrne BM. Structural equation modeling with AMOS. Mahwah: Lawrence Erlbaum Associates; 2001.
  24. Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6(1):1-55.
  25. Collins A, Brown JS, Newman SE. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In: Resnick LB, editor. *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.; 1989. p. 453-494.
  26. American Education Research Association and American Psychological Association. Standards for educational and psychological testing. Washington, DC: American Education Research Association. 1999 [cited 2018 July]. Available from: <https://www.directtextbook.com/isbn/9780935302356>.
  27. Beckman TJ, Cook DA, Mandrekar JN. Factor instability of clinical teaching assessment scores among general internists and cardiologists. *Med Educ* 2006 Dec;40(12):1209-1216.