



Vahid Ghavami<sup>1</sup>, Jamshid Jamali<sup>1\*</sup>, Maryam SarKhosh<sup>2</sup>

<sup>1</sup>Department of Biostatistics and Epidemiology, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran

<sup>2</sup>Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

\*Mashhad University of Medical Sciences, Daneshgah St. Mashhad, 9137673119 Iran

Tel: +98 5131892706

Fax: +98 5131892151

Email: Jamalij@mums.ac.ir

### Assessment of the Construct Validity of the Evaluation of Faculty Members by the Students Questionnaire at the School of Health of Mashhad University of Medical Sciences

**Background:** From the students' viewpoint the evaluation of teachers is one of the most important tasks of faculty evaluation aiming to strengthen the abilities, modify teachers' weaknesses, and improve the quality of teaching system. The present study aimed to determine the suitable structure to evaluate faculty members by the students' questionnaire, at the Health School of Mashhad University of Medical Sciences (MUMS), in the first semester of 2018-2019.

**Methods:** The present research is a descriptive-analytic cross-sectional study. The statistical population consisted of students from the School of Health of MUMS, in the first semester of 2018-2019. The Evaluation of Faculty Members by the Students Questionnaire (EFSQ) used at the School of Health of MUMS consisted of 12 questions with a 10-point Likert scale and 4 dimensions. This study used the exploratory and confirmatory factor analysis to assess the suitable structure of the EFSQ.

**Results:** The mean of the total score of the EFSQ was  $8.77 \pm 1.30$  out of 10. The four-dimensional structure was not respectable, and the three-dimensional structure was appropriate.

**Conclusion:** The current EFSQ used in MUMS is inadequate, and it is necessary to be revised.

**Keywords:** Education, Evaluation, Faculty members, Student, Mashhad University of Medical Sciences, Reproducibility of results

مناقشة صحة هيكل استبيان تقييم الأساتذة من قبل طلاب كلية الصحة بجامعة مشهد للعلوم الطبية

الخلفية والهدف: يعتبر تقييم أعضاء هيئة التدريس من قبل الطلاب أحد أهم طرق تقييم الأساتذة، والتي تهدف إلى تعزيز نقاط القوة وإصلاح نقاط الضعف لدى أعضاء الهيئة التعليمية وبالتالي تحسين النظام التعليمي في نهاية المطاف. كان الهدف من هذه الدراسة هو تحديد الهيكل المناسب لنموذج تقييم أعضاء هيئة التدريس من وجهة نظر الطلاب في كلية الصحة في جامعة مشهد للعلوم الطبية في عام 2019. الطريقة: هذه دراسة مقطعية وصفية تحليلية، والسكان الإحصائيون لهذه الدراسة هم طلاب كلية الصحة بجامعة مشهد للعلوم الطبية في النصف الأول من العام الدراسي 2019-2020. يتكون نموذج التقييم المستخدم لتقييم أعضاء هيئة التدريس من قبل طلاب كلية الصحة العامة في مشهد من 12 سؤالاً بمقياس ليكرت من 10 نقاط وأربعة مجالات. لتقييم الهيكل المناسب لنموذج التقييم، تم استخدام تحليل عامل الاستكشاف والتأكيد.

النتائج: بلغ متوسط الدرجة الكلية لنموذج تقييم أعضاء هيئة التدريس من قبل الطلاب  $8.77 \pm 1.30$  من 10، وكشفت نتائج تحليل عامل التأكيد على عدم تأكيد مجالات التقييم الأربعة وتأكيد بنية المجال الثلاثة.

الخلاصة: الشكل الحالي لتقييم أعضاء هيئة التدريس غير مناسب في وجهة نظر طلاب كلية الصحة العامة في مشهد، وهناك حاجة لمراجعة هذا النموذج.

الكلمات المفتاحية: التعليم، التقييم، أعضاء الهيئة التدريسية، طالب، جامعة مشهد للعلوم الطبية، تكرار النتائج

### ارزیابی روایی ساختار پرسشنامه ارزشیابی استاد توسط دانشجو دانشکده بهداشت دانشگاه علوم پزشکی مشهد

**زمینه و هدف:** ارزشیابی اساتید از دیدگاه دانشجویان، از مهمترین شیوه های ارزشیابی اساتید محسوب می گردد که هدف آن تقویت نقاط قوت و اصلاح نقاط ضعف اساتید و در نهایت ارتقاء سیستم آموزشی است. مطالعه حاضر با هدف تعیین ساختار مناسب فرم ارزشیابی اساتید از دیدگاه دانشجویان دانشکده بهداشت مشهد در سال ۱۳۹۷ انجام شد. **روش:** مطالعه حاضر یک مطالعه مقطعی و از نوع توصیفی-تحلیلی می باشد که جامعه آماری آن را دانشجویان دانشکده بهداشت دانشگاه علوم پزشکی مشهد در نیمسال اول سال تحصیلی ۹۸-۱۳۹۷ تشکیل می دهند. فرم ارزشیابی مورد استفاده برای ارزشیابی اساتید توسط دانشجویان دانشکده بهداشت مشهد، مشتمل بر ۱۲ سؤال با طیف لیكرت ده امتیازی و چهار حیطه است. برای ارزیابی ساختار مناسب فرم ارزشیابی، از تحلیل عاملی اکتشافی و تأییدی استفاده گردید.

**یافته ها:** میانگین نمره کل فرم ارزشیابی اساتید توسط دانشجویان  $8.77 \pm 1.30$  از ۱۰ بود. نتایج تحلیل عاملی تأییدی بیانگر عدم تأیید فرم چهار حیطه ای ارزشیابی و تأیید ساختار سه حیطه ای بود.

**نتیجه گیری:** فرم کنونی ارزشیابی اساتید از دیدگاه دانشجویان دانشکده بهداشت مشهد، دارای ساختاری نامناسب می باشد و نیاز به بازنگری دارد.

**واژه های کلیدی:** آموزش، ارزشیابی، اعضای هیأت علمی، دانشجو، دانشگاه علوم پزشکی مشهد، تکراریذیری نتایج

### مشهد یونیورسٹی آف میڈیکل سائنسس میں اساتذہ کی توانائیوں کا جائزہ لینے کے لئے طلباء کا تیار کردہ سوالنامے کے علمی معیار کا جائزہ

**بیک گراؤنڈ:** طلباء کی نظر سے اساتذہ کی علمی توانائیوں کا جائزہ لینا، اساتذہ کی علمی صلاحیتوں کو جانچنے کا ایک اہم طریقہ ہے۔ اس طریقے کا مقصد اساتذہ کے مثبت پہلوؤں کی تقویت اور نقائص کو دور کرنا ہے۔ دوہزار اٹھارہ میں مشهد یونیورسٹی آف میڈیکل سائنسس میں ہیلتھ فیکلٹی نے اساتذہ کی علمی توانائیوں کی جانچ کرنے کے لئے سوالنامہ مرتب کیا تھا۔

**روش:** اس تحقیق میں گذشتہ برس کے پہلے سال کے طلباء کو شامل کیا گیا۔ انہیں اساتذہ کی صلاحیتوں کے تعلق سے سوالنامہ دیا گیا۔ سوالنامے میں بارہ سوال تھے جنہیں لائیکرٹ اسکیل سے دس رینکوں اور چار اصلی طبی موضوعات پر مشتمل رکھا گیا تھا۔

**نتیجے:** سوالنامہ کے ذریعے حاصل ہونے والے اوسط نمبر دس میں سے آٹھ اعشاریہ سات سات تھے۔ اس سے یہ بھی نتیجہ حاصل ہوا کہ تین اصلی موضوعات کا نظام تعلیم بہتر ہے۔

**سفرش:** اساتذہ کی علمی توانائیوں کا جائزہ لینے کے لئے طلباء کا تیار کردہ فارم ناقص ثابت ہوا، اس پر نظر ثانی کی ضرورت ہے۔

**کلیدی الفاظ:** تعلیم، اصلی موضوعات، لائیکرٹ اسکیل، مشهد

## INTRODUCTION

One of the major goals of universities is to train specialist staffs. One of the important pillars of this process is university lecturers, who educate students and prepare them for the work market. The study of university professors' characteristics and merits is inevitable owing to their high duty. The evaluation system of faculty members at universities of medical sciences is more important, as faculty members and their graduates are responsible for humans' health and life (1, 2).

Evaluation can identify professors' strengths and weaknesses and leads the system from static to dynamic with high quality. In addition, it attempts to strengthen the educational system by providing appropriate strategies (3). The assessment of professors' educational environment is one of the most complex types of assessment (4). Universities generally use different methods to evaluate their professors; however, their ultimate goal is to improve the teaching method, enhance and promote effective activities, and eliminate ineffective and undesirable methods (5). Professors' evaluation is one of the most complex types of evaluation in academic environments, since the characteristics of the community to be evaluated are unique and their tasks are diverse. The process of evaluating professors can be influenced by various parameters such as executive responsibilities, characteristics of the lesson presented, conditions and facilities of the university and class, and economic and livelihood issues; therefore, considering each of these cases seems to be impossible in practice (6). Thus far, several methods, including evaluation by faculty colleagues, evaluation by students, evaluation by educational experts, and evaluation by checking the educational documents have been proposed to evaluate professors (7). The most common method used by Iranian universities is the assessment of faculties by their students.

Researchers have employed different parameters to evaluate faculties by students; some of these parameters are students'

educational level, level of difficulty or simplicity of the class, number of students in the class, students' personality, teachers' academic level, students' semester and their expected and actual grades, philosophy of education, teachers' demographic characteristics such as gender, age, educational level, background and teaching experience, as well as physical and social attractions (8-10).

Considering that the main group of activities of teaching staff is related to students; therefore, it is important to check the tools used in this section. Although many important parameters cannot be evaluated in practice, the low credibility of the measurement tool can be detrimental to the process. Based on the present knowledge, a standard form for student evaluation has not been designed so far, and universities and faculties have designed and used a separate form according to their needs.

The validity and reliability of the questionnaire used in the evaluation of professors are highly important to achieve credible results. Thus far, various methods have been proposed to assess the validity of the questionnaires, including content validity, construct validity, and criterion validity. In this study, we aim to investigate the validity of the EFSQ used at the School of Health of MUMS using confirmatory factor analysis (CFA).

## METHODS

The present research is a descriptive-analytic cross-sectional study. The statistical population consists of students at the School of Health of MUMS in the first semester of 2018-2019. All the students who completed the EFSQ were considered. The inclusion criteria were studying at the School of Health of MUMS in 2018, and the exclusion criteria were incomplete completion of the questionnaire. We consider the ethical issues such as confidentiality of information, and non-prejudice of researchers in interpreting the results of the study. The data-gathering tool in this study was the EFSQ at the School of Health of MUMS, which was completed by students on their portal. The EFSQ consists of 12 questions

**Table 1. The EFSQ at the School of Health of MUMS**

No.	Questions	Domain
1	The power of expression and the ability to communicate and convey content	Education
2	Teaching in accordance with the educational goals and time defined in the curriculum	Education
3	Correct and fair evaluation using modern approaches	Education
4	Observing ethical and professional principles in the educational environment	Ethic
5	Advice, guidance and help to solve learning problems	Ethic
6	The professor is known as a suitable professional model	Ethic
7	Regular and active presence in educational activities (theoretical and practical classes)	Discipline
8	Attention to students' attendance and observance of educational regulations	Discipline
9	Access to outside the classroom (in-person or non-attendance)	Discipline
10	Optimal use of time allocated to the classroom or practical training program with proper planning (the lecture time control, question and answer, teamwork, etc.)	Management
11	Providing conditions for students' active participation in the training process	Management
12	The power of student management and leadership in the educational environment	Management

with a Likert scale of 10 points. The range of scores is between 1 and 10 in which higher scores are a sign of the better quality of teacher training. The following table lists the questions in four sections, including education, discipline, ethics, and management.

This study used SPSS and AMOS software programs to analyze the data in descriptive and analytical formats. Also charts and tables were employed to describe the data and evaluate the structural validity by exploratory and confirmatory factor analysis. The normality of data was confirmed using skewness and kurtosis indices and Shapiro-Wilk test. The Mahalanobis distance was used to evaluate outlier data. The significant level was considered 5%.

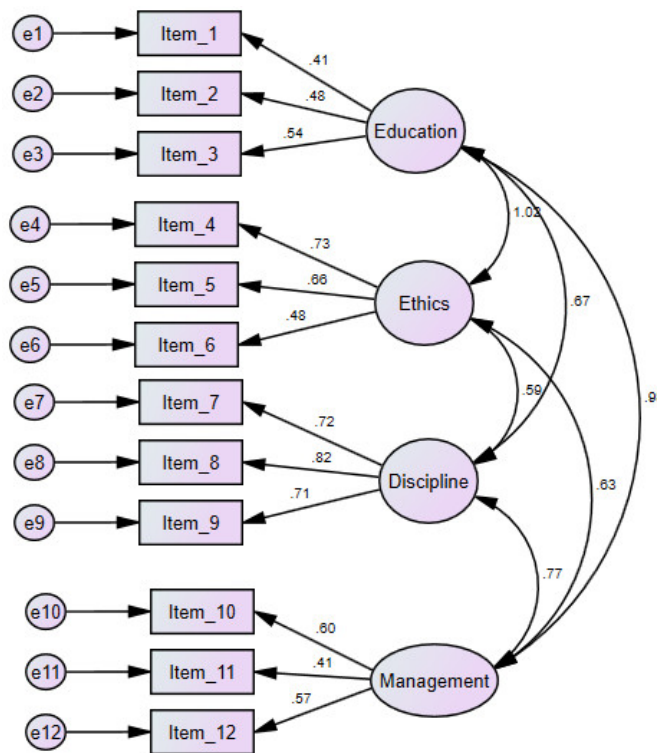
**RESULTS**

The rate of responding to the EFSQ was 92% (2044 out of 2219). The average total score of the EFSQ was  $8.77 \pm 1.30$  out of 10. Table 2 presents the descriptive statistics for the EFSQ dimensions.

Figure 1 which shows the CFA results using standard factor loading values and goodness of fit indices, indicates the weak fit of this model. Therefore, the current EFSQ used to evaluate faculty members is inappropriate and needs to be reviewed. Given that the minimum sample size required in a CFA is 10 to 20 cases per parameter, the sample size used in this study is sufficient. In addition, the adequacy of the sample size after fitting the CFA was also confirmed using the HOLTER test.

**Table 2. Descriptive statistics of scores in the EFSQ**

Dimension	N	Mean (SD)	(Minimum, Maximum)	Median (Interquartile Range)
Education	2044	8.79 (1.54)	(1 , 10)	9.33 (2.00)
Ethics	2044	8.84 (1.63)	(1 , 10)	9.33 (1.67)
Discipline	2044	8.77 (1.86)	(1 , 10)	9.67 (1.67)
Management	2044	8.70 (1.66)	(1 , 10)	9.33 (2.00)
Total	2044	8.77 (1.30)	(1.92 , 10)	9.21 (1.50)



Chi square=2095.656; df=48; p<0.001  
 RMSEA=.145; CFI=.738; TLI=.640

**Figure 1. Results of the CFA of the initial model on the entire data using standard factor loading values**

To achieve the proper structure, the data was divided into two parts. In the first part, to obtain the proper structure, exploratory factor analysis (EFA) was performed, and in the second part, the proposed structure obtained from the previous step was evaluated by using a CFA. The value of the KMO index (0.805) indicates the adequacy of the sample size for this analysis. In the EFA with Quartimax rotation, the model with three factors was evaluated as the best model. Table 3 presents the factor loading values.

Figure 2 shows the results of the CFA of the proposed model on 50% of the data (Test Data), indicating the good fit of this model. Table 4 presents the goodness of fit indices for the CFA, indicating that the model is proper.

The values of various internal consistency indices, including Cronbach's alpha and Composite Reliability, were almost equal to and greater than 0.7, confirming the internal consistency of the model.

**DISCUSSION**

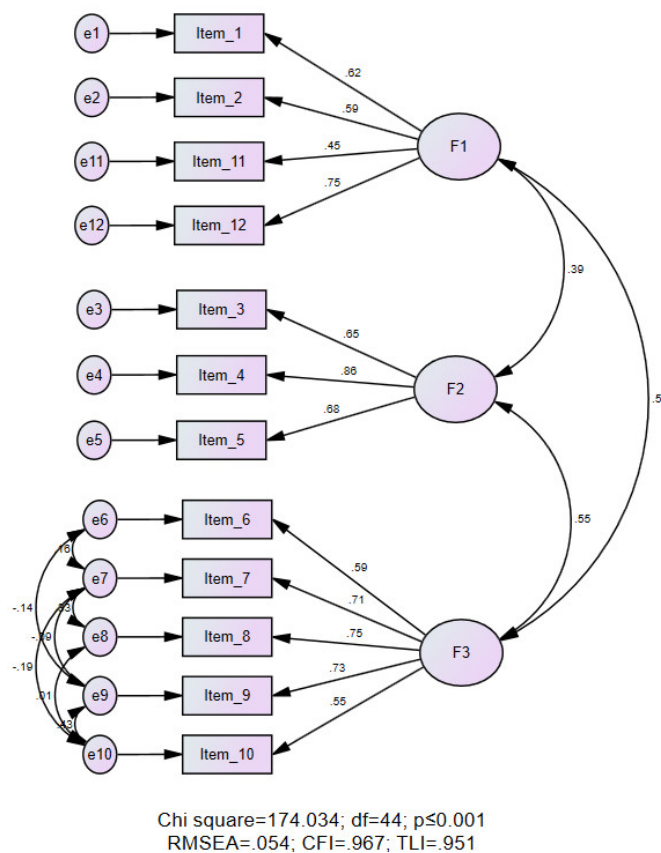
One of the most controversial issues in the teaching process is the evaluation of the educational system, particularly the evaluation of professors (12). Undoubtedly, the results of accurate evaluation illustrating the strengths and weaknesses provide a basis for decision-making on educational programs and arrange a framework to improve the quality of university education. From the students' viewpoint, one of the most important methods to evaluate the

faculty members to show the professors' performance among the most important community of their target is students.

**Table 3. EFA results on 50% of the data (Train Data)**

	Rotated Component Matrix <sup>a</sup>		
	Component		
	1	2	3
Item 1		.760	
Item 2		.584	
Item 3			.789
Item 4			.831
Item 5			.703
Item 6	.599		
Item 7	.748		
Item 8	.832		
Item 9	.792		
Item 10	.676		
Item 11		.663	
Item 12		.787	

Extraction Method: Principal Component Analysis.  
Rotation Method: Quartimax with Kaiser Normalization.<sup>a</sup>



**Figure 2. The results of the CFA of the proposed model on 50% of the data (Test Data) using standard factor loading values.**

**Table 4. Goodness of fit indices for the CFA on 50% of the data (Test Data)**

Measure	Estimate	Threshold (11)	Interpretation
Chi-Square ( $\chi^2$ )	174.034	--	--
DF	44.000	--	--
$\chi^2/DF$	3.955	Between 1 and 3	Questionable
CFI	0.967	>0.95	Excellent
SRMR	0.047	<0.08	Excellent
RMSEA	0.054	<0.06	Excellent
PClose	0.190	>0.05	Excellent

**Table 5. Indices of internal consistency assessment on 50% of the data (Test Data)**

	Cronbach's Alpha	Composite Reliability
F1	0.683	0.696
F2	0.765	0.778
F3	0.814	0.803

Although the aim of the evaluation is to promote the education, as well as growth and development of the faculty members; however, in many universities, it is merely a criterion for promotion of professors (3, 13). The aim of this study was to determine the proper structure of the EFSQ at the School of Health of MUMS. The results of this study indicated that the questionnaire with four dimensions was not suitable, and the most appropriate structure for these questions was the form with three dimensions. Based on the CFA and EFA results, the questions 1, 2, 11 and 12 determined the education scope.

Since questions 3, 4, and 5 were in one scope, the title of commitment was proposed for this dimension. Although the set of questions 6 to 10 was in one area, it seemed that the sixth question was ambiguous. Regarding the lack of a precise definition for the statement "professional style", this study suggested the removal or modification of this question. Furthermore, the present researchers thought that the title of discipline was appropriate for this domain. Previous studies have indicated that in addition to a teacher's academic and practical skills, the attitudinal, moral, as well as personality traits are also factors which introduce a teacher as an appropriate professional pattern (14). In addition, given that the eleventh question had the lowest standard factor loading (less than 0.5), it was suggested that this question be removed and replaced with another appropriate question.

The third question of the EFSQ was concerning the correct and fair assessment by the professors; however, it was noticeable that the evaluation of students by teachers has been generally conducted at the end of the semester, and the EFSQ was completed before the final exam; therefore, the students' answer to this question was doubtful; so the deletion of this question was suggested (15).

However, some studies have reported that the EFSQ scores

are associated with students' expected score so that if students expect a high score, they will give a high score to the teacher and vice versa (16).

In one study, Asgharpour et al. examined the factors affecting the student's evaluation of the professor from faculty members and students' perspective at the Faculty of Nursing and Midwifery of Tehran University of Medical Sciences. In this study, professors and students agreed and approved that 18 qualities in the field of individual-personality traits of the teacher and 32 items relating to evaluation of the professor were high value items (17). Comparing the questions of the present study with the above high value items shows that many of them are collectively referred to one question; for example, Question 2 of this study "Teaching in accordance with educational goals and time defined in curriculum" covers many issues; however, other issues such as the use of visual and audiovisual equipment and attention to students' understanding determined in the study of Ali Asgharpour et al., were not available in the evaluation form of the School of Health of MUMS.

The results of this study demonstrated that some students used the same pattern to answer questions, which could be due to lack of motivation to participate in the evaluation process. This result is consistent with the results of studies that do not support teachers' evaluation by students. These studies stated some reasons for not approving this evaluation method. Some of these reasons are lack of justification of the students before completing the form, lack of responsibility of students, lack of feedback to the professors and consequently, their impact on the teaching method, as well as lack of influence on promoting the professors (8, 15, 18). Voluntary participation, considering incentive privileges, and qualitative evaluations can be a step toward solving this problem.

The EFVS is a 10-option Likert spectrum that can cause confusion and biased results. Previous studies have shown that the five-choice questions in Iran are the most appropriate ones (19); however, the present study suggests that the number of options be reduced to five.

Student assessment scores have been used in many studies to examine the effectiveness of education, which is based on the belief that students learn more from professors with higher grades (20, 21). However, a meta-analysis study indicated that student grades should not be considered a

factor associated with the effectiveness of professors' teachings (22).

Based on the factor analysis results, the instrument validity was not approved. It is necessary to modify the evaluation process, for example, by voluntary participation in the evaluation. In general, due to mismatch of the EFSQ at the School of Health of MUMS with the EFSQ in other universities, the present researchers recommend that the questions be revised and the content validity be evaluated through the panel of experts.

#### Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double

publication and/or submission, redundancy, etc.) have been completely observed by the authors.

#### ACKNOWLEDGEMENT

We thank all the students who participated in this study for generously helping us in this research. The authors would like to thank all the members of the Educational Development Office (EDO) for their valuable and constructive comments.

**Financial Support:** The present study was financially supported by the Research Council of Mashhad University of Medical Sciences with code number: 971908.

**Conflict of interest:** None declared.

#### REFERENCES

- Sharifinia H, Ebadi A, Hekmat Afshar M. The characteristics of good teacher based on view points of students and teachers: A qualitative study. *J Health Breeze*. 2014;2(1):1-10. Persian.
- Kamali F, Yamani N, Changiz T, Zoubin F. Factors influencing the results of faculty evaluation in Isfahan University of Medical Sciences. *J Educ Health Promot*. 2018;7:13.
- Zabbah I, Mirzadeh S, Jafari S. Improving Teacher Evaluation using Fuzzy Logic. *Research in Curriculum Planning* 2018;15(30):94-113. Persian.
- Causeman R, Hermen J. Strategic planning in educational system: reevaluating, reconstructing the structures, regenerating. [F. Mashayekh, A. Bazargan Harandi, trans]. Tehran: Madreseh. 2005. Persian.
- Bastani P, Amini M, Tahernejad A, Rouhollahi N. The Tehran University of Medical Sciences Faculty Members' Viewpoints about the Teachers' Evaluation System: A Qualitative Study. *J Torbat Heydariyeh Univ Med Sci*. 2014;2(1):7-16. Persian.
- Theall M, editor Valid faculty evaluation data: Are there any? An interactive symposium exploring issues in evaluation and student ratings. 85th annual meeting of the American Educational Research Association, Montreal; 2005.
- Marquis B, Hoston C. Leadership roles and management functions in nursing: theory and application. 7th ed. Philadelphia: Lippincott Williams & Wilkins; 2011.
- Aghasi N, SHahsavari F. Academic staff and student's opinion about evaluation of academic staff by students in Tehran Dental Branch, Islamic Azad University. *J Res Dent Sci*. 2018;14(4):237-43. Persian.
- Yunker PJ, Yunker JA. An investigation into the validity of student evaluations of teaching in accounting education. *Accounting Educators' Journal* 2000;12:1-16.
- Dargahi H, Mohammadzadeh N. Faculty members' evaluation by students: Valid or invalid. *Iranian Journal of Medical Education* 2013;13(1):39-48. Persian.
- Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 1999;6(1):1-55.
- Gien LT. Evaluation of faculty teaching effectiveness--toward accountability in education. *The Journal of nursing education* 1991;30(2):92-94.
- Mousavi S, Mouodi S, Kheirkhah F, Azadfrouz S, Sefidchian A, Bijani A. Evaluation of faculty members in the psychiatry group of Babol University of Medical Sciences: Viewpoints of internship medical students. *Biannual Journal of Medical Education Development Center (edc) Babol University of Medical Sciences* 2014;2(2):37-42. Persian.
- Heidari AA, Khooei A, Dadgarmoghadam M. Content analysis of educational assistants views regarding the evaluation of Mashhad University of Medical Sciences' professors in educational clinical departments: a qualitative study. *Future Med Educ J*. 2018;8(4):3-7.
- Shakurnia A, Torabpour M, Elhampour H. Correlation between student evaluation of teaching and students' grades. *Iranian journal of Medical Education*. 2006;6(1):51-58. Persian.
- Millea M, Grimes PW. Grade expectations and student evaluation of teaching. *College Student Journal*. 2002;36(4):582-91.
- Aliasgharpour M, Monjamed Z, Bahrani N. Factors affecting students' evaluation of teachers: Comparing viewpoints of teachers and students. *Iranian Journal of Medical Education* 2010;10(2):186-95. Persian.
- Greenwood GE, Bridges Jr CM, Ware WB, McLean JE. Student evaluation of college teaching behaviors instrument: a factor analysis. *The Journal of Higher Education* 1973;44(8):596-604.
- Jamali J, Ayatollahi SMT, Jafari P. The effect of small sample size on measurement equivalence of psychometric questionnaires in MIMIC model: A simulation study. *BioMed Research International*. 2017; Article ID 7596101.
- Fenderson BA, Damjanov I, Robeson MR, Rubin E. Relationship of students' perceptions of faculty to scholastic achievement: are popular instructors better educators? *Hum pathol*. 1997;28(5):522-25.
- Cohen PA. Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Rev Educ Res*. 1981;51(3):281-309.
- Uttl B, White CA, Gonzalez DW. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Stud Edu Eval*. 2017;54:22-42.