ORIGINAL ARTICLE

Reliability and validity of the faculty evaluation instrument used at King Saud bin Abdulaziz University for Health Sciences: Results from the Haematology Course

Fahad Al-Eidan,[1] Lubna Ansari Baig,[2] Mohi-Eldin Magzoub,[3] Aamir Omair[4]

Abstract

Objectives: To assess reliability and validity of evaluation tool using Haematology course as an example.

Methods: The cross-sectional study was conducted at King Saud Bin Abdul Aziz University of Health Sciences, Riyadh, Saudi Arabia, in 2012, while data analysis was completed in 2013. The 27-item block evaluation instrument was developed by a multidisciplinary faculty after a comprehensive literature review. Validity of the questionnaire was confirmed using principal component analysis with varimax rotation and Kaiser normalisation. Identified factors were combined to get the internal consistency reliability of each factor. Student's t-test was used to compare mean ratings between male and female students for the faculty and block evaluation.

Results: Of the 116 subjects in the study, 80(69%) were males and 36(31%) were females. Reliability of the questionnaire was Cronbach's alpha 0.91. Factor analysis yielded a logically coherent 7 factor solution that explained 75% of the variation in the data. The factors were group dynamics in problem-based learning (alpha0.92), block administration (alpha 0.89), quality of objective structured clinical examination (alpha 0.86), block coordination (alpha 0.81), structure of problem-based learning (alpha 0.84), quality of written exam (alpha 0.91), and difficulty of exams (alpha0.41). Female students' opinion on depth of analysis and critical thinking was significantly higher than that of the males (p=0.03).

Conclusion: The faculty evaluation tool used was found to be reliable, but its validity, as assessed through factor analysis, has to be interpreted with caution as the responders were less than the minimum required for factor analysis.

Keywords: Haematology course, Reliability and validity, Saudi Arabia. (JPMA 66: 453; 2016)

Introduction

Evaluation is an approach used to measure the quality of effectiveness of a programme and it is an essential part of medical education process.[1,2] Medical teaching requires evaluation as a part of their quality assurance and improvement procedures which provides evidence whether the teaching standards are being improved and how well course objectives are being achieved.[3] Evaluation should be multi-dimensional, involving subjective and objective data to gather comprehensive qualitative and quantitative information on teaching processes and learning outcomes.[3,4] Hence, there is a need for developing valid and reliable instruments for course evaluations.

Student evaluations of teaching (SET) generally using Lickert-type scales are the most commonly used methods of evaluation of teaching in higher education.[5] Despite questions regarding students competency in evaluating faculty, it is generally agreed that only students are in a position to provide faculty and course evaluation.[6-8] Using this approach to evaluate quality of course as a whole can be misleading as student ratings might be biased by the initial interest of students,[9] instructor reputation,[10] and instructor enthusiasm.[11,12]

The knowledge and ability of the supervisors in conducting the course has been identified as an important factor that affects the scores of the students in that course.[13,14] Harris et al. identified curricular design, administrative skills of the supervisors, learning resources and environment as important factors for success of curriculum.[15] Srinivasan et al. in 2011 identified six core competencies for medical educators through a systematic review which are: medical knowledge, learner centeredness, interpersonal and communication skills, professionalism and role modelling, practice-based reflection and improvement, and system-based learning.[16]

A systematic review by Beckman et al. on the reliability and validity of instruments used for clinical teaching found that majority of the instruments used internal structure for validating the instruments.[17] The most frequently used domains included evaluation of clinical teaching and interpersonal skills of tutor, with the least being motivation, delegation, punctuality and availability.[17] The Boerboom et al. study from the

[1]Department of Pharmacology, [3,4]Department of Medical Education, King Saud Bin Abdul Aziz University of Health Sciences, Riyadh, Saudi Arabia, [2]APPNA Institute of Public Health, Jinnah Sindh Medical University, Karachi Pakistan.
Correspondence: Lubna Baig. Email: Lubna.shakil1983@gmail.com

veterinary school in Netherlands got a five factor solution for the Mastricht Clinical Teaching Questionnaire (MSTQ). All the five factors had reliability ranging from 0.87 to 0.96 and included general learning climate (GLC), modelling, coaching, articulation and exploration.[18] Kirschling et al. validation of teaching effectiveness tool for nurses also got a five factor solution with some similarities with the Boerboom et al. study and included: knowledge and expertise, facilitative teaching, communication style, use of own experience and feedback.[18,19] Broomfield and Bligh validated the Course Evaluation Questionnaire (CEQ) for undergraduate medical education and found good reliability for the 6 factors, out of which five factors were validated by Steele et al.[20,21]

The College of Medicine, King Saud bin Abdul Aziz University of Health Sciences (KSAU-HS) is constantly trying to improve the quality of courses and relying heavily on course evaluation instruments. There has been no study so far in the Kingdom of Saudi Arabia that assessed the reliability and validity of the course evaluation instruments. The current study was planned to assess reliability and validity of the course evaluation tool using the Haematology course as an example.

Subjects and Methods

The cross-sectional study was conducted at KSAU-HS, Riyadh, Saudi Arabia, in 2012, while data analysis was completed in 2013. The College of Medicine in KSAU-HS has a four-year problem-based learning (PBL) programme adapted from the University of Sydney, Australia. The programme is delivered as organ-system based courses (referred to as blocks) in the first two years (called stage 1 and 2) and then in discipline-based courses/blocks (called stage 3) inclusive of clerkship in Medicine, Surgery, Paediatrics, Gynaecology and Obstetrics, and Family Medicine in the latter two years. The block is managed through a preceptor who is called the coordinator. The responsibilities of the block coordinator include organisation and smooth functioning of the block, assigning tutors and attachments for the students, developing assessment tools with other faculty members in the block, and dealing with students' learning issues during the block. This study was done with the haematology block that is offered in the first year over six weeks. The first five weeks of the block are for teaching and the last week is the exam week. The teaching is organised around the problem used for PBL sessions in that week.

After approval from the institutional review committee, the 27-item block evaluation instrument was developed for the study to assess various components of the block, including organization (5 items), duration (1 item), quality of problems (9 items), performance of block coordinator (3 items), quality of written exams (3 items), and quality of objective structured clinical examination (OSCE) (6 items). The questionnaire was developed after an in-depth review and validation by faculty and in the light of block evaluation instruments available in the literature.[17-21] Typically, block evaluation is conducted at the end of each week through a structured questionnaire on a five-point Likert scale (5 = excellent, 4 = very good, 3 = good, 2 = fair, 1 = poor). The results of all the weeks are then aggregated to get a comprehensive view of the entire block.

The study comprised male and female students of first year in the medical school who went through the haematology block. The internal consistency reliability of the instrument was assessed through Cronbach's alpha. To assess the validity of the questionnaire, factor analysis was done using principal component analysis with varimax rotation and Kaiser normalisation. Loadings below 0.3 were suppressed and Eigen value was kept at 1. The items loading on to the respective factors were combined to get the internal consistency reliability of each factor. To compare the significant differences between males and females, students t-test was used to compare mean ratings for the faculty and block evaluation. Levine's test for equality of variances was done to ensure that the appropriate statistical test was applied. Levine's test was only significant for quality of block and, hence, t-test, which does not assume that variances are equal, was used.

Results

Of the 116 subjects in the study, 80(69%) were males and 36(31%) were females. The reliability of the questionnaire was Cronbach's Alpha 0.91. Factor analysis yielded a 7 factor solution that explained 75% of the variation in the data. The seven factors, which were logically coherent and matched with the items on the faculty evaluation tool, were: group dynamics in PBL (alpha 0.92), block administration (alpha 0.89), quality of OSCE (alpha 0.86), block coordination (alpha 0.81), structure of PBL (alpha 0.84), quality of written exam (alpha 0.91), and difficulty of exams (alpha 0.41). There were some double loadings which were logically connected, like sequence of activities loaded with block administration and coordination. There was one isolated loading of block duration with difficulty of exam which we could not explain and therefore was not included with any factor for calculating reliability (Table-1).

There was a significant difference between the perception of knowledge gained in the first week, "Always Tired" ($p=0.03$) and in the third week "A Swollen Knee" ($p=0.007$) (Table 2). There were no significant differences between males and females in their perception of the relevance and

Table-1: Factor analysis of the faculty evaluation tool.

	Group dynamics in PBL	Block administration	Quality of OSCE	Block coordination	Structure of PBL	Quality of written exam	Difficulty of exams
1.	Amount Learned				0.766		0.305
2.	Relevance to KSA for the PBL					0.771	
3.	Stimulating Problem					0.743	
4.	Group Rapport, Cooperation	0.776					
5.	Depth of analysis, Critical thinking	0.772					
6.	Balance of participation	0.773	0.386				
7.	Division of work	0.861					
8.	Function of chairman	0.841					
9.	Function of secretaries	0.761				0.315	
10.	Block Duration						0.814
11.	Quality of the block book content		0.883				
12.	Quality of the block clarity		0.874				
13.	Reasonable sequence of activities within the block		0.764		0.340		
14.	Schedule maintenance: lectures stayed on schedule	0.366	0.487		0.475		0.404
15.	Schedule maintenance skills demonstration	0.449			0.573		0.403
16.	Accessibility of the block coordinator			0.356	0.725		
17.	Block coordinators helpfulness				0.580		
18.	Attention of the block coordinator				0.841		
19.	Comprehensive coverage of content						0.862
20.	Quality of items in written exam						0.855
21.	Difficulty of written exam	0.308	-0.544		0.354		0.376
22.	Comprehensive coverage of content in OSCE			0.741			
23.	Quality of station's set up			0.903			
24.	Flow between stations			0.824			
25.	Quality of station monitors			0.818			
26.	Quality of standardized patients		0.461	0.459			
27.	Level of OSCE difficulty						0.600

KSA: Kingdom of Saudi Arabia
PBL: Problem-based learning
OSCE: Objective Structured Clinical Examination

Table-2: Comparison between perception of male and female students for the amount of knowledge gained in different weeks[a].

Problems	Amount learned			Relevance to KSA			Problems were Stimulating		
	Male (n=80)	Female (n=36)	p-value	Male (n=80)	Female (n=36)	p-value	Male (n=80)	Female (n=36)	p-value
Always Tired	4.2 ±1.0	4.3 ±0.7	0.76	4.3 ±1.0	3.9 ±1.3	0.14	4.1 ±1.1	4.0 ±0.9	0.69
While I'm Here	4.0 ±1.1	4.3 ±0.9	0.27	4.1 ±1.1	3.7 ±1.3	0.18	4.0 ±1.1	3.7 ±1.3	0.28
A Swollen Knee	4.2 ±1.0	4.6 ±0.6	0.007*	4.1 ±1.1	3.9 ±1.3	0.44	4.1 ±1.1	4.2 ±0.9	0.49
Pale and Feverish	4.3 ±1.0	4.4 ±0.7	0.66	4.1 ±1.2	3.8 ±1.3	0.19	4.3 ±1.0	4.1 ±1.0	0.40
Hind's Painful Calf	4.4 ±0.8	4.5 ±0. 7	0.66	4.2 ±1.1	3.7 ±1.3	0.04*	4.1 ±1.1	4.1 ±1.0	0.85

KSA: Kingdom of Saudi Arabia
[a]The mean and standard deviation values are rounded off to one decimal place
*Statistically significant at p<0.05.

characteristics (problem was stimulating) of the problem used for PBL sessions (p>0.05).There was no statistically significant difference in the opinion of male and female students about the quality of problems used for PBL Sessions (p>0.05).

Male students gave a significantly high rating to block book's clarity (p<0.001) and content (p=0.007) (Table-3), and thought that block coordinator was more helpful (p<0.001) (Table-4) compared to females. Male students' gave a significantly high rating (3.8±1.1) to the quality of

Table-3: Students perception of the Quality of Problem-Based Learning Sessions[a].

Opinion of PBL sessions	Male (n=80)	Female (n=36)	p-value
Group Rapport, Cooperation	4.2 ±1.1	4.1 ±1.0	0.88
Depth of analysis, critical thinking	3.9 ±1.1	4.0 ±0.9	0.52
Balance of participation	4.0 ±1.1	3.6 ±1.2	0.056
Division of Work	4.0 ±1.1	3.7 ±1.2	0.12
Function of Chairman	4.2 ±0.1	4.3 ±0.8	0.52
Function of Secretary	4.1 ±1.0	4.3 ±0.8	0.48

PBL: Problem-based learning

[a]The mean and standard deviation values are rounded off to one decimal place

* Statistically significant at p<0.05.

items in the written exam compared to female students' (3.3 ±1.3) (p=0.04). There were no significant differences among male and female students for the quality and difficulty of examination (Table-5).

Discussion

The key finding of our study was that the faculty evaluation tool was a reliable and valid instrument. Also, there were no major differences among male and female students although the instructors and block coordinators were different for the two groups.

Most studies in literature have validated the teaching instruments and found five to seven factor solutions with 1-43 items on the quationnaires.[17-19] The seven-factor solution yielded by the present study, had strong loadings from the 27 items on the questionnaire which is an empirical evidence of its reliability. The six factors had high reliability ranging from 0.81 to0.92 with the exception of one, "difficulty of exams", which could be due to the fact that there were only two variables that did not load on to any other factor. The method and process of adducing evidence of reliability that we used for the data is similar to the studies that have validated their teacher evaluation instruments.[17-19] The items on the questionnaire used by two other studies[20,21] assessed students' perception of learning from the course whereas our instrument assessed students' perception of block organisation and management.

There were generally no statistically significant differences between the perceptions of males and females on any of the factors evaluating the course of study/block, which is

Table-4: Students perception about the Administration and Coordination of the Block[a].

Block Administration	Male (n=80)	Female (n=36)	p-value
Quality of the Block Book - Content	4.4 ±0.7	3.9 ±1.2	0.003*
Quality of the Block Book - Clarity	4.5 ±0.7	3.7 ±1.3	0.001*
Reasonable Sequence of the Activities within the Block	4.5 ±0.8	3.9 ±1.0	0.003*
Schedule Maintenance: Lecture stayed on the schedule	4.5 ±0.8	4.0 ±0.9	0.006*
Block Duration			
Appropriate length for the Block	4.0 ±1.4	3.5 ±1.3	0.047*
Block Coordination			
Attention of the Block Coordinator to emerging Block needs	4.7 ±0.9	4.5 ±0.7	0.30
Schedule Maintenance: Skills Presentation (Demonstration) on the schedule	4.4 ±0.8	4.2 ±1.0	0.22
Helpfulness of the block Coordinator	4.8 ±0.5	4.5 ±0.7	0.03*
Accessibility of the Block Coordinator	4.7 ±0.6	4.6 ±0.6	0.20

[a]The mean and standard deviation values are rounded off to one decimal place

* Statistically significant at p<0.05

Table-5: Perception of the students about the Difficulty and Quality of Written and Objective Structured clinical Examination[a].

	Male (n=80)	Female (n=36)	p-value
Written			
Comprehensive Coverage of content	3.8 ±1.2	3.7 ±1.1	0.79
Quality of Items	3.8 ±1.1	3.3 ±1.3	0.04*
OSCE			
Comprehensive coverage of content	4.0 ±1.1	4.0 ±0.9	0.89
Quality of stations' set up	3.9 1.0	3.9 ±0.9	0.96
Flow between station	4.1 ±1.0	4.1 ±0.8	0.75
Quality of station monitor	4.0 ±1.0	3.9 ±1.2	0.49
Quality of standardized patients	3.7 ±1.3	3.6 ±1.2	0.61
Difficulty in Exams			
Level of OSCE Difficulty	3.7 ±0.8	3.5 ±0.7	0.16
Level of Difficulty	3.9 ±0.8	3.5 ±1.2	0.07

[a]The mean and standard deviation values are rounded off to one decimal place

* Statistically significant at p<0.05.

not in concordance with the Dundee Ready Educational Environment Measure (DREEM) inventory that was used in the Kingdom and found a positive inclination of females towards the learning environment.[22] The same inventory when used in Sweden showed no difference in the perception of males and females.[23] Although these studies assessed the learning environment, but we stress that this is the closest comparison at this time of the instrument that we used for block evaluation. The differences in the opinion/perceptions of male and female students are similar to what is found internationally even though the classes for male and female students are held separately in Saudi Arabia and more than 90% of the times they are taught by different instructors, and even the buildings are separate. Although our data is from one cohort of students from one block/course of study, but we can be confident in saying that both male and female students had similar opinion on the quality of the organisation and management of the Haematology Block.

In terms of limitations, factor analysis requires a minimum of 5 responses for each variable and the current study was short of 19 responses. Hence, the results have to be interpreted with caution as they may be unstable. Besides, the results of the study are preliminary as they are from one block of study, but this instrument can be used confidently for other blocks as well. As a continuation of this study we intend to use the data from other blocks to do a confirmatory factor analysis to adduce evidence of construct validity for this instrument.

Conclusions

The faculty evaluation tool was found to be reliable, but its construct validity has to be evaluated with caution as there were 13% less respondents than required. No major difference was found between male and female students about the perceptions of the quality of block organisation, except that male students thought that the block book was better, instructors were more helpful and the items on the exam were of high quality.

References

1. Morrison J. ABC of learning and teaching in medicine: Evaluation. BMJ. 2003; 326:385-7.
2. Cohen L, Manion L. Research methods in education. 4th ed. London: Routledge, 1994.
3. McOwen KS, Bellini LM, Morrison G, Shea JA. The development and implementation of a health-system-wide evaluation system for education activities: build it and they will come. Acad Med. 2009; 84:1352-9.
4. 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; 34:862-70.
5. McKeachie W. Student ratings; the validity of use. Am Psychol. 1997; 52: 1218-25.
6. Coffey M, Gibbs G. The evaluation of the student evaluation of educational quality (SEEQ) questionnaire in UK higher education. Ass Eval High Edu. 2001; 26:89-93.
7. Cohen PA, McKeachie WJ. The role of colleagues in the evaluation of teaching. Improving college and university teaching. Teache Evaluation second edition ed - Kenneth D. Peterson. Corwin Press, Inc. California 2000; pp 147-54.
8. Kember D, Leung D, Kwan K. Does the use of student feedback questionnaires improve the overall quality of teaching? Assess Eval Higher Educ. 2002; 27:411-25.
9. Prave RS, Baril GL. Instructor ratings: Controlling for bias from initial student interest. J Educ Bus. 1993; 68:362-6.
10. Griffin BW. Instructor reputation and student ratings of instruction. Contemp Educ Psychol. 2001;26:534-52.
11. Naftulin DH, Ware JE, Donnelly FA. The Doctor Fox lecture: A paradigm of educational seduction. J Med Educ. 1973; 48:630-5.
12. Marsh HW, Ware JE. Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. J Educational Psychol. 1982; 74:126-34.
13. Gathright MM, Thrush C, Jarvis R, Hicks E, Cargile C, Clardy J, et al. Identifying areas for curricular program improvement based on perceptions of skills, competencies, and performance. Acad Psychiatry. 2009; 33:37-42.
14. Steinert Y. Mapping the teacher's role: The value of defining core competencies for teaching. Med Teach. 2009; 31:371-2.
15. Harris DL, Krause KC, Parish DC, Smith MU. Academic competencies for medical faculty. Fam Med. 2007; 39:343-50.
16. Srinivasan M, Li ST, Meyers FJ, Pratt DD, Collins JB, Braddock C, et al. "Teaching as a Competency": competencies for medical educators. Acad Med. 2011; 86:1211-20.
17. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. J Gen Intern Med. 2004; 19:971-7.
18. 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:e84-91.
19. Kirschling JM, Fields J, Imle M, Mowery M, Tanner CA, Perrin N, et al. Evaluating teaching effectiveness. J Nurs Educ. 1995; 34:401-10.
20. Broomfield D, Bligh J. An evaluation of the 'short form' course experience questionnaire with medical students. Med Educ. 1998; 32:367-9.
21. Steele G, West S, Simeon D. Using a modified course experience questionnaire (CEQ) to evaluate the innovative teaching of medical communication skills. Educ Health (Abingdon). 2003; 16:133-44.
22. Mojaddidi MA, Khoshhal KI, Habib F, Shalaby S, El-Bab ME, Al-Zalabani AH. Reassessment of the undergraduate educational environment in College of Medicine, Taibah University, Almadinah Almunawwarah, Saudi Arabia. Med Teach. 2013; 35:S39-46.
23. Edgren G, Haffling AC, Jakobsson U, McAleer S, Danielsen N. Comparing the educational environment (as measured by DREEM) at two different stages of curriculum reform. Med Teach. 2010; 32:e233-8.