Education in Medicine Journal

ORIGINAL ARTICLE

Title: The Development of 360-Degree Evaluation Model on Medical Curriculum with Kirkpatrick Hierarchy Approach

Authors: Mia Kusmiati, Suhaila Sanip, Rafidah Bahari

Submitted Date: 31-08-2022

Accepted Date: 07-09-2023

To cite this article: Kusmiati M, Sanip S, Bahari R. The development of 360-degree evaluation model on medical curriculum with Kirkpatrick hierarchy approach. Education in Medicine Journal. (early view).

ORIGINAL ARTICLE

ARTICLE INFO

Submitted: 31-08-2022
Accepted: 07-09-2023

The Development of 360-Degree Evaluation Model on Medical Curriculum with Kirkpatrick Hierarchy Approach

Mia Kusmiati[1], Suhaila Sanip[2], Rafidah Bahari[3]

[1]Department of Medical Education, Bioethics and Humanity, Faculty of Medicine, Bandung Islamic University, INDONESIA
[2]Department of Medical Education, Faculty of Medicine, University Sains Islam Malaysia, Kualalumpur, MALAYSIA
[3]Department of Psychiatry, Faculty of Medicine, University of Cyberjaya, Selangor, MALAYSIA

To cite this article: Kusmiati M, Sanip S, Bahari R. The development of 360-degree evaluation model on medical curriculum with Kirkpatrick hierarchy approach. Education in Medicine Journal. (early view).

ABSTRACT

A 360-degree evaluation model of the medical curriculum was developed by adhering to the Kirkpatrick Hierarchy. In order to bridge the research gap pertaining to instrument development of 360-degree evaluation, a set of tools derived from various stakeholders' perspectives was used. This mixed-method study in the sequential exploratory design involved 797 participants for two years, 2017-2019. The three study phases involved were: the qualitative phase (18 informants), the item construction phase with exploratory factor analysis (EFA) (298 participants), and the validation phase via confirmatory factor analysis (CFA) (481 participants by random sampling). In-depth interview sessions were conducted with two stakeholder groups; lecturers and preceptors. Focus group discussion (FGD) sessions were conducted with two other groups; medical students and patients at the teaching hospital. The item construction phase was executed based on the themes that emerged from qualitative findings. Lastly, the final validation stage was performed based on the EFA results. In total, 23 themes were derived from four stakeholder groups. In the item construction phase, 13 and 10 factors were identified for lecturer and student instruments, respectively. To build the preceptor and patient tools, 10 scales were used to validate the item constructs. Thus, the 360-degree evaluation model has 4 levels of Kirkpatrick Hierarchy with 4 instrument models. The 360-degree evaluation model is valid, well-constructed, and accurately reflects the indicator variables. The evaluation model is feasible and acceptable to assess the medical curriculum.

Keywords: *Evaluation, Factor analysis, Model approach, Medical curriculum, Stakeholders*

CORRESPONDING

Mia Kusmiati, Department of Medical Education, Bioethics and Humanity, Faculty of Medicine, Bandung Islamic University, Bandung-40116, Indonesia

Email: dr.mia74@gmail.com

INTRODUCTION

Curriculum evaluation is designed to assess crucial curriculum elements, along with an emphasis on several aspects of the new curriculum (1). Several tools have been developed to evaluate, but limited only to a number of aspects embedded in the curriculum components. Medical education, since the last decade, has prescribed multisource feedback to examine the academic performance of medical students (2). In fact, the three approaches to curriculum evaluation are reductionist, systems theory, and complexity (3). The reductionist curriculum evaluation assesses the relationship between curriculum components and causal approach among educational inputs and outputs, while the complexity approach looks into the overall aspects in a holistic manner from various stakeholder perspectives. Curriculum evaluation of the system theory consists of components related to the interactions and interrelations among the elements, all existing within and interacting with the program setting. Complexity theory views medical curricula as complex systems, given that they are made up of diverse components with interactions among those components (3).

A number of institutions have implemented the reductionist method to assess one component of the curricula. Nonetheless, the assessment of all curriculum components from various academic stakeholders is in scarcity; which should be given due attention to ascertain social accountability for medical education. By building a multisource evaluation of the Kirkpatrick Hierarchy (4), all curriculum components ranging from teaching and assessment methods, learning outcomes, curriculum structures, content knowledge to educational program outcomes can be evaluated, which have been declared as professional behaviour and clinical competence of medical graduates (5). The evaluation of educational programs from multiple stakeholders (6) is crucial because a comprehensive evaluation is built from a stakeholder's understanding of the most effective strategy to yield the expected outcomes of education (3). Stakeholder perspective is, undeniably, an integral factor in identifying the developmental needs of the curriculum (6,7).

Some studies have evaluated merely a single fraction of the curriculum, such as the following evaluations: clinical teaching efficacy (8,9), musculoskeletal module in a university (10), professional exercise (11), and learning outcomes on general areas of expertise (1). Meanwhile, the evaluation of clinical competency was performed to assess effective teaching skills (12), intervention programs(13), clinical teachers' supervisory skills (14), learning environment (15), and medical professionalism (16).

The 360-degree evaluation model is one of the best assessment methods that can be deployed to examine professionalism and communication skills competencies (2). In line with the complexity theory, a 360-degree curriculum evaluation refers to a thorough evaluation of the curriculum components and the interactions between those components. The research gap reflects the fact that although many evaluation tools have been developed, they are still separate per curriculum components, such as the development of tools to assess simply learning reactions or educational outcomes only. Therefore, this present study outlines the holistic aspects of the four levels of Kirkpatrick's model. In order to bridge the research gap on the instrument development of 360-degree evaluation, a set of tools from various stakeholders' perspectives is required. As such, this study substantially contributes to the body of knowledge regarding curriculum evaluation methods, particularly for the development of multisource evaluation or multi-source feedback. Effective evaluation methods that determine the efficacy of educational programs are needed to enable various stakeholders to thoroughly overview the medical curriculum.

The Kirkpatrick Hierarchy refers to an evaluation approach that revolves around four aspects: reaction, learning, behaviour, and results. Level 1-reaction measures how one reacts to a program (students' satisfaction). Level 2-learning assesses if one truly understands the course (increase in

knowledge, skills or experience assessed by lecturer). Level 3- behaviour denotes aspects used from the learned result at the workplace (change in behaviour as a performance at the hospital), and level 4- result determines if learning is transferred into practice at the workplace (medical professional behaviour) (17). Hence, an evaluation model of medical curriculum is proposed in this study from four stakeholder groups by adhering to the Kirkpatrick Hierarchy. In this context, the evaluation that amalgamates with the Kirkpatrick model is representative of the 360-degree evaluation model.

METHODOLOGY

Study Context

The mixed-method approach based on sequential exploration design was deployed in this study. Three phases were executed to meet the study objective, namely: exploration phase via qualitative approach (18 informants), questionnaire items development (298 participants), and validation phase (481 participants). A total of 797 individuals participated in this study. The data analysis conducted during the qualitative phase adhered to the seven-step content analysis prescribed by Dahlgren (18). Next, item construction and instrument validation phases were executed by performing exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), respectively.

Setting

In the first phase, qualitative data were captured by identifying important themes and specific statements from four stakeholder groups; lecturers, medical students, preceptors, and patients at a teaching hospital in Indonesia. In-depth interview sessions were conducted by the researcher in a semi-structured manner with two stakeholder groups (lecturers & preceptors), whereas focus group discussion (FGD) sessions were conducted with medical students and patients. The semi-structured in-depth interviews lasted for 45-105 minutes with 4 lecturers and 4 preceptors. The FGD sessions lasted for 90-120 minutes with 6 medical students and 4 inpatients from the surgical department who were already recovering at the teaching hospital. All interview and FGD sessions were recorded using an audiotape recorder and transcribed verbatim after gaining consent from the participants.

In the second phase, the questionnaire items were determined based on the themes identified from the initial phase. The EFA was deployed to select factors for each evaluation instrument. The sample size for each stakeholder group was based on sample formulation to items ratio of 2-3 (19) for lecturers and preceptors, whereas sample formulation to factors ratio for students and patients. The ratios used for students and patients were 15:1 and 10:1, respectively (20). The sample size for each stakeholder group is given as: 74 lecturers (2 subjects for 37 items), 69 preceptors (3 subjects for 23 items), 105 students (15 subjects for 7 factors), and 50 patients (10 subjects for 5 factors). The total number of participants involved in the second phase was 298.

Besides qualitative findings, the questionnaire items were compiled from definitions found in the literature based on the desired domains (21,22). For example, the professional behaviour of internship students from the stances of patients and preceptors was adapted to refer to predefined terminology.

In the final phase, lecturers and patients were selected using the purposive sampling technique, while students and preceptors were selected using the simple random sampling method. The inclusion criteria of lecturers include full-time academic members of the UNISBA medical faculty, have been employed as a lecturer for at least 2 years, and meet the required academic qualifications. Meanwhile, the inclusion criteria of patients are as follows: recorded as a patient of inpatient and outpatient clinic

in departments of surgery and aged 21-60 years. The difference in the sampling methods used for patients, lecturers, and students with preceptors stemmed from the limited number of lecturers in the faculty who were eligible to be selected as research participants in this study.

The sample size and factors used in the ratio to execute CFA are as follows: 12 for lecturers (6 factors), 14 for preceptors (5 factors), 26 for students (7 factors), and 30 for patients (5 factors) (20). The following states the number of participants according to the categories: 73 preceptors, 73 lecturers, 182 students, and 153 patients. The number of participants involved in phase 3 was 481.

Statistical Methods

The qualitative phase (phase 1) deployed in this study integrated the seven steps of content analysis, which required the researcher herself to transcribe and analyse the outcomes captured from the semi-structured interview and FDG sessions, assisted by a research assistant. The seven steps consist of 1) familiarisation by reading all transcripts, 2) condensation by identifying meaning units for purposes of further scrutiny, 3) comparison by comparing units about similarities and differences, 4) grouping by capturing the essential meaning of each category, 5) articulating by expressing the core meaning of each category, and 6) labelling. Steps 3-6 were carried out in an iterative procedure to ensure that the similarities and differences within the categories were discerned and formulated distinctly. The final step (7) involved contrasting by comparing the categories in terms of similarities and differences.

All the participants were interviewed until data saturation was attained (no emergence of additional theme related to the study goals). The participants were recruited until a saturation point was achieved. This was estimated to range between 3 and 25 participants, as depicted in the literature (23,24). However, an empirical argument is absent for selecting that particular range and not any other ranges or numbers (23).

The data analysis of the second phase employed the reduction and extraction methods of the factors using the principal component analysis, principal axis factoring, and maximum likelihood with Kaiser Normalisation via SPSS Ver24 program. The loading factor value was grouped into a single factor. If the value of the loading factor exceeds 0.4 and 0.5 prior to the factor analysis for each instrument, both Kaiser Meyer Olkin (KMO) and Bartlett Sphericity tests were performed. If KMO value exceeds 0.5, the sample is considered adequate. Internal consistency was determined via Cronbach's alpha, whereby the observed variable data are considered consistent if Cronbach's alpha value exceeds 0.7 (25).

The data analysis for the third phase employed CFA with a structural equation modelling approach. The CFA is an important method to determine instrument validation by using LISREL 8.7 for the window program. The validity of the instrument was determined based on the t-value with the LISREL program. An item is considered valid if its t-value > 1.96, while the goodness of fit of a model is determined based on one of the following criteria: chi-square test (or $0 < Cmin / df < 3$, Root Mean Square Error of Approximation (RMSEA)) or comparative fit index (26,27). In this present study, the instrument model is considered good if the RMSEA value < 0.08 and $0 < Cmin / df < 3$.

Sample Size

Table 1 presents the number of participants involved in this study.

Table 1: Summary of participants' sample size

No.	Participant Type	Participants		
		Qualitative	EFA	CFA
1	Preceptor	4	69	73
2	Patient	4	50	153
3	Lecturer	4	74	73
4	Medical student	6	105	182
	Total	18	298	481

A total of 797 participants from 4 stakeholder groups were recruited. Subsequently, they were grouped into 3 cohorts. After getting the consent of each participant, they were asked to participate in each phase, namely the qualitative phase, the item construction phase, and the validation phase. Eight participants from two stakeholder groups (lecturer and preceptor) were interviewed for about 45-105 minutes. Two FGD sessions were conducted for 90-120 minutes in separate time and place, one session with 6 medical students in the university and another session with 4 patients in the teaching hospital. A total of 298 participants were invited to participate in the second phase, whereas the other 481 participants were required to complete the disseminated questionnaires in the third phase.

Ethical approval was granted by the research ethics committee from the Medical Faculty of Universitas Islam Bandung (UNISBA) No. 005/Ethic Committee FK/VI/2017.

RESULTS

Phase 1

In Phase 1, 23 themes were yielded from all stakeholder groups. Themes derived from lecturers and students were teaching-learning elements, assessment methods, and lecturer's teaching ability. Next, the patients and preceptors led to themes related to professional behaviour and clinical competence. Themes related to teaching-learning elements, assessment methods, and lecturer's teaching ability derived from students and lecturers represented level 1-reaction and level 2-learning in the Kirkpatrick Hierarchy. Seven factors that emerged from the students to assess level 1-reaction and six factors from lecturers in assessing level 2-learning are presented in Table 2.

Table 2: Similarity and differences factors amongst lecturer' and students' perspective

No.	Lecture's factor	Student's factor
1	Curriculum structure	Learning material & body of knowledge
2	Assessing learning outcome	Lecturer capacity
3	Lecturer competencies	Exam effectiveness
4	Content of the learning module	Learning experience
5	Assessment capacity	The difficulty level of exam
6	Learning method & material	Suitability of learning material with test
7	-	Integrated material

The subthemes retrieved from the teaching-learning process include tutorial method, clinical laboratory, and lecture method. As for lecturer competency, the subthemes obtained encompass pedagogic competency, guide and facilitator, and knowledge delivery (interactive lecture). The subthemes of curriculum structure embody the sequential of the block system, learning module content, and cases presented in the module. The subthemes obtained from the assessment method are type of examination, assessment capacity, and exam efficacy.

The five themes obtained from the preceptors regarding medical competencies and professional behaviour of medical doctors to construct level 3-evaluation are 1) professional conduct, 2) communication, 3) attitude towards patient and punctuality, 4) clinical skill competence, and 5) initiative. Next, the patients propounded five themes: altruism, communication-empathy, humanism, pleasant manner, and responsibility, for the professional behaviour of medical doctors and these correspond to level 4-evaluation based on the Kirkpatrick Hierarchy.

The subthemes of professional behaviour include the humanity aspect and the way of communication. As for responsibility-accountability, the subthemes are reliability in managing patients and experiencing medical doctors. The subthemes of clinical skill expertise encompass demonstrated performance, cognitive ability, and patient handling. The subtheme retrieved from professional conduct is demonstrated attitude, whereas the subtheme from attitude towards patient refers to positive interaction. Lastly, the subtheme of initiative is interaction with the patient and his family.

Level 4-result determines if learning is transferred into practice at the workplace. In this case, patient assessment of professional behaviour attributes is relevant to the work context in the hospital. Meanwhile, the evaluation by the preceptor about medical competence development and professionalism is categorised as level 3. This is because; the preceptor's assessment involved young doctors who were still in the learning process at the clinical stage.

Phase 2

The item constructs used to test the components of educational outcomes in terms of professional behaviour from the stances of patients and preceptors (28) were good and valid.

Item construction for lecturer and student tools involved 13 factors that met the psychometric property requirements with a sufficient number of participants in each group (74 lecturers & 105 students). The values of KMO and Bartlett tests for lecturer instrument were 0.845 and $p < 0.000$ ($X2 = 2070.29$), respectively. Meanwhile, the student instrument scored 0.705 and $p < 0.001$ ($X2 = 2584.12$) for KMO and Bartlett's test of sphericity, respectively. Seven factors were retained for student instrument as the

initial Eigenvalue exceeded 1 (see Figure 1) and accounted for 53.77% of the total variance after 9 iterations (see Table 2).

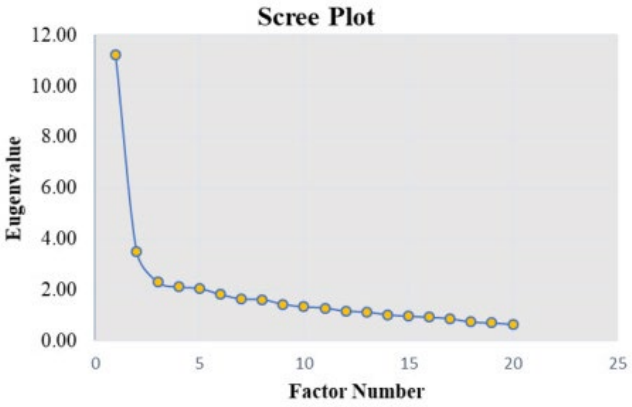

Figure 1: Factors in seven components that explained the variables in student instrument

Table 3: Structure matrix of maximum likelihood with varimax rotation (student instrument with 105 participants)

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	H2	Mean	SD
Q1		0.415						0.613	3.04	0.84
Q4				0.542				0.632	3.41	0.95
Q9				0.499				0.583	4.16	0.48
Q10		0.516						0.691	3.40	0.84
Q11	0.490	0.542						0.731	3.16	0.91
Q12	0.506					0.518		0.709	3.50	0.83
Q13	0.434					0.421		0.674	3.43	0.69
Q14	0.647							0.781	3.70	0.75
Q15	0.699							0.757	3.82	0.69
Q16	0.718							0.705	3.81	0.79
Q17	0.684							0.814	3.89	0.67
Q18	0.852							0.890	3.98	0.62
Q19	0.701			0.428				0.861	4.02	0.55
Q20	0.725							0.827	3.90	0.66
Q21		0.471						0.650	2.52	1.00
Q23	0.521							0.723	3.78	0.69
Q24	0.688							0.777	3.74	0.71
Q25			0.445			0.468		0.739	3.57	0.76
Q26			0.757					0.796	3.69	0.85
Q27			0.530					0.780	3.97	0.60

Q28			0.439					0.687	3.89	0.81
Q29			0.793					0.816	3.64	0.75
Q32					0.903			0.873	3.35	0.80
Q33					0.568			0.833	3.30	0.87
Q36						0.502		0.785	3.70	0.59
Q37						0.434		0.608	3.66	0.62
Q38				0.407				0.762	3.78	0.72
Q41		0.694						0.752	2.38	1.05
Q42		0.685						0.730	2.90	1.01
Q43		0.549						0.747	3.05	0.88
Q44							0.593	0.668	3.75	0.63
Q45							0.744	0.692	3.57	0.66
Q46				0.532		0.419		0.673	4.01	0.56

Notes: Factor 1: Lecturer ability and competency; Factor 2: Learning material and body of knowledge; Factor 3: Exam effectiveness; Factor 4: Learning experience; Factor 5: The difficulty level of the exam; Factor 6: Suitability of learning material with a test; Factor 7: Integrated material/module

Thirty-three items were retained as indicator variables (loading values > 0.4), while 13 items were discarded because their loading values were below 0.4 (20,29). The delete items are items q2-q3, q5-q8, q22, q30-q31, q34-q35, and q39-q40. The Cronbach's alpha values for all 33 items were 0.913; signifying very high consistency. All the retained items represented all indicator variables to assess both the teaching-learning process and the assessment methods.

A review of the factor structures in the lecturer instrument demonstrated that the seven-factor composition was the most appropriate structure based on the initial Eigenvalue of > 1 (see Figure 2).

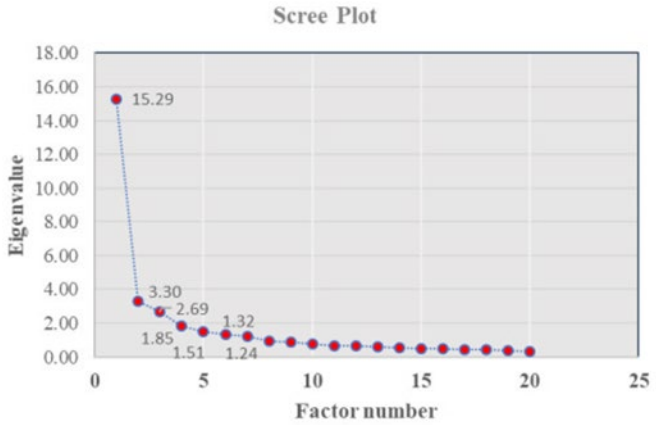

Figure 2: Factors in seven components that explained the variables in lecturer instrument

The factors are: 1) curriculum structure and its implication, 2) learning outcome assessment, 3) lecturer competency, 4) learning content, 5) assessment capacity, 6) mastery of learning material, and 7) learning method. As the sixth factor had 1 item (item 18) and overlapped with factor 1, this factor was combined with the curriculum structure factor. Six out of seven factors were retained in the instrument and accounted for 73.21% of the total variance after 14 iterations (see Table 3) with Cronbach's alpha value of 0.958 for 37 items.

Table 4: Structure matrix of the Oblimin rotation method (37 items & 74 lecturers)

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	h2	Mean	SD
1							-0.848	0.747	3.73	0.896
2							-0.840	0.758	3.54	0.838
3				-0.574		0.591		0.740	3.96	0.766
4	0.637							0.627	4.01	0.785
5	0.667							0.636	4.23	0.713
6	0.755							0.629	4.18	0.709
7	0.521				-0.600		-0.505	0.660	4.07	0.709
8							-0.775	0.652	3.39	0.948
9	0.561						-0.616	0.653	4.20	0.662
10				-0.771				0.678	3.89	0.694
11				-0.818				0.745	3.92	0.614
12	0.558						-0.750	0.754	3.73	0.727
13			0.858					0.785	4.18	0.558
14			0.880					0.796	4.32	0.552
15			0.776					0.684	4.24	0.637
16	0.769						-0.704	0.769	3.96	0.696
17	0.853						-0.529	0.779	4.03	0.702
18	0.626					0.658		0.760	3.93	0.865
19	0.745						-0.505	0.651	4.09	0.878
20					-0.886			0.813	4.03	0.810
21					-0.880			0.822	3.92	0.717
22	0.509				-0.622			0.680	4.16	0.550
23	0.644				-0.572		-0.538	0.734	3.85	0.715
24	0.777				-0.517		-0.621	0.800	3.97	0.702
25	0.756						-0.616	0.782	3.88	0.781
26	0.836						-0.556	0.768	3.78	0.786
27	0.793			-0.566				0.815	3.97	0.721
28	0.709			-0.519			-0.547	0.701	3.93	0.669
29	0.651	0.562						0.669	3.73	0.804
30	0.872						-0.517	0.797	3.95	0.724
31		0.822						0.734	3.25	0.703
32		0.781						0.723	3.25	0.683
33		0.860						0.834	3.16	0.722
34		0.729						0.690	3.38	0.757
35		0.888						0.813	3.12	0.682
36		0.850						0.790	3.27	0.708
37	0.719	0.507						0.620	3.73	0.764

Note: Factor 1 = curriculum structure, Factor 2 = learning outcome assessment, Factor 3 = lecturer competencies, Factor 4 = learning content, Factor 5 = assessment capacity, and Factor 6 = learning method

The items were grouped into the same component based on factor loading > 0.5 (30). For instance, items 3, 10, and 11 were grouped into the same component as learning content (factor 4) because the scored factor loading was > 0.5. Item 27 was excluded from factor 4 because it scored 2 factor loading values (0.793 & -0.566). Hence, item 27 was grouped into a factor with higher factor loading values - factor 1 (curriculum structure).

The item construction for preceptor and patient tools showed that 10 factors met the psychometric property requirements with an adequate number of participants in each group (69 preceptors & 50 patients) (29). The results of KMO and Bartlett's test of sphericity for the preceptor tool were 0.834 and $p < 0.001$ ($X2 = 1243.69$), respectively. Meanwhile, the patient tool scored 0.810 and $p < 0.001$ ($X2 = 1018.13$) for KMO and Bartlett's test of sphericity, respectively.

Table 5: Structure matrix of the Varimax rotation method for the 23 items from 69 preceptors

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	h2	Mean	SD
cs1		0.490		0.530		0.632	3.14	0.896
cs2		0.539		0.496		0.641	2.86	0.809
cs3				0.815		0.736	3.17	0.822
cs4				0.727		0.672	3.03	0.874
cs5			0.710			0.822	3.77	0.789
com1		0.553				0.691	3.17	0.822
com2					0.723	0.817	3.62	0.688
com3			0.734			0.821	3.86	0.827
pb1			0.805			0.867	4.06	0.725
pb2					0.700	0.691	3.72	0.616
pb3					0.816	0.788	3.68	0.675
pb4	0.576		0.541			0.758	3.83	0.766
pb5	0.558					0.630	3.55	0.814
pb6		0.813				0.783	2.93	0.792
pb7		0.497			0.577	0.744	3.59	0.754
ic1		0.832				0.799	3.12	0.814
ic2	0.710					0.726	3.49	0.851
ic3	0.679					0.692	3.52	0.917
ic4	0.711					0.871	3.83	0.804
ic5			0.609			0.755	4.04	0.794
at1		0.660				0.660	3.33	0.780
at2	0.500					0.593	3.84	0.678
at3	0.689			0.497		0.813	3.49	0.868

Note: Factor 1 = professional behaviour, Factor 2 = communication and commitment to work, Factor 3 = attitude towards patient and punctuality, Factor 4 = clinical skill, and Factor 5 = initiative-accountability

Table 4 lists the pattern matrix of psychometric properties from the stance of preceptors. Five factors were retained in the preceptor instrument as their initial Eigenvalue exceeded 1 and supported by 73.9% of the total variance after eight iterations (see Table 4) with Cronbach's alpha value of 0.951 for 23 items (28).

Table 6: The pattern matrix of Principal axis factoring of 20 items for patient tool

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	h2	Mean	SD
Exc1	.616				.927	.859	3.26	.443
Exc2					.800	.761	3.30	.463
Exc3	.517	-.547			.761	.848	3.28	.454
Exc4	.595				.636	.801	3.20	.452
Exc5	.673		.564	-.721	.654	.921	3.28	.454
H1		-.842			.652	.937	3.30	.463
H2	.525	-.895			.501	.896	3.30	.505
H3	.600	-.762	.530			.831	3.26	.487
H4	.553	-.552	.644		.563	.775	3.34	.479
H5	.865		.524		.595	.941	3.26	.443
Ac1	.559	-.709			.508	.851	3.20	.495
Ac2	.911	-.574			.639	.911	3.22	.418
Ac3	.781					.730	3.16	.510
Ac4	.693	-.500		-.794	.661	.928	3.34	.479
Ac5	.634	-.691		-.699	.563	.882	3.36	.525
Alt1	.759		.664		.726	.891	3.18	.388
Alt2	.529		.837			.886	3.24	.476
Alt3			.573			.586	3.08	.528
Alt4	.564		.928			.842	3.18	.482
Alt5	.870		.666		.639	.908	3.24	.431

Note: Factor 1 = altruism, Factor 2 = humanism, Factor 3 = accountability, Factor 4 = professional attitude and empathy, and Factor 5 = pleasant manner

Table 5 shows the pattern matrix of items for the patient instrument with five factors based on the initial Eigenvalue that exceeded 1 and accounted for 81.27% of the total variance after 15 iterations and Cronbach's alpha value of 0.947 for 20 items. The items were grouped into the same factor because the factor loading values exceeded 0.5 (30). For instance, items H1-H3 and Ac1 were grouped in the same component (humanism factor) due to their highest factor loading values from the same group. Based on the pattern matrix of the patient tool (see Table 5), factor 1-altruism has five scales, while factor 2-humanism and factor 3-accountability have four scales each. Meanwhile, factor 4-professional attitude and factor 5-pleasant manner have three and four scales, respectively.

Phase 3

The CFA was performed in Phase 3 to ensure the validity of the four instruments and the suitability of the model with the observed variables. The demographic factor analysis for the preceptor instrument based on 73 participants showed that male participants (54.79%) were more than females (45.21%), with a mean age of 47 years, and most of them were from the Department of Surgery in UNISBA.

The CFA results for the preceptor tool are presented in Figure 3. The 23 items were categorised into five analysis factor groups. The loading values of all items in each group of analysis factor exceeded

1.96 (95% confidence level); signifying that all items are indeed valid, reliable, and reflect the indicator of latent variable.

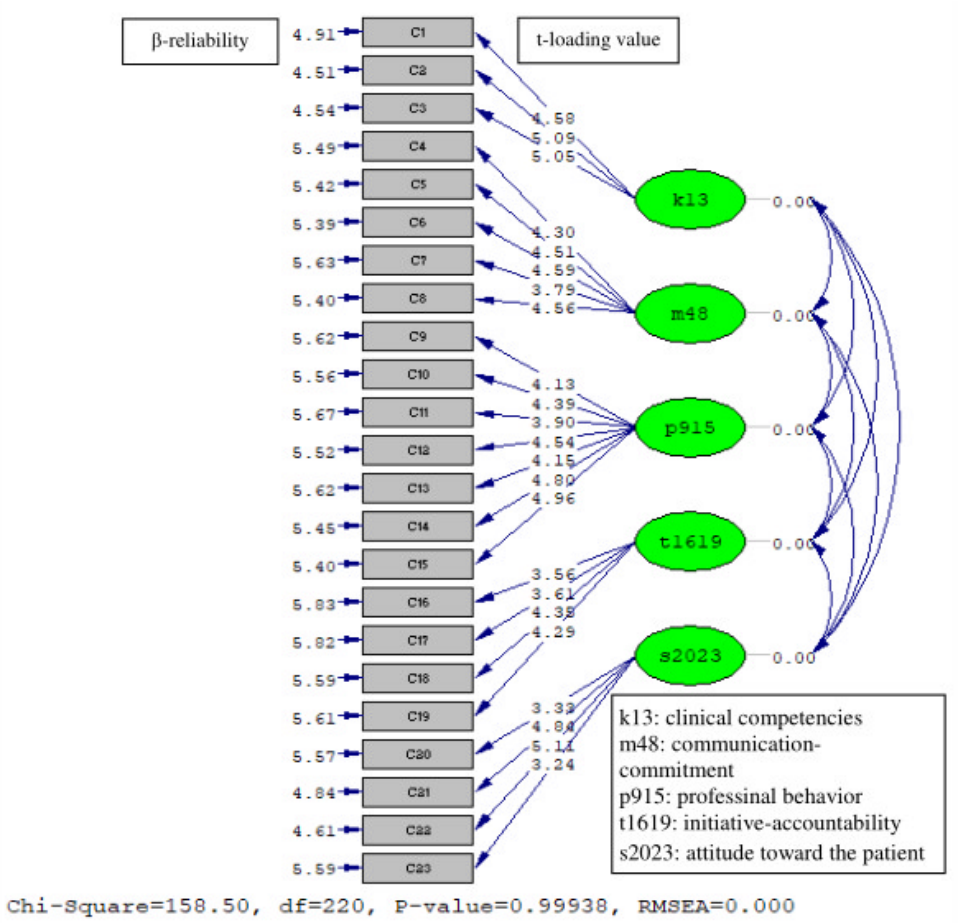

Figure 3: Results of confirmatory factor analysis for preceptor instrument with 73 subjects

The validation of the patient instrument revealed a significant indicator variable for the latent variable (31). Referring to Figure 4, the altruism factor has 5 items and item c14 scored the highest value of 9.22. The item consists of involving their family in decision-making during treatment as consideration from the family is required.

The CFA results of the demographic profile for the student model showed that female students were more prominent than male students with a proportion of 2.5:1. The average age of the students was 20.5 years old with a mean of GPA 3.20. The results of t-value and β-reliability of the selected 183 medical students based on CFA are tabulated in Figure 4.

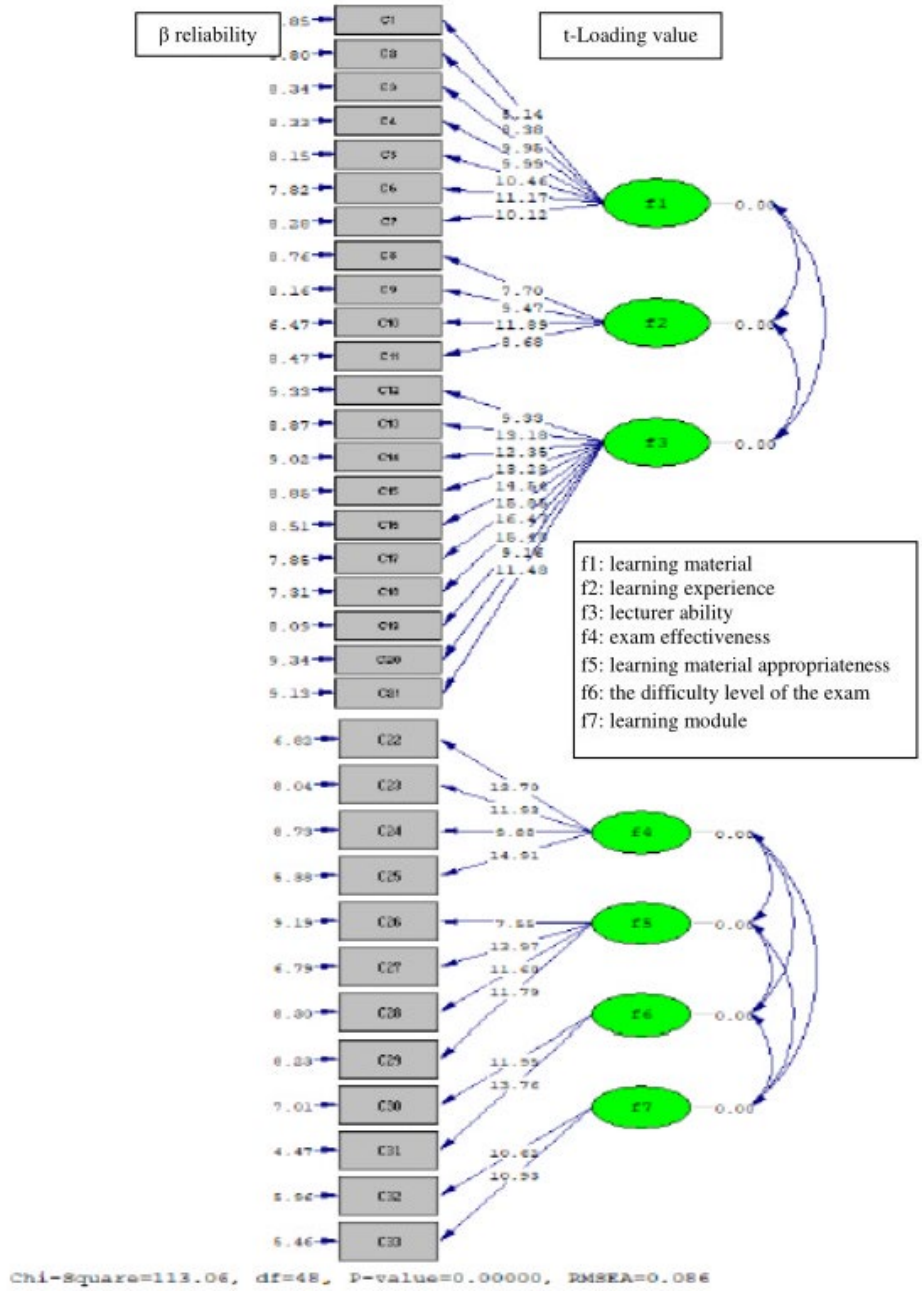

Figure 4: Results of confirmatory factor analysis for student instrument with 182 subjects

The CFA results of the demographic profile for lecturers showed that female lecturers were higher in number than male ones with a ratio of 2:1 (48 & 25). About 43.83% of the lecturers were in the age range of 28-38 years old and mostly had earned a master's degree.

Figure 5 illustrates the outcomes derived from the lecturer model. The 37 items were divided into six groups of analysis factors. The t-value of all analysis factor groups had loading factor values and measurement errors that exceeded 1.96. All the 37 items revealed a significant correlation with their respective factor group.

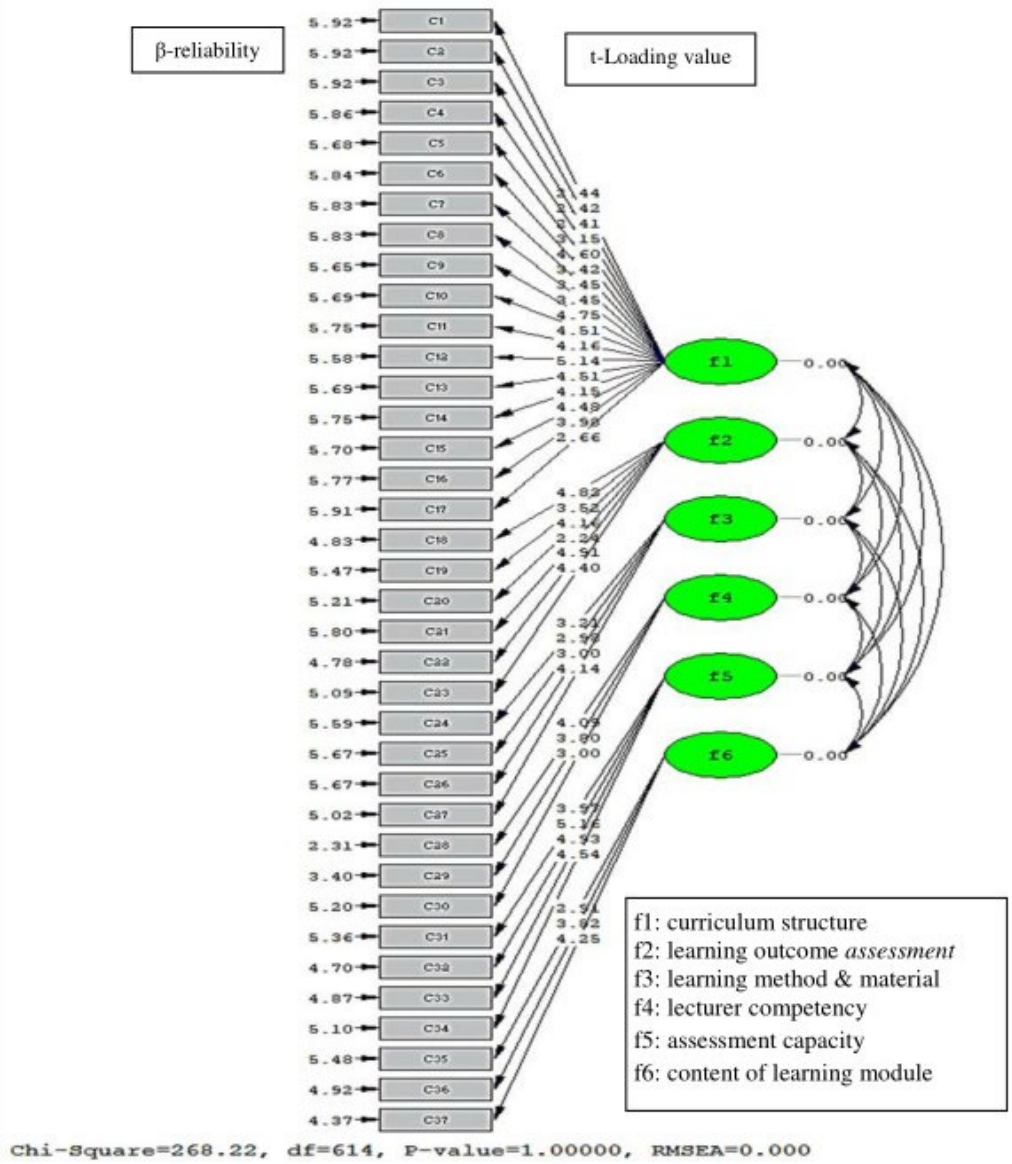

Figure 5: Results of confirmatory factor analysis of lecturer instrument with 73 subjects

The results of t-values and β-reliability for 153 patients are presented in Figure 6. The t-values of all 20 items showed that the loading factor and the measurement error values were above 1.96. All the 20 items displayed a significant correlation with their respective factor group.

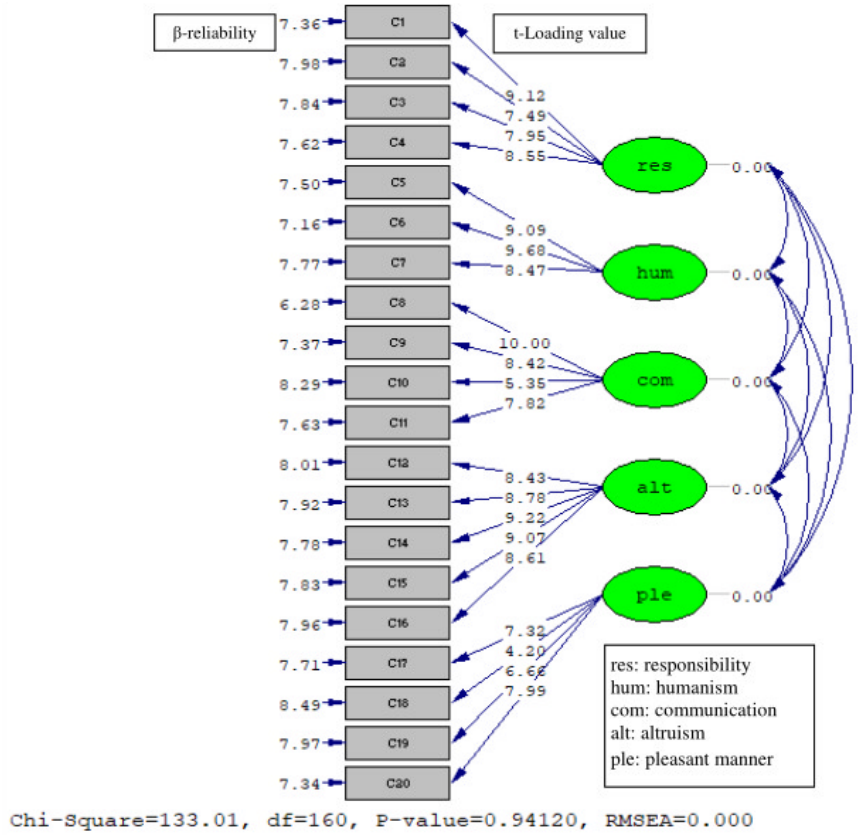

Figure 6: Results of confirmatory factor analysis to 5 latent variables for 153 patients

Summary of CFA Results

Referring to the CFA results, four instruments displayed goodness of fit between the model and their observed data. These findings are tabulated in Table 7.

Table 7: The result summary of confirmatory factor analysis of four model instruments.

Instrument	RMSEA	Cmin/df	Conclusion
Students	0.086	2.35	accepted
Lecturer	0.000	0.437	accepted
Preceptor	0.000	0.718	accepted
Patient	0.000	0.831	accepted

DISCUSSION

Many quarters have improved the assessment of curriculum components. However, such assessment typically concentrates on the teaching-learning process and its efficacy (32) or clinical teaching. For instance, a survey conducted in Canada evaluated the efficacy of clinical teaching (8) as a single component of the learning module (10). Turning to this present study, a 360-degree evaluation model was developed for the medical curriculum based on 4 stakeholder groups via Kirkpatrick Hierarchy.
The themes identified in phase 1 were used for item construction in phase 2. In total, 13 domains of qualitative findings were captured from 2 stakeholder groups related to teaching-learning process and

assessment methods. Lecturers and students considered the components of learning process, assessment methods, and teacher capability in the faculty as crucial factors to determine the success of a curriculum. The teaching-learning process reflected level 1-reaction in Kirkpatrick Hierarchy. The student group identified 7 vital factors to assess curriculum efficacy.

Level 2-learning was measured through the learning program of faculty from the stance of lecturers (17,33). The lecturer group determined 6 factors to evaluate curriculum efficacy. The curriculum structure denotes the arrangement of planned learning experiences to pave the direction, to support the activities, as well as to determine the outline of presenting subjects and teaching methods that best fit a subject (34).
The curriculum content refers to a series of modules that contains learning materials that serve as guide in the teaching-learning process. Both curriculum structure and content were considered essential by both students and lecturers mainly because the two aspects are driven in a successful learning process. Similarly, assessment methods were deemed as a crucial aspect of determining the achievement of learning outcomes (34,35).

Interestingly, the results derived from both student and lecturer instruments differed in terms of the importance of lecturer capacity. The lecturers issued their competencies with 3 items, while the students proposed 10 items for lecturer ability (phase 2). This can be understood because the students are very hopeful that their lecturers can provide plenty of support to achieve their knowledge competence. In addition, a match was noted between lecturers' perceptions and students' points of view about lecturer competency. Similarly, Duvivier stated that in order to achieve good teaching quality, lecturers must possess high didactic and interpersonal skills (36). According to the students, a lecturer's competency or capacity was viewed as a crucial factor in measuring the success of delivering the curriculum. Pedagogical competence refers to a lecturer's competency in education theory, educational psychology, and educational jurisdiction (37). These competencies are important components of teaching. A teacher's ability to manage the learning process is composed of five aspects: learning climate, modelling, coaching, exploration, and articulation (14).

In line with meaningful assessment of learning, both lecturers and students considered that the assessment was aimed at measuring the progress of learning outcomes based on standard criteria (35). The difference in perspectives between lecturers and students on the assessment methods lies in the capacity and the purpose of the assessment. Lecturers placed more emphasis on assessment in light of formative function, while students perceived that assessment is heavy on summative function. As the students believed that the assessment methods were more focused on reaching the advanced level, they identified two domains: exam efficacy and exam difficulty level. Meanwhile, the lecturers considered that assessment is to address its function and capacity in predicting the achievement of students' performance.

The 10 themes derived from two stakeholder groups were related to professional behaviour and medical competencies. These components reflect the assessment of level 3-behavior and level 4-outcome of the Kirkpatrick Hierarchy. Level 3-behaviour describes what the student internship does in the practice of medicine as the preceptors proposed. Next, assessing professional behaviour from the patients' stance indicated the level 4-outcome of education at the top of the Kirkpatrick Hierarchy since they assessed the competencies of junior doctors at the workplace (38). Level 4 is the result of a program, whereby, in this case, it is represented by the performance of medical graduates at the workplace (3,4).

Professional behaviour in this study refers to a set of qualities and attributes of one's behaviour that is observable and related to medical practice (39,40). A similarity was noted in the terminology regarding professional behaviour among alumni and preceptors in terms of serving patients, appropriate medical knowledge, and time management.

Both patients and preceptors viewed communication as a crucial domain due to its link with the contextual factor of paternalistic pattern and cultural consideration in Southeast Asia (41,42). In a similar vein, Wilkinson asserted that good communication indicates an effective interaction with patients,(43) primarily because good interaction with patients can satisfy them (44).

Responsibility refers to completing tasks and meeting the requirements of a contract dealing with the doctor-patient relationship, professionalism, and society (21,43). The patients asserted responsibility as a crucial aspect because they had more expectations from a medical doctor; in line with high-standard professionalism (45). Hence, a physician should take his work with trustworthiness that involves sincerity of intentions, quality of work, and social responsibility (22,40).

This result is in line with other studies that medical professionalism is built by a number of aspects. The eight attributes of medical professionalism are self-awareness, self-management, excellence and commitment to professional development, reflective practice, respect for patients and colleagues, keeping professional confidentiality, accountability, and self-motivation (40,46,47).

A comparison of medical professionalism factors among those derived from preceptors and patients in this study, a blueprint of systematic review, the Islamic stance, and the American Board of Internal Medicine (ABIM) formulation framework is presented in Table 8. The study outcomes showcase some similarities with blueprint professionalism (43) and Islamic stance (40) without disregarding other important features stipulated in the ABIM formulation (48).

Table 8: Comparison of medical professionalism factors extracted in this study with blueprint, Islamic Perspective, and ABIM formulation framework

Professionalism factors extracted	**Blueprint professionalism**	**Islamic perspective**	**ABIM framework**
Professional behaviour	Self-management Reflective practice	Excellence performance	Excellence
Communication skill	Effective interaction with patients	-	-
Clinical skill competence	Continuous improvement of competence	Self-accountability	Accountability
Initiative-commitment	Complete task Commitment to professional development	Strife toward perfection	Duty
Altruism	Balance availability to others with care for oneself	Consciousness	Altruism
Empathy-Attitude	Caring/compassion		Integrity
Humanism	Adherence to ethical practice principles	Faith	Honour
Pleasant manner	Effective interaction with another health workforce	Best character	Respect for others
Responsibility	Take responsibility	Responsibility (*Amanah*)	-

Referring to the CFA results, all latent variables reflected the observed variable indicators. Notably, four instruments (preceptors, students, patients, & lecturers) scored lower RMSEA values (preceptor = 0.000, lecturer = 0.00, patient = 0.000, & student = 0.086).

All the five latent variables in the preceptor instrument significantly influenced their indicator variable. This means; the model instrument is suitable to assess professional behaviour and medical competency among junior doctors in a teaching hospital. Professional behaviour in this study entails enhancing one's capability, having initiative, respecting others, being committed to the best quality, and serving patients; as depicted in studies within the contexts of Arab (47) and Malaysia (49). Similarly, the five latent variables significantly represented the 20 items in the patient instrument.

All six latent variables of the lecturer instrument influenced their indicator variables significantly. This instrument represents an evaluation of the curriculum from the stance of lecturers mainly because some factors constructed a device, whereas the curriculum contains learning content, teaching-learning methods, assessment methods, and learning outcomes. Interestingly, the perspective from lecturers about learning methods, particularly clinical skill training, shared some similarities with a study conducted in Maastricht (36). The study asserted that clinical teaching skills reflect accomplishments that must be achieved by the students.

All the seven latent variables in the student instrument significantly influenced their indicator variables. Hence, this model instrument is suitable to assess the components of teaching-learning process and assessment methods as level 1-evaluation in the Kirkpatrick Hierarchy. The seven items in the student instrument entail curriculum structure, learning experience, teacher competencies, two aspects of assessment methods, and two aspects of learning content. In contrast with the lecturers' perspective, the students believed that lecturer competencies comprised of 10 items, namely: knowledge competency (4 items), pedagogical competency (3 items), ability of learning facilitation (2 items), and one aspect of personality competency.

A learner's experience with evaluation and assessment processes determines how the student approaches learning (50). The major strength of this study lies in the methodical manner in which it was conducted. The development and validation phases in this study adhered to several guidelines, which are standard in some articles pertaining to psychometric properties and analysis factors that are commonly applied to assess behaviour and educational psychology. The development of tools involved some crucial stakeholders within the medical education domain by engaging with both internal and external stakeholders (application of good educational psychology). As for the qualitative segment in this study, the related processes were meticulously recorded and documented.

Study Limitations

Several limitations were noted in this study. First, the collection of qualitative data was conducted at the earlier phase of the research for questionnaire development (exploratory approach). Therefore, a deeper understanding of the research findings could not be sought from the participants. Future research may use our questionnaire as an explanatory approach to look for a deeper understanding of the findings. Second, this study was performed on one faculty of medicine in Bandung. This presents some limitations when generalising the results to other institutions. Third, the interview sessions were conducted in the Indonesian native language, which could lead to misinterpretation of meaning when the emerging themes were translated into the English language.

CONCLUSION

The 360-degree curriculum evaluation model of medicine immensely contributes to the expansion of medical education, especially in terms of the practice of tool development. This evaluation model has been proven to be feasible and acceptable; thus beneficial for curriculum evaluation. However, as any good research work will reveal, there is always more research work to be conducted.

Direction for Further Work

Despite the meticulous steps taken to develop the 360-degree evaluation model, the study outcomes were carefully converted into a survey and later tested on true sample. The validity of the results should be verified in future studies, primarily to determine the generalizability of the outcomes to other institutions and study subjects. While this study yielded a set of variables deemed crucial by students, lecturers, preceptors, and patients; quantitative advance studies are needed to estimate the actual influence these factors have on student and lecturer ratings for teaching quality and learning outcomes.

ACKNOWLEDGEMENTS

We would like to thank the Director of Main Teaching at Hospital UNISBA and the Chairman of the student executive body for their support throughout the data collection process. In addition, we would like to extend our gratitude to the Rector of UNISBA. This work was supported by the Medical Faculty of UNISBA Indonesia and the Research Resources Centre at the University of Cyberjaya Malaysia [grant numbers CRG/04/11/2017].

REFERENCES

1. Spiel C, Schober B, Reimann R. Evaluation of curricula in higher education: Challenges for evaluators. Eval Rev. 2006;30(4):430–50.
2. Oktay C, Senol Y, Rinnert S, Cete Y. Utility of 360-degree assessment of residents in a Turkish academic emergency medicine residency program. Turkish J Emerg Med [Internet]. 2017;17(1):12–5. Available from: http://dx.doi.org/10.1016/j.tjem.2016.09.007
3. Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE Guide No. 67. Med Teach. 2012;34(5).
4. Sanchez-Reilly S, Ross JS. Hospice and palliative medicine: Curriculum evaluation and learner assessment in medical education. J Palliat Med. 2012;15(1):116–22.
5. Swanwick T. Understanding Medical Education. 2nd editio. London: Wiley Blackwell; 2014. 41–44 p.
6. Ruhe V, Boudreau JD. The 2011 Program Evaluation Standards: A framework for quality in medical education programme evaluations. J Eval Clin Pract. 2013;19(5):925–32.
7. Thomas PA, Kern DE, Hughes MT, Chen BY. Curriculum development for medical education: A six-step approach. Curriculum Development for Medical Education: A Six-Step Approach, Third Edition. 2015.
8. Fluit C, Bolhuis S, Grol R, Ham M, Feskens R, Laan R, et al. Evaluation and feedback for effective clinical teaching in postgraduate medical education: Validation of an assessment instrument incorporating the CanMEDS roles. Med Teach. 2012;34(11):893–901.
9. Stalmeijer RE, Dolmans DHJM, Wolfhagen IHAP, Muijtjens AMM, Scherpbier AJJA. The maastricht clinical teaching questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. Acad Med. 2010;85(11):1732–8.
10. Clark ML, Hutchison CR, Lockyer JM. Musculoskeletal education: A curriculum evaluation at one university. BMC Med Educ. 2010;10(1).

11. Tsugawa Y, Tokuda Y, Ohbu S, Okubo T, Cruess R, Cruess S, et al. Professionalism mini-evaluation exercise for medical residents in Japan: A pilot study. Med Educ. 2009;43(10):968–73.
12. Duvivier RJ. Mastering the Art of Medicine. Maastricht; 2012.
13. Biggs JS, Farrell L, Lawrence G, Johnson JK. A practical example of Contribution Analysis to a public health intervention. Evaluation. 2014;20(2):214–29.
14. Boerboom TBB, Dolmans DHJM, Jaarsma ADC, Muijtjens AMM, Beukelen P van, Scherpbier AJJA. Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations. Med Teach. 2011;33(2):84–91.
15. Schönrock-Adema J, Visscher M, Janet Raat AN, Brand PLP. Development and validation of the scan of postgraduate educational environment domains (SPEED): A brief instrument to assess the educational environment in postgraduate medical education. PLoS One. 2015;10(9):1–12.
16. Kwan YH, Png K, Phang JK, Leung YY, Goh H, Seah Y, et al. A Systematic Review of the Quality and Utility of Observer-Based Instruments for Assessing Medical Professionalism. J Grad Med Educ. 2018;10(6).
17. Heydari MR, Taghva F, Amini M, Delavari S. Using Kirkpatrick's model to measure the effect of a new teaching and learning methods workshop for health care staff. BMC Res Notes [Internet]. 2019;12(388):1–5. Available from: https://doi.org/10.1186/s13104-019-4421-y
18. Stenfors-Hayes T, Hult H, Dahlgren LO. What does it mean to be a good teacher and clinical supervisor in medical education? Adv Heal Sci Educ. 2011;16(2):197–210.
19. Anthoine E, Moret L, Regnault A, Sbille V, Hardouin JB. Sample size used to validate a scale: A review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes. 2014;12(1):1–10.
20. Shirali G, Shekari M, Angali KA. Assessing Reliability and Validity of an Instrument for Measuring Resilience Safety Culture in Sociotechnical Systems. Saf Health Work [Internet]. 2018;9(3):296–307. Available from: https://doi.org/10.1016/j.shaw.2017.07.010
21. Mueller PS. Teaching and Assessing Professionalism in Medical Learners and Practicing Physicians. Rambam Maimonides Med J. 2015;6(2):e0011.
22. Arnold L. Assessing professional behavior: Yesterday, today, and tomorrow. Acad Med. 2002;77(6):502–15.
23. Mason M. Sample size and saturation in PhD studies using qualitative interviews. Forum Qual Sozialforsch. 2010;11(3):1–19.
24. Miles M, Huberman A, Saldana J. Qualitative data analysis: A methods sourcebook. 3rd editio. Los Angeles: Sage Publication; 2014.
25. van der Vleuten CPM. Revisiting 'Assessing professional competence: from methods to programmes.' Med Educ. 2016;50(9):885–8.
26. Widarjono A. Applied multivariate analysis with SPSS program, AMOS, and SMARTPLS. 2nd editio. Yogyakarta, Indonesia: UPP STIM YKPN; 2015. 190–197 p.
27. Marsh HW, Morin AJS, Parker PD, Kaur G. Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. Annu Rev Clin Psychol. 2014;10(1):85–110.
28. Kusmiati M, Hamid NAA, Sanip S, Emilia O. Development of an Instrument for Preceptor Evaluation of Medical Graduates ' Performance : the Psychometric Properties. MedSciEduc. 2019;(02 August):1–6.

29. Cabrera-Nguyen P. Author Guidelines for Reporting Scale Development and Validation Results in the *Journal of the Society for Social Work and Research*. J Soc Social Work Res. 2010;1(2):99–103.
30. Costello AB, Osborne JW. Best Practices in Exploratory Factor Analysis : Four Recommendations for Getting the Most From Your Analysis. Pract Assess Res Eval. 2005;10(7):1–9.
31. Kusmiati M, Bahari R, Hamid NAA, Sanip S, Emilia O. Validation of Patient Perception Instruments for Junior Doctor Performance: a Factor Analysis. Glob Med Heal Commun. 2019;7(1):71–80.
32. Guolla M. Assessing the Teaching Quality to Student Satisfaction Relationship : Applied Customer Satisfaction Research in the Classroom. J Mark Theory Pract. 2015;7(February):1–12.
33. Bates R. A critical analysis of evaluation practice : the Kirkpatrick model and the principle of beneficence. Eval Program Plann. 2004;27(2004):341–7.
34. Oliver R, Kersten H, Vinkka-puhakka H, Alpasan G, Bearn D. Curriculum structure : principles and strategy. Eur J Dent Educ. 2008;12(suppl 1):74–84.
35. Lockyer J, Carraccio C, Chan MK, Hart D, Smee S, Touchie C, et al. Core principles of assessment in competency-based medical education. Med Teach. 2017 Jun 3;39(6):609–16.
36. Duvivier RJ, Dalen J Van, Vleuten CPM Van Der, Scherpbier AJJA, Dalen J Van, Vleuten CPM Van Der, et al. Teacher perceptions of desired qualities , competencies and strategies for clinical skills teachers Teacher perceptions of desired qualities , competencies and strategies for clinical skills teachers. Med Teach. 2009;31(July):634–41.
37. Sherbino J, Frank JR, Ed MA, Snell L. Defining the Key Roles and Competencies of the Clinician – Educator of the 21st Century : A National Mixed-Methods Study. Acad Med. 2014;89(5):1–7.
38. Tavakol M, Gruppen LD, Torabi S. Using evaluation research to improve medical education. Clin Teach. 2010;7(3):192–6.
39. Lesser CS, Lucey CR, Egener B, Braddock CH, Linas SL, Levinson W. A behavioral and systems view of professionalism. JAMA - J Am Med Assoc. 2010;304(24):2732–7.
40. Kasule OH. Medical professionalism and professional organizations. J Taibah Univ Med Sci [Internet]. 2013;8(3):137–41. Available from: http://dx.doi.org/10.1016/j.jtumed.2013.09.003
41. Claramita M, Nugraheni MDF, van Dalen J, van der Vleuten C. Doctor-patient communication in Southeast Asia: A different culture? Adv Heal Sci Educ. 2013;18(1):15–31.
42. Irby DM, Hamstra SJ. Parting the Clouds: Three Professionalism Frameworks in Medical Education. Acad Med. 2016;91(12):1606–11.
43. Wilkinson TJ, Wade WB, Knock LD. A blueprint to assess professionalism: Results of a systematic review. Acad Med. 2009;84(5):551–8.
44. Francis MD, Warm E, Julian KA, Rosenblum M, Thomas K, Drake S, et al. Determinants of Patient Satisfaction in Internal Medicine Resident Continuity Clinics: Findings of the Educational Innovations Project Ambulatory Collaborative. Grad Med Educ [Internet]. 2014;471–7. Available from: http://dx.doi.org/10.4300/JGME-D-13-00398.1
45. Joynt GM, Wong W, Ling L, Lee A, Matthew G, Wong W, et al. Medical students and

professionalism – Do the hidden curriculum and current role models fail our future doctors ? Med Teach [Internet]. 2017;0(0):1–5. Available from: https://doi.org/10.1080/0142159X.2017.1408897

46. Al-Eraky MM, Donkers J, Wajid G, Van Merrienboer JJG. A Delphi study of medical professionalism in Arabian countries: The Four-Gates model. Med Teach. 2014;36(SUPPL.1):S8–16.
47. Al-Rumayyan A, Van Mook WNKA, Magzoub ME, Al-Eraky MM, Ferwana M, Khan MA, et al. Medical professionalism frameworks across non-Western cultures: A narrative overview. Med Teach [Internet]. 2017;39(0):S8–14. Available from: http://dx.doi.org/10.1080/0142159X.2016.1254740
48. Agha S, Gooda MR. Undergraduate Clinical Students' Perception of Professionalism in Their Learning Environment. Int J Arts Sci [Internet]. 2015;8(6):593–615. Available from: https://search.proquest.com/docview/1764688846?accountid=26688%0Ahttp://link.periodicos.capes.gov.br/sfxlcl41?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Amaterialsscijournals&atitle=UNDERGRADUATE+CLINICAL+STU
49. Haque M, Zulkifi Z, Zohurul Haque S, Kamal Z, Salam A, Bhagat V, et al. Professionalism perspectives among medical students of a novel medical graduate school in Malaysia. Adv Med Educ Pract. 2016;Volume 7:407–22.
50. Ginns P, Prosser M, Barrie S. Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. Stud High Educ. 2007;32(5):603–15.