ORIGINAL ARTICLE

Design, Development & Validation of an Assessment Tool for Undergraduate Medical Students of Community Health Sciences Rotation

FARHAT R MALIK[1], SAIMA ALI[2], IRFAN ULLAH[3]

[1]Professor Department of Community Health Sciences, Peshawar Medical College, Warsak Road, Peshawar. Riphah international University, Islamabad. Pakistan.

[2]Professor Department of Paediatrics- Kuwait Teaching hospital, Peshawar Medical College, Warsak Road, Peshawar. Riphah international University, Islamabad. Pakistan.

[3]Department of Life Sciences, School of Science, University of Management and Technology (UMT), Lahore, Pakistan

Corresponding authors: Irfan Ullah, Email: irfan_btn@hotmail.com

ABSTRACT

Background: Questionnaires are the most commonly used data collection methods in applied research for assessment of inputs. It is a useful instrument if valid and reliable.

Objectives; To establish, design and appraise the reliability and validity tool for measuring knowledge with skills among undergraduate students.

Methods: An observational study conducted at Peshawar Medical College in six months' duration through three stage process, after having approval of Institutional Review Board of Prime Foundation. Process initiated by slot regulation, component development and questionnaire generation with judgement analysis of instrument by an expert panel of five public health specialists for relevance, representativeness and transparency of each item based on Likert rating scale. Validity and reliability measured in the final steps. Suggestions put forward by the experts with item impact scores corresponded to face validity. Rewording, combination and elimination resulted in final 35 item instrument.Data was analysed through SPSS Version- 21 with computation of content validity ratio, content validity index, item content validity, scale validity, Kappa statistics and Cronbach's Alpha values.

Results: Mean years of experience for the panellists was 14.2 years with S. D + 5.2 (n= 5). Excellent CVR, I CVI, S CVI, Percent Agreement and Kappa statistics were calculated for the entire questionnaire as 1.The final 24 item knowledgesection had 0.732 Inter Class Correlation and acceptable Cronbach Alphaas 0.743, while the 11 item skill portion had 0.819 Inter Class Correlation with good Cronbach Alpha 0.890.

Conclusions: The findings support the face and content validity of the questionnaire.

Keywords: Instrument development and Validation, Competency, Knowledge, Skill, Assessment, Content Validity, Content Validity Ratio, Reliability, Questionnaire

INTRODUCTION

Medical education developed remarkably with new concepts to its portfolio. Tutoring became meticulous with pedagogical principles and problem based curriculum that enhance effectivelearning. Instructorsproceeded from problem-identifier to a solution-provider. The efficacious healthcare delivery requires proficiency with interpretative and communication skills, that sounds comprehensive and robust enough assessment systems. These systems estimate the required aspect with leading knowledge and skills. A purposeful driven evaluation has a supreme conclusive governing response upon learning. It is crucial and execute as the most relevant instigator of student learning. Exams developed to valuate many interspersed competence, for instance accurate intelligence, cogent, inquiry with integration of information can be problematic to measure the progress. Students may show progress in one trait so diverse abilities be measured through progress tests [1]. Educationists persuadeassessment, course objectives and intended outcomes be aligned with the feedback to improve competencies [2].

Promotion of health and welfare of the community through primary health-care approach is known as Community Health Sciences. It gives holistic with comprehensive training approach to competence in dealing with primary health care, evidence-based practiceand teamwork with professional humane behavior to endorse population's health [3]. Student's clinical reasoning is efficient in areas with basic systematic knowledgewhereas quite low in the territories strange to them [3, 4].

"The perpetual and factitiouscommunication use, competence, scientific proficiency, logical explanation, empathy, morals, and consideration in individuals and community prosperityis termed as competence [4, 5]. Medical colleges, postgraduate couching programs, and licensing authorities over the past decapodhave made new efforts to present steady and proper valuation of student's proficiency [5, 6, 7].

Modern research is complicated with multiple sets of skills like medical, social, technological, mathematical and statistical. Suitable instruments give unbiased error free results of the indicators. Questionnaire is the most popularand frequently used technique to evaluate applied research. The significance of veracity and suppleness measurement of the tools is validity and reliability cited in researches but their capacity is not accomplished in developing world. It is associated to the paucity of tests familiarity[8, 9]. Content validation exemplify that intends to give pledge that the tool gauges the area supposed to check [10].

Face validity is allegedly linked to basic study design. It is the agreement with items as well as the wording in an instrument aligned with the research intentions. Validity is pertinent to the magnetism of a research tool affecting respondent's opinion. It does not acknowledge what to test rather targets the appearance of tool, moreoverit is seen as a weak form of construct validity, but still useduniversally in developing countries [9, 11, 14].

It is essential to measure tools content validity to assure construct validity as well as viewer's assurance. Variables are tested by content validity alias content related validity, intrinsic validity, representative validity, relevance validity and convincing effectiveness. It is used to test the relevantdiscipline items in a questionnaire [11]. Competent personal decision is needed to ascertain the degree of constructed tool to quantify attributes[12]. Atleast five experts have enough authority to decide the content domains of an instrument through rating scales.Professional's number have always been capricious. It is mandatoryto have 10 competent judges because when number of experts rises, final verdict decreases [13, 14, 15].

Researchers use expert's notions. Interviews are executedin qualitative research with the target groups. Dialogues crucial lists include Items level difficulty, appropriateness, associationwith the major intention in relation to the study tool, dubiety, item delusion and un-comprehensibility of the words essence. To record study populationviews is a crucial part of content validity which yield desirable outcome as students are familiar with the construct. They identify crucial items and grade them on Likertrating scale by pickingstep wise through the most important to non- important. In quantitative method, item impact score calculations are done. Mostlypeople who score four or five to item frequency importance

is calculated with itemsmean scores and finally item impact scores of the entire instrument[16, 17, 18, 19].All the terms used in this study are outlined as flowchart [20] in Figure- 1.

Quantification of human behaviors is an important element of all the researches done in social sciences sector using instrument through observation but the tool should be valid and reliable.

The rationale behind carrying out this research is to give the concerned researchers an insight of two important concepts being widely used in social sciences with detailed steps in designing and developing a tool and secondly to introduce the relevant methods to assess validity with reliability in relation to behavioral research and Cronbach's Alpha model and interpretation.

MATERIALS AND METHODS

A study conducted at Peshawar Medical College in six months to appraise the validity and accuracy of a tool for undergraduate medical students. Ethical endorsement from the Institutional Review Board of Prime Foundation was taken. Panel selected with a post-graduation in public health, more than 5 years teaching experience, familiarity with the thematic domains, curriculum development guidelines as per PMDC rules and policies [21]and 100% response rate during three rounds.

Instrument design: was through a three stage process. First stage included domain determination. Item generation and instrument construction done through focus group discussions. Explored thematic domains were;

Knowledge;Common abbreviations in Community Health Sciences, vaccine preventable diseases, hand washing, rehabilitation, malnutritionand MUAC tape measurements, vaccines and their required temperature, oral rehydration and waste management, delays of maternal mortality, levels of prevention, differentiation of the terms in reference to immunization and matching figures with the statements related to the taught concepts.

Skill; Nutritional status through body mass index and expected date of delivery calculations from the given scenarios, interpretations of scenarios to identify the nutritional status through mid- upper arm circumference tape, visual acuity cut off limits in blindness, drawing and labelling of communication cycle and health care delivery model and growth chart plotting.

Results generated a 56 item assessment sheet. Instrument's constructionwas done by refining and sequence organizing.

Judgment: required five public health experts selected randomly with modified Delphi technique. Experts were requested to rate individual item on a Likert scale from (1 as Strongly Disagree, 2 as Disagree, 3 as Neutral, 4 for Agree and finally 5 for Strongly Agree) as well as rating individual items on (Congruity, Accuracy, Integrity and Obscurity) on a four-point scale (1-4) and advocate adjustments in wording, identify prolixity and recommend further items.Qualitative and quantitative expert's suggestions were assembled in relation to item's relevancy, accuracy, representativeness and comprehensiveness to gauge constitute; optionally delineated by these items to establish the content validity [2 & 3]. In relation to this phase, three Delphi rounds were conducted [22].

Item Selection: Experts quantified content validity for appropriateness by computing each item's content validity ratio through CVR= (Ne - N/2) / (N/2), Ne being number of panellists indicating "essential" and N is the total members in the panel, varying between 1 and -1. Item necessity in the scale was depicted by high scores. Lawshe table decided for the arithmetic assessmentof content validity ratio [23]. A five-member panel required a minimum CVR of .99, to satisfy the 5% level of significance. When everyonein the panel say "essential," the CVR is calculated to be 1.00. However, if the panel membersare more than half but less than all, then the CVR lies between zero and 0.99 [21], then the items or questionswere included in the final draft. To retain an item, agreement level 1 must be achieved and everyone in the panel must agree[7, 8, 9,23].

The Content Validity Index: CVI determinationretainsitems, then the content validity indexfor the entire assessment tool is computed [11]. CVI is the mean of the included items CVR.Item rating and scale level rating are crucial for content validity. The itemrated content validity indices are connoted as I-CVI, while the scale content validity index as S-CVI. I-CVI calculated with level of agreement among raters with ≥ 0.78 as significant level for item inclusionfor S- CVI [14, 15, 24].

Relatedness and articulateness of each item (I-CVIs) was calculated byrelevant items divided by the total number of experts. Content validity index congruence was determinedfor item level (I-CVIs) as well as the scale-level (S-CVI). Item-CVI explicit the agreement proportion on the relevancy of every item between 0 and 1and the S- CVI is characterized as "the proportion of items on an instrument that achieved a rating of 3 or 4 by the content experts" [3, 24].

Gadget planners never give details for computing the scale-level index (S-CVI) [24].Universal agreement among experts is S-CVI/UA that is calculated by summing up all relevant ratings put forward by experts and dividing them by the total items. But a less traditional approach equates the item-level CVIs (S-CVI/Ave). Scale is dichotomized and only two dichotomous groups are formed for every item as "relevant and not relevant" [3, 20]. The total of items treated compatible or relevantby all the professionals (CVI equal to 1) is divided by the grand total of items and is termed as universal approach whereas in the average approach, the sum of I-CVIs is divided by the total number of items.Three ways are in practice to calculate the S-CVI/Ave. The first by averaging proportion of items rated relevant among experts and can be calculated by summing up all expert ratings divided by number of experts, yet another way is to average the I-CVIs by summing them and dividing by the number of items. Thirdly this can be calculated by counting the total number of relevant items rated by experts and to then divide by the total number of ratings. Same results will be the outcome of all the three computations. Researchers think to give average approach values as it focuses on average item quality as compared to average performance by the professional experts. Moreover, average approach is same as average congruency percentage [10].

Researchers must describein detail both methods as it might generate different values and consider 80 % compliance or above among the professionals for brand new tools [24, 27]. Each item is inferred upon values of I-CVI. The item is considered appropriate with value higher than 79 percent, needs revision with values between 70 and 79 percent and eliminated with values less than 70 percent[24]. Researchers use CVI extensively to estimate content validity, but it does not acknowledge the probability of bloated values due to chance agreement. Therefore, the researcher's recommendation of content validity index and kappa statistic in the study [24].

Expert's recommendations with item impact scores measured the face validity. Item impact score were done by first calculating percent, who scored 4 or 5 to item importance, frequency and the mean importance score of item and then item impact score of instrument items by; Item Impact Score= frequency × Importance[16, 17, 18, 19]. Item impact equal to or greater than 1.5 corresponds to a mean frequency of 50% and an importance mean of 3 on the 5-point Likert scale, is retained otherwise eliminated [20].

Cronbach's Alpha assesses reliability that refers to the amount of variance attributed to the final scores of the construct. This uses inter item correlation to check if the domain was correctly measured by the constructed items and show homogeneity. This value must be as high as possible because low values will show low reliability and will be dropped out. Cronbach Alpha should exceed 0.70 for a developing tool and 0.80 for a more established tool with unit measurements necessary for each of the domain rather entire tool. Reliability is essential for validity. Cronbach's Alpha(**α**), was interpreted on the basis of the results produced in a study as (α > 0.9; Excellent, 0.8 α >0.9; Good, 0.8 α

> 0.7; Acceptable, 0.7 α > 0.6; Questionable, 0.6 α> 0.5; Poor, 0.5 > α; Unacceptable)[28].

Data Analysis: was done through SPSS Version- 21 with computation of Content Validity Ratio [5], Content Validity Index [6, 7, 11], Item Content Validity, Scale Validity [8], Kappa Statistics and Cronbach's Alpha [10] values respectively. The minimum acceptable CVI between five experts is > 0.78 at 0.05 level of significance there by implying everyonemust agree to retain the items in the questionnaire. In addition, according to Lawshe[20, 23, 25, 26, 27], nonetheless it is accepted with a value < 0.78 of a question and mean judgments > 1.50.

RESULTS

Panel members were post graduate public health experts, working as faculty members, researchers and instructors having rich experience as teachers as well as in curriculum development (5-20 years). Mean years of experience for all the panellists was 14.2 years with S.D+ 5.2 (n= 5).

The final tool contained 35 items (24 in knowledge and 11 in skill domain), after review was consolidated and analysed from the original having 56 items. Items with low agreement among the reviewers (CVI< 1) were removed and suggestions were accommodated in all the rounds.Items of the scale not aligned with the domain and values less than acceptable in case of CVI were removed.

Scale Content Validity Index/ Universal Agreement for knowledge came to be 0.571 and for skill as 0.561.

CVR, I- CVI, S- CVI, Percent Agreement and Kappa statistics were calculated for the entire questionnaire and values as 1were termed as excellent to be retained within the sample questionnaire as depicted in Table- 1, 2 and 3 whereas rest of the

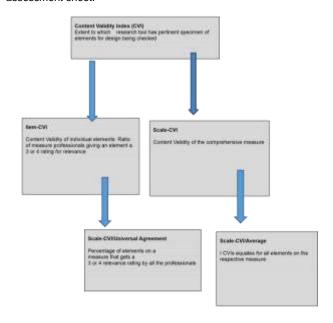items with less than 1 value were rejected and removed from the assessment sheet.

Figure 1: Definitions of the terms used and calculated for Validation in present study [20]

Table 1: Content Validity Ratios, Item Scale Content Validity, Average Scale Content Validity, Universal Agreement Content Validity, Probability of Chance Agreement, Kappa Statistics Calculated Values for the final tool with interpretations of Knowledge Items.

Questions	CVR	I CVI	S CVI/Ave By all three methods[*]	Probability of chance agreement(Pc)	Kappa Statistics(K)	Interpretation
1	1	1		3125	1	
2	1	1		3125	1	
3	1	1		3125	1	Excellent
4	1	1		3125	1	
5	1	1		3125	1	
6	1	1		3125	1	
7	1	1		3125	1	
8	1	1		3125	1	
9	0.6	0.8		Dropped Out		
10	0.6	0.8				
11	1	1		3125	1	
12	1	1		3125	1	
13	1	1		3125	1	Excellent
14	1	1		3125	1	
15	1	1		3125	1	
16	0.6	0.8	1. 0.912[*]			
17	0.6	0.8	2. 0.914[*]			
18	0.6	0.8	3. 0.910[*]			
19	0.6	0.8				
20	0.6	0.8		Dropped Out		
21	0.6	0.8				
22	0.6	0.8				
23	0.6	0.8				
25	0.6	0.8				
26	1	1		3125	1	
27	1	1		3125	1	
28	1	1		3125	1	Excellent
29	1	1		3125	1	
30	1	1		3125	1	
31	1	1		3125	1	
31	1	1		3125	1	
32	1	1		3125	1	
33	1	1		3125	1	
34	1	1		3125	1	

35	0.6	0.8				
36	0.6	0.8		Dropped Out		
37	0.6	0.8				
38	0.6	0.8				
39	0.6	0.8				
40	0.6	0.8				
41	1	1		3125	1	Excellent
42	0.6	0.8		Dropped Out		
Grand Total	0.829	0.900	0.911			

Table 2: Content Validity Ratios, Item Scale Content Validity, Average Scale Content Validity, Universal Agreement Content Validity, Probability of Chance Agreement, Kappa Statistics Calculated Values for the final tool with interpretations of Skills Items.

Questions	CVR	I CVI	S CVI/Ave By all three methods*	Probability of chance agreement (Pc)	Kappa Statistics (K)	Interpretation
1	1	1	0.957* 0.957* 0.957*	3125	1	Excellent
2	1	1		3125	1	
3	1	1		3125	1	
4	1	1		3125	1	
5	1	1		3125	1	
6	0.6	0.8		Dropped Out		
7	0.6	0.8				
8	0.6	0.8				
9	1	1		3125	1	Excellent
10	1	1		3125	1	
11	1	1		3125	1	
12	1	1		3125	1	
13	1	1		3125	1	
14	1	1		3125	1	
Grand Total	0.914	0.960	0.957	3125	1	

Table 3: Internal Class Correlation Coefficient &Cronbach's Alpha values with interpretations for Knowledge and Skill domains

DOMAINS	Items (n= 56)	Internal Class Correlation Coefficient (ICC)	Cronbach's Alpha	Interpretation
Knowledge	24	0.732	0.743	Acceptable
Skill	11	0.819	0.890	Good

The final 24- item knowledge section had 0.732 Inter Class Correlation with acceptable Cronbach Alpha as 0.743, while the 11 item skill portion had 0.819 Inter Class Correlation with good Cronbach Alpha 0.890. These values correlate towards a reliable, credible and valid tool for knowledge and skill assessment among the students.

DISCUSSION

The present study designed, developed and validated an essential assessment tool for undergraduate students of Community Health Sciences rotation. The main findings revealed 24 assessment questions in knowledge domain and 14 for skills measurement. Evaluated items were considered content valid for testing knowledge and skills of undergraduate medical students according to CHS and PMDC requirements.

A PhD scholar reviewed 38 articles and found only 20% researchers who mentioned content validity in their articles, claiming content validity is not considered important however interpretation of results become precise. He further explained that content validity is an important factor in determining the measuring concept; however, it is not enough indication that the instrument amplify what it is supposed to ascertain. A single approach is not enough but variety of approaches must be tested [11].

A researcher demonstrated indices for domain accuracy regarding a brand new tool and discussed with them during composition and psychiatric patient centred communication measuring tool. He added that affirmation is a tedious course by assessing accuracythrough intramural firmness and check recheck, design accuracy (through factor analysis) and benchmark accuracy [24].

A study conducted in 2018 was about designing, validating and applying the questionnaire to assess ability, competence, societal obligation and practical training. Calculated CronbachAlpha was within acceptable range.They used public health teaching, learning and skill development with risk assessment among communities which was lacking in present study, however present study measured community health sciences discipline which is a part of public health and it was created only for 3rd year students not yet applied. Whereas the said study applied it on 3 and 4th year undergraduate students. CA was acceptable for present study questionnaire as well [29].

Ashok Kumar's review article presented a systematic and logical approach in validating an assessment tool, using a framework and illustrations to support with factor and item analysis. However, the present study did not take into account factor analysis, as it is yet to be applied. The researchers only validated the assessment tool for undergraduate students in a specific discipline and outlined the whole process [30].

Limitations include subjective errors by experienced feedback. Some questions had limited validity from knowledge and skill domain categories. They were modified according to the expert's wishes.Delphi rounds need at least 20 people in a panel, but only five experts here. Generalization and transfer of these results to other locations need to be explored and dealt with caution as done in one institute. This instrument does not cover all dimensions of public health and community health sciences subject. As it would have resulted into a many items questionnaire making it very difficult for the ones taking it. However, needs of the undergraduates are fully met here with the designed tool however lacks factor analysis and discriminatory index as it is yet to be applied on students.

CONCLUSION & RECOMMENDATION

Validity and reliability of the designed assessment tool was sufficient enough to evaluate knowledge and skills of the undergraduate medical students of Community Health Science's discipline.

Training regarding content validity and the process of validation must be provided to teachers, students and researchers that will enable them to understand better, criticize and use research tools with a more authentic approach.Careful consideration of augmenting accuracy, representativeness and

effectiveness of the themes of the questionnaires will produce purposeful studies having scientific results and interpretations.

REFERENCES

1. Tabish SA. Assessment Methods in Medical Education. International Journal of Health Sciences. 2008; 2 (2): 3-7.
2. Shunmugam J, Gurupatham D. Assessment of competency of medical students in community medicine. Int J Community Med Public Health 2016; 3:99-103.
3. Murugavel J, Chellaiyan VG, Krishnamoorthy D. Attitude toward learning of community medicine: A cross-sectional study among medical school students. Journal of Family Medicine and Primary Care. 2017; 6 (1):83-7.
4. Epstein RM. Assessment in Medical Education. NEJM. 2007; 356: 387-96.
5. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002; 287:226-235
6. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. Health Aff (Millwood) 2002;21:103-111
7. Leung WC. Competency based medical training: review. BMJ 2002;325:693-96
8. Singh AS. Common Procedures for Development, Validity and Reliability of a Questionnaire. International Journal of Economics, Commerce and Management. 2017; 5 (5):790- 801.
9. Bolarinwa OA. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. Niger Postgrad Med J. 2015; 22 (4):195-201.
10. Frank-Stromberg, M., & Olsen, S. J. (2004). Instruments for clinical health-care research. London, England: Jones &Bartlette.
11. Yaghmale F. Content Validity and its estimation. Journal of Medical education 2003; 3 (1): 25- 7.
12. Nunnally JC. Psychometric theory, 2nded. New York: McGraw-Hill book Company; 1978.
13. Wilson HS. Research in nursing, 2nd ed. California: Addison-Wesley Publishing company; 1989.
14. Sangoseni O, Hellman M, Hill C. Development and validation of a questionnaire to assess the effect of online learning on behaviors, attitude and clinical practices of physical therapists in United States regarding of evidencebased practice. Internet J Allied Health SciPract 2013; 11: 1-12.
15. Development of an instrument for measuring Patient-Centered Communication Journal of Caring Sciences, Jun 2015; 4 (2), 165-178
16. De Von HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ. et al. A psychometric toolbox for testing validity and reliability. J NursScholarsh. 2007; 39 (2):155–64.
17. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986; 35 (6): 382–5.
18. Waltz CF, Strickland O, Lenz ER. Measurement in nursing and health research. 4th ed. New York: Springer Publishing Company; 2010.
19. Newman I, Lim J, Pineda F. Content validity using a mixed methods approach: Its application and development through the use of a table of specifications methodology. Journal of Mixed Methods Research. 2013; 7 (3): 243–60.
20. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 2006; 29: 489e97.
21. Pakistan Medical & Dental Council; Curriculum of MBBS- Revised 2011 (Page; 53), [Internet][Copyright 2006; Cited 2018 Sep 29] Available from:http://www.pmdc.org.pk/LinkClick.aspx?fileticket=EKfBIOSDTkE%3D
22. Kikukawa et al. An instrument for evaluating clinical teaching in Japan: Content validity and cultural sensitivity, BMC Medical Education 2014 14: 179
23. Lawshe CH. A quantitative approach to content validity. Personnel psychology 1975; 28 (4): 563-75.
24. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar A-R. Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. Journal of Caring Sciences. 2015; 4 (2):165-78.
25. Shirali G, et al., Assessing Reliability and Validity of an Instrument for Measuring Resilience Safety Culture in Sociotechnical Systems. Saf Health Work 2018; 9; DOI: 296- 307; http://dx.doi.org/10.1016/j.shaw.2017.07.010
26. Charter RA. Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. J ClinExpNeuropsychol 1999; 21: 559e66.
27. Davis L. Instrument review: Getting the most from your panel of experts. ApplNurs Res 1992; 5:194e7.
28. Singh A. Common Procedures for Development, Validity and Reliability of a Questionnaire. International Journal of Economics, Commerce and Management. 2017; V (5).
29. Vackova D, Chen CK, Lui JNM, Johnston JM. A validation study of public health knowledge, skills, social responsibility and applied learning. International journal of medical education. 2018;9:175-81.
30. Kumar A. Review of the Steps for Development of Quantitative Research Tools. Advance Practices in Nursing. 2015; 1(1):1- 7.