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

FARHAT R MALIK¹, SAIMA ALI², IRFAN ULLAH³
¹Professor Department of Community Health Sciences, Peshawar Medical College, Warsak Road, Peshawar. Riphah international University, Islamabad. Pakistan.
²Professor Department of Paediatrics- Kuwait Teaching hospital, Peshawar Medical College, Warsak Road, Peshawar. Riphah international University, Islamabad. Pakistan.
³Department 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 knowledge section had 0.732 Inter Class Correlation and acceptable Cronbach Alphaa 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 effective learning. Instructors proceeded from problem-identifier to a solution-provider. The efficacious healthcare delivery requires proficiency with interpretative and communication skills, that sound 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 persuade assessment, 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 practice and teamwork with professional humane behavior to endorse population’s health ³. Student’s clinical reasoning is efficient in areas with basic systematic knowledge whereas quite low in the territories strange to them ³-⁴.

The perpetual and factitious communication use, competence, scientific proficiency, logical explanation, empathy, morals, and consideration in individuals and community prosperity is termed as competence ⁴, ⁵. Medical colleges, postgraduate couching programs, and licensing authorities over the past decapohade 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 popular and frequently used technique to evaluate applied research. The significance of veracity and supleness 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. Content validation exemplify that intends to give pledge that the tool gauges the area supposed to check ⁹.

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, moreover is seen as a weak form of construct validity, but still used universally in developing countries ¹⁰, ¹¹, ¹². 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 relevant discipline 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 mandatory to have 10 competent judges because when number of experts rises, final verdict decreases ¹³, ¹⁴, ¹⁵.

Researchers use expert’s notions. Interviews are executed in qualitative research with the target groups. Dialogues crucial lists include items level difficulty, appropriateness, association with the major intention in relation to the study tool, dubiety, item delusion and un-comprehensibility of the words essence. To record study population views 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 Likert rating scale by picking step wise through the most important to non-important. In quantitative method, item impact score calculations are done. Mostly people who score four or five to item frequency importance.
is calculated with item mean scores and finally item impact scores of the entire instrument. All the terms used in this study are outlined as flowchart 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 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, malnutrition and 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 construction was 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 (Strongly Disagree) to 5 (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 construct; optionally delineated by these items to establish the content validity. In relation to this phase, three Delphi rounds were conducted.

Item Selection: Experts quantified content validity for appropriateness by computing each item’s content validity ratio through CVR = (Ne - N2) / (N2), Ne being number of panelists 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 1. Item impact score of instrument items by; Item Impact Score = frequency × Importance. Cronbach’s Alpha model and interpretation.

The Content Validity Index: CVI determination involves items, then the content validity index for the entire assessment tool is calculated. CVI is a statistical index of the item rating and scale level rating are crucial for content validity. The itemized content validity indices are computed 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 inclusion or S-CVI ≥ 0.8. 

Relatedness and articulateness of each item (I-CVIs) was calculated by the arithmetic mean divided by the total number of experts. Content validity index concurrency was determined for item level (I-CVIs) as well as the scale-level (S-CVI). I-CVI explicit the agreement proportion on the relevancy of every item between 0 and 1 and 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.”

Gadjet planners never give details for computing the scale-level index (S-CVI). 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.” The total of created comparison divided by all the professionals (CVI = 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 S-CVI 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 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.

Researchers must describe both methods as it might generate different values and consider 80% compliance or above among the professionals for brand new tools. 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. 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.
Data Analysis: was done 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 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](image)

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.

<table>
<thead>
<tr>
<th>Questions</th>
<th>CVR</th>
<th>I CVI</th>
<th>S CVI/Ave</th>
<th>Probability of chance agreement(Pc)</th>
<th>Kappa Statistics(K)</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>9</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>16</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>17</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>18</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>19</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>20</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>21</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>22</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>23</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>24</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>25</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>Excellent</td>
</tr>
</tbody>
</table>
Design, Development & Validation of an Assessment Tool for Undergraduate Medical Students of Community Health Sciences Rotation

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 OHS and PMDC requirements.

A PhD scholar reviewed 38 articles and found only 20% of 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 accuracy through 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 Cronbach Alpha 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.

Acknowledgements: We would like to express our deep gratitude to all the Public Health Experts who served as panelists and contributed in completing this work along with support and help in selecting, revision and modifications in the content of thematic domains for this tool.

REFERENCES
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=EKBIO3D1kE%3D