

Assessment Implementation Measures (AIM): The Journey Towards the Tool Validity

Amara Butt¹, Yawar Hayat Khan² and Muhammad Zeeshan Baig³

¹Department of Medical Education, Rawal Institute of Health Sciences, Islamabad, Pakistan

²Department of Medical Education, RIPAH International University, Islamabad, Pakistan

³Department of Oral and Maxillofacial Surgery, Frontier Medical and Dental College, Abbottabad, Pakistan

ABSTRACT

Objective: To establish the construct validity of the Assessment Implementation Measures (AIM) tool to accurately assess faculty perspectives on implemented assessment systems, facilitating the alignment with set standards.

Study Design: Qualitative Study.

Place and Duration of the Study: The study was carried out at RIPAH International University and data were collected from participants (senior faculty members) involved in teaching and assessment of undergraduates from various medical and dental colleges in Pakistan. Google form questionnaire was distributed using email and WhatsApp. The data were analysed using IBM AMOS SPSS version 24.

Methodology: It was a 30-item tool. The sample size was calculated using 1:10 item-to-participant ratio. Hence data were collected from 313 participants. Confirmatory factor analysis was done to establish construct validity.

Results: The indices of confirmatory factor analysis of the tool showed a discrepancy. To remove this discrepancy, items with weak item-loading values were removed. In the end, the tool was reduced to 13 items belonging to three domains.

Conclusion: The final model was improved by excluding items from the original model. A re-validation study with a careful selection of experienced participants from various institutional backgrounds having baseline knowledge of medical education is suggested to improve the results.

Key Words: AIM Tool, Confirmatory factor analysis, Medical education, Assessment.

How to cite this article: Butt A, Khan YH, Baig MZ. Assessment Implementation Measures (AIM): The Journey Towards the Tool Validity. *J Coll Physicians Surg Pak* 2025; **35(01)**:90-94.

INTRODUCTION

Assessment is crucial in medical education, as it evaluates student capabilities and fosters learning through continuous feedback.¹ Formative assessment is gaining importance over summative assessment due to its ongoing nature. A curriculum must clearly define assessment policies, including fail/pass criteria, promotion/demotion cases, and external scrutiny, to ensure effective preparation by stakeholders.² Additionally, balanced weightage across knowledge, skill, and attitude domains is essential for a comprehensive assessment strategy. Without these elements, the curriculum fails to serve its intended purpose.³

Effective student evaluation requires faculty orientation on institutional assessment policies and procedures.

Faculty development programmes ought to instruct instructors in the use of assessment instruments, the creation of challenging multiple-choice questions, the creation of question banks, and the analysis of post-exam items.⁴ Though many are ignorant of important principles such as the pass/fail criteria and exam retakes, teachers' opinions on assessment are quite important. Their selection of assessment instruments is impacted by this ignorance or ambiguous communication, which may jeopardise the accuracy of the evaluation of students' knowledge and proficiency.

There has been no significant study assessing faculty perceptions of assessment measures in undergraduate medical curricula. The only identified tool for this purpose is the Assessment Implementation Measures (AIM) tool, which evaluates perceptions across four domains: Assessment policies (AP), assessment quality measures (AQM), purpose of assessment (PA), and assessment methods (AM).⁵ These domains include 8, 9, 5, and 8 items, respectively, making up a 30-item tool. Responses are measured on a 5-point Likert scale from strongly agree (score 4) to strongly disagree (score 0). The maximum score of 120 indicates ideal assessment quality as perceived by the respondent, supporting institutional self-evaluation and enhancement of assessment practices.

Correspondence to: Dr. Muhammad Zeeshan Baig, Department of Oral and Maxillofacial Surgery, Frontier Medical and Dental College, Abbottabad, Pakistan
E-mail: 7zeeshanbaig@gmail.com

Received: March 09, 2024; Revised: June 19, 2024;

Accepted: November 12, 2024

DOI: <https://doi.org/10.29271/jcpsp.2025.01.90>

The important factor analysis required for tool validation was not carried out during the AIM tool's initial piloting. To make sure the tool measures what it is supposed to and yields correct findings, validation is crucial. In the absence of this, the gathered data may be skewed and inaccurate.⁶ By using this method to better understand faculty perceptions, assessment quality can be improved and problems with assessment design can be found, facilitating self-evaluation. By establishing the construct validity of the AIM tool using factor analysis, this study aims to close the gap that exists between established assessment standards and real-world undergraduate medical programme practices.⁷

The purpose of this study was to establish construct validity by doing factor analysis of the AIM tool which would enable us to bridge the gap between set standards for assessment and implemented practices in medical undergraduate programmes in each institution. Through the AIM tool, which was developed to evaluate teachers' perspectives, an insight about what are the various factors that contribute to the successful implementation of assessment procedures as well as the issues that hinder the process.

METHODOLOGY

The study commenced after approval from the Institutional Review Board (IRB) under Ref. No. Riphah/IIMC/IRC/22/2010. The study was completed in 7 months. The questionnaire was circulated to the participants through email and WhatsApp using Google forms. Detailed description of this study along with informed consent form was furnished to every participant in the start of the Google form. Anonymity and confidentiality of data were maintained at all levels of this research project. As this was a tool-validation study, written permission was obtained by the author to use and validate the tool. Participants' identity was kept anonymous and only the master researcher had access to the data. It is a tool-validation study, based on the guidelines laid by Association for Medical Education in Europe (AMEE) guide no. 87. Tool validation involves step 7 of the AMEE guide.

The non-probability purposive sampling technique was used. In this type of sampling technique, the researcher himself selects the participants because of their location, position, job specification, education, or socioeconomic status etc. (<https://www.cambridge.org/core/journals/prehospital-and-disaster-medicine/article/population-research-convenience-sampling-strategies/B0D519269C76DB5BFFB84ED7031267>). For this study, all the participants were faculty members involved in undergraduate medical and dental teaching, working in various private and government colleges across Pakistan. Participants comprised of faculty members from basic medical and dental sciences as well as clinical sciences. However, junior lecturers and non-teaching clinical faculty were excluded from the study. The suggested sample size for factor analyses ranges from 50 to over 1,000 samples. Additionally, the recommended item-to-response ratio falls between 1:3 and 1:20.

There were total 30 items and using an average item-to-response ratio of 1:10, the sample size of 300 was calculated.⁸ As this tool is comprised of 30 items, so the sample size was calculated to be 300. Data analysis involved factor analysis of the items which was done using IBM AMOS (analysis of a moment structures) SPSS version 24. AMOS is indeed a plug-in for SPSS that is used for structural equation modelling (SEM), available from: (<https://www.taylorfrancis.com/books/mono/10.4324/9781003018414/applied-structural-equation-modeling-using-amos-joel-collier>).

This software is used to analyse data by doing factor analysis and SEM.⁹ A confirmatory factor analysis (CFA) with maximum likelihood method was conducted. In evaluation of confirmatory factor analysis, absolute and relative fit indices are used. Absolute fit indices determine how well the prior model fits or reproduces data. The principal component analysis (PCA) is used to explore underlying structures. (<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119111931.ch158>).

They include chi-square test, root mean-square error of approximation (RMSEA), goodness of fit index (GFI), adjusted goodness of fit index (AGFI), root mean-square residual (RMR), and standardised root mean-square residual (SRMR). The data analysis process using AMOS involved initial data reduction. It was performed in SPSS by putting in all variables into the data reduction option. Principal components analysis with Varimax rotation was applied to explore underlying structures and more interpretable factor structure in PCA. The results obtained from the rotation matrix were transferred to the pattern matrix within the AMOS model builder interface. Refinement of the path diagram was carried out by incorporating guidance from model fit measures, enhancing the clarity and accuracy of the structural model. Counts and percentages were used to express categorical variables.

RESULTS

Of all the 313 faculty members who submitted completed questionnaires, 159 (50.8%) were males and 154 (49.2%) were females. One hundred and forty-five (46.3%) faculty members had more than 11 years of teaching experience, whereas 93 (29.7%) had 8-11 years, 32 (10.2%) had 4-7 years, and 41 (13.1%) had up to 3 years of teaching experience. Of all the respondents, 59 (18.9%) were senior registrars, 132 (42.2%) were assistant professors, 67 (21.4%) were associate professors, and 53 (17.0%) were professors.

Table I shows indices such as minimum discrepancy function by degrees of freedom divided (CMIN/DF), comparative fit index (CFI), SRMR, root mean square error of approximation (RMSEA), and PClose showed acceptable values for modified AIM tool. The internal consistency of the tool was estimated through Cronbach's Alpha, and it came out to be 0.93 which is in acceptable limit (0.1-1). Figure 1 shows factor analysis of original AIM tool with item loading in 4 domains i.e., (A) assessment methods, (B) purpose of assessment, (C) assessment policies, and (D) assessment quality measures.

Table I: Model fit measures of AIM tool.

Measure	Estimate	Threshold	Interpretation
CMIN/DF	2.463	Between 1 and 3	Excellent
CFI	0.844	>0.95	Terrible
SRMR	0.075	<0.08	Excellent
RMSEA	0.069	<0.06	Acceptable
PClose	0.06	>0.05	Acceptable

Table II: Validity measures of AIM tool.

Measure	CR	AVE	MSV	MaxR (H)	A	B	C	D
A	0.845	0.382	0.580	0.854	0.618			
B	0.859	0.360	0.582	0.869	0.763***	0.601		
C	0.836	0.470	0.339	0.875	0.582***	0.576***	0.685	
D	0.808	0.530	0.339	0.885	0.583***	0.567***	0.410***	0.728

Table III: Model fit measures of assessment methods.

Measure	Estimate	Threshold	Interpretation
CMIN/DF	2.394	Between 1 and 3	Excellent
CFI	0.925	>0.95	Excellent
SRMR	0.058	<0.08	Excellent
RMSEA	0.067	<0.06	Acceptable
PClose	0.111	>0.05	Excellent

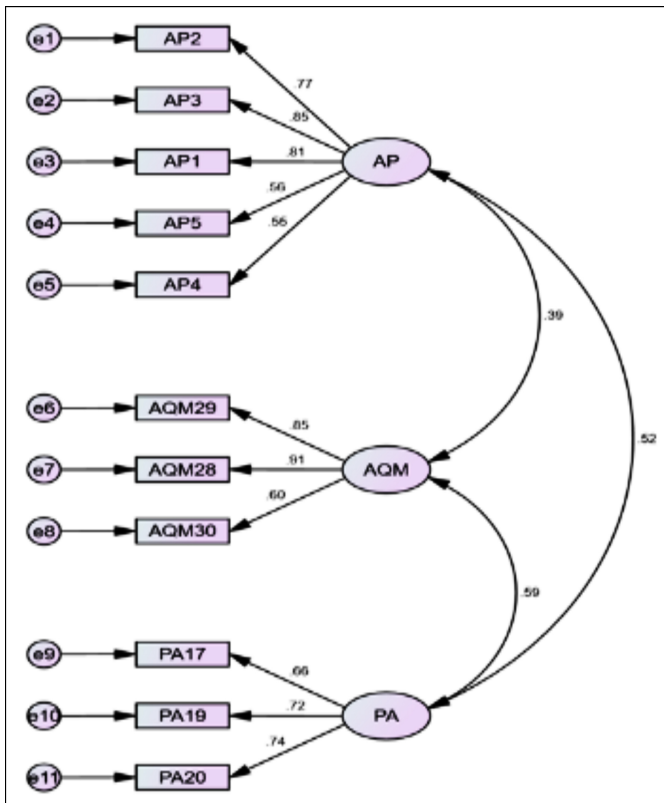


Figure 1: Factor analysis of original AIM tool with item loading in four domains.

Composite reliability (CR) and average variance extracted (AVE) which are measures of internal reliability and consistency are presented in Table II. AVE of each domain or construct must be at least 0.5. The value of AVE in original tool for domain A, B, and C showed values less than 0.5. The tool was modified using AMOS software and items with low item-loading were removed and the AVE was improved.

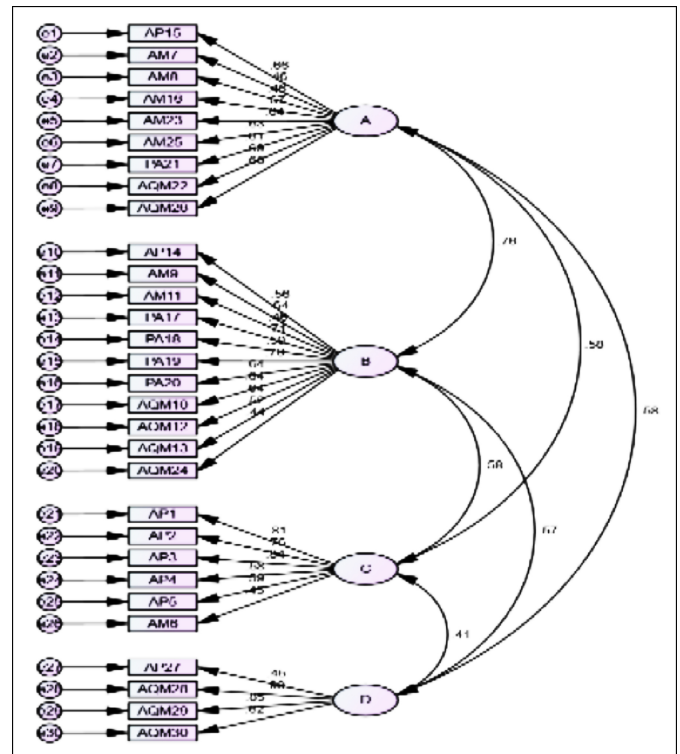


Figure 2: The factor analysis of modified AIM tool.

As a result, the domain of assessment methods was removed. Items with low item-loading were removed from the remaining three domains. Modified tool with good AVE is shown in Table II and model fit measures is shown in Table III. The factor analysis of the modified AIM tool is shown in Figure 2.

DISCUSSION

The main objective of the study was to validate AIM tool that can be used to assess the perspectives of teachers on

prevailing assessment system in their institutions. The results showed that out of 30 items in four domains, only 13 items in three domains could be validated. AMOS software designed for SEM, path analysis, and, CFA, provided proof of item-loading into each domain and eliminated the items with poor item-loading. In CFA, the researcher specifies the proposed relationships among latent factors and their observed indicators using a SEM framework. (<https://www.tandfonline.com/doi/abs/10.1080/00273171.2019.1602503>). It defines the relationship between latent factors (items) and observed variables (indicator loadings), as well as the covariance among latent factors. Fit indices are used to evaluate how well the hypothesised model fits the observed data. Common fit indices include the Chi-square statistic, CFI, and SRMR.¹⁰ Lower values of RMSEA and SRMR and higher values of CFI and TLI, indicate better fit model. If the initial model fit is unsatisfactory, modifications may be necessary to improve fit. This could involve adding or removing paths, allowing correlated errors between indicators, or re-specifying the model based on theoretical considerations. Here, the items with lower factor loading were removed to improve the values.¹¹

This tool can be used by teachers, college administration, and examination authority for assessing the quality of assessments and comparing it with the international standards. (https://dro.deakin.edu.au/articles/journal_contribution/Challenges_in_reforming_higher_education_assessment_a_perspective_from_afar/20691955/1). There are many studies reported in literature that measure medical undergraduate students' perceptions regarding assessment in various aspects such as tools of assessment, grading system, and level of difficulty (<https://asmepublications.onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2923.2002.01291.x>) but the perceptions of teachers related to various aspects of assessment are not explored in depth. Sajjad *et al.* constructed assessment implementation measures tool for assessing faculty knowledge and perspectives about assessment practices in their institutions compared to national and international standards.⁵ This tool was developed using a mixed-method approach. Literature review was done to make a primary questionnaire and was given to 10 medical educationists for three rounds modified-delphi technique. Panel agreement of $\geq 75\%$ was considered for inclusion of items. Cognitive pre-testing and later piloting was done by randomly selecting 30 faculty members. The value of Cronbach's alpha was estimated to be 0.93 which is a measure of internal consistency and reliability of the tool (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4205511/>).

Internal consistency pertains to how closely all items within a test align with the same underlying concept or construct, thus reflecting the inter-relatedness of the items within the test or tool.¹² The original AIM tool covers assessment policies, assessment purpose, assessment methods, and assessment quality measures. The domain of assessment methods was removed during factor analysis which led to an incidental

finding that the responses to various items in that domain needed knowledge of health professionals education. Lack of knowledge of faculty members who did not have any medical education background led to faulty responses in that domain that led to lower values of covariance and hence their item loading was also not up to the mark. (<https://www.science-direct.com/science/article/abs/pii/S000296102100341X>).

The main strength of this study was to establish construct validity and reliability of AIM tool on the basis of an empirical approach with SEM, which is a popular technique for establishing the validity and reliability of the instrument. Another strength of this study was to engage faculty members from varying backgrounds and experiences, i.e., from lecturers, assistant professors, associate professors, and professors to heads of institutions. However, this study is not without limitations. The sample of the study was 311 cases collected from different designations which had different exposure about current assessment practices and looked at assessment with different lenses. A more in-depth study with a greater number of faculty members, especially with health professions education background should be done for better results.

CONCLUSION

Knowing the faculty's perception about standards for the assessment and their state of implementation in the undergraduate medical curriculum helps the institutions to improve their teaching and assessment strategies. Factor-loading indicated the need to remove many items whose model-fit values were not up to the mark. Hence the final modified model with all best indices values showed only three dimensions i.e., assessment policies (AP) with five items, Assessment quality measures (AQM) with three items, and purpose of assessment (PA) with three items. In this model, assessment methods (AM) were removed.

ETHICAL APPROVAL:

The study commenced after approval from the Institutional Review Board (IRB) under Ref. No. Riphah/IIMC/IRC/22/2010.

PARTICIPANTS' CONSENT:

Informed consent was taken from all the participants included in the study.

COMPETING INTEREST:

The authors declared no conflict of interest.

AUTHORS' CONTRIBUTION:

AB: Conceived and designed, data collection, manuscript drafting and editing.

YHK: Critical reviewing and editing.

MZB: Data analysis and correspondence.

All authors approved the final version of the manuscript to be published.

REFERENCES

1. Singh T, Gupta P, Dhir SK. Understanding clinical competence: Understanding student assessment. *Indian Pediatr* 2023; **60(4)**:267-71.
2. Naeem NK, Hadie SNH, Ismail IM, Ullah S, Yusoff MSB. Buzz or fuss: Gauging the online learning environments in undergraduate medical education – A mixed-method study. *Educ Med J* 2024; **16(2)**:63-83. doi: 10.21315/eimj2024.16.2.5.
3. Alkhateeb NE. Tools of formative assessment. In: Zaidi SH, Hassan S, Bigdeli S, Zehra T, Eds. *Global medical education in normal and challenging times. Advances in science, technology & innovation.* Cham: Springer; 2024. doi: 10.1007/978-3-031-512445_17.
4. Garcia-Ros R, Ruescas-Nicolau MA, Cezon-Serrano N, Flor-Rufino C, Martin-Valenzuela CS, Sanchez-Sanchez ML. Improving assessment of procedural skills in health sciences education: A validation study of a rubrics system in neurophysiotherapy. *BMC Psychol* 2024; **12(1)**:147. doi: 10.1186/s40359-024-01643-7.
5. Sajjad M, Khan RA, Yasmeen R. Measuring assessment standards in undergraduate medical programs: Development and validation of AIM tool. *Pak J Med Sci* 2018; **34(1)**:164-9. doi: 10.12669/pjms.341.14354.
6. Aslam K, Khan RA, Aslam MA, Zaidi FZ. Curriculum implementation challenges: Development and validation of an integrated curriculum implementation challenges tool. *Pak J Med Sci* 2024; **40(1 Part-I)**:89-94. doi: 10.12669/pjms.40.1.7258.
7. Khan HF, Yasmeen R, Ali S. Validation of an assessment tool for professionalism in medical students: A mixed method study. *J Pak Med Assoc* 2023; **73(11)**:2177-82. doi: 10.47391/JPMA.7770.
8. Yilmaz Kogar E, Kogar H. A systematic review and meta-analytic confirmatory factor analysis of the perceived stress scale (PSS-10 and PSS-14). *Stress Health* 2024; **40(1)**:e3285. doi: 10.1002/smi.3285.
9. Gogia EH, Shao Z, Khan K, Rehman MZ, Haddad H, Al-Ramahi NM. Exploring the relationship of organizational virtuousness, citizenship behavior, job performance, and combatting ostracism through structural equation modeling. *BMC Psychol* 2024; **12(1)**:384. doi: 10.1186/s40359-024-01873-9.
10. Byrne BM. *Structural equation modeling with AMOS: Basic concepts, applications and programming.* ed. 3rd, San Diego; Taylor and Francis/Routledge; 2016. doi: 10.4324/9781315757421.
11. Liao Y, Yang J. Status of nutrition labeling knowledge, attitude, and practice (KAP) of residents in the community and structural equation modeling analysis. *Front Nutr* 2023; **10**:1097562. doi: 10.3389/fnut.2023.1097562.
12. Raheel H, Naeem NJPMA. Assessing the objective structured clinical examination: Saudi family medicine undergraduate medical students' perceptions of the tool. *J Pak Med Assoc* 2013; **63(10)**:1281-4.

