Open Access

Original Article

Developing an Instrument to Assess Teaching Skills of Lab Instructors in Medical Education: Initial Validation

Sadia Ahsin, Muhammad Iqbal*, Hira Ashraf

Department of Physiology, Foundation University Islamabad Pakistan, *Department of Surgery, Shifa Tameer-e-Millat University Islamabad Pakistan

ABSTRACT

Objective: To devise and validate a direct observation tool for assessing lab teaching skills of laboratory instructors namely 'Direct Observation of Teaching Skills' (DOTS)

Study Design: Mixed method study

Place and Duration of Study: Foundation University Islamabad and Shifa Tameer-e-Milat University Islamabad, Pakistan from Apr 2021 to Apr 2022.

Methodology: More than three hundred and fifty qualified medical educationists who were registered with Pakistan Medical and Dental Council and having five years and above teaching experience participated through purposive sampling. Tool development was done in systematic seven recommended steps of AMEE Guide 87 including 1) Literature Search, 2) Focus group discussion, n= 7, 3) Synthesis of literature review and focus group discussions, 4) Item development, 5) Expert Validation, 6) Cognitive pretesting n= 5 and 7) Pilot testing n= 310. SPSS and AMOS were used for data analysis.

Results: A 29-item tool evaluating laboratory teacher performance including laboratory instructions, activities and management was validated for content, response process and construct with Scale Content Validity Index/ Average (S-CVI) of 0.867 and S-CVI Universal agreement of 0.989. Confirmatory analysis showed factor loading of > 0.50 for all items. The absolute fit value of Observed normed $\chi 2$ was 2.99, goodness of fit index was 0.79 and root mean square error of approximation was 0.08. Internal consistency for lab instruction, lab activities and lab management showed Cronbach's α values of 0.909, 0.899, 0.912 respectively.

Conclusion: The devised tool, DOTS appears reliable and valid in content, response process and construct.

Keywords: Delphi's method, factor analysis, Laboratory teaching evaluation, validity

How to Cite This Article: *Ahsin S, Iqbal M, Ashraf H. Developing an Instrument To Assess Teaching Skills of Lab Instructors In Medical Education: Initial Validation. Pak Armed Forces Med J 2025; 75(3): 538-544. DOI: https://doi.org/10.51253/pafmj.v75i3.10610*

INTRODUCTION

Various approaches and guidelines are recommended for teacher evaluation through classroom observations. The fundamental concepts of these methods and guidelines can be applied in evaluation of teaching skills in various disciplines.[1-3] The available in-depth evaluation tools used by supervisors and principals are reliable and have been validated for evaluation of those teachers' performance who are involved in classroom teaching.[1,3] These tools are not directly applicable in laboratory teaching set up; however, it has been suggested that basic concepts of these tools can be adjusted as per the requirement of settings where teaching evaluation is being done.[1] Here it is proposed that for evaluation of teachers who are exclusively involved in laboratory teaching, an observation tool based on available guidelines, with tangible and focused performance expectations needs to be designed. This study was conducted to develop and validate a comprehensive lab teaching assessment tool, by "Direct Observation of Teaching Skills" (DOTS) which can be used to assess, guide, and improve teaching skills of laboratory instructors.

Correspondence: Dr Sadia Ahsin, Department of Physiology, Foundation University Islamabad Pakistan

Received: 12 Jul 2023; revision received: 05 Apr 2024; accepted: 01 Aug 2024

METHODOLOGY

This mixed method study was carried out at Foundation University Islamabad and Shifa Tameer-e-Milat University Islamabad, Pakistan over a period of one year from April 2021-April 2022 after approval by Ethical Review Board Shifa Tameer-e Milat University (vide ref no "IRB # 079-21) and Foundation University Ethical Review Committee (vide ref no "FF/FUMC/215-113/Phy-21")

Inclusion Criteria: Pakistan Medical and Dental Council registered medical educationists with minimum 5 years of teaching experience were included.

Exclusion Criteria: Recent graduates and health professionals without medical teaching experience were excluded from the study.

Following 7 systematic steps/phases as per AMEE Guide 87: Developing questionnaires for educational research were adopted. [4]

Phase 1: Literature Review

Following BEME guidelines, articles published in databases including Google Scholar, PubMed, Eric, and Pakmedinet, in last fifteen years were retrieved. Three open ended questions covering themes of effective lab instruction delivery, lab teaching activities/strategies and lab management skills were prepared for phase 2 of the study.

Phase 2: Focus Group Discussion (FGD)

Seven experts participated in FGD with the aim of developing statements that were in accordance with current study context. Pre- prepared open-ended questions derived from first phase were emailed to these participants to deliberate upon before FGD. Discussion continued till participants response saturation was reached. Responses were recorded. Manual data transcription and concept-driven open coding of text statements was done.[5] Statements along with codes were then arranged under predefined item constructs. Data triangulation was done by getting codes, statements and constructs rechecked by the participants as well as an independent expert researcher in medical education.

Phase 3: Combining Literature Review and FGD

A comprehensive list included statements which were a blended replication of literature review and views of FGD participants was prepared.

Phase 4: Item Development

Under three constructs including Lab instructions, Lab activities and Lab management, 30 items were developed. These were given codes LI, LA and LM. Each item belonging to each construct was given an item code, with 11 items in LI starting from LI1, LI2, LI3 to LI11, 8 items in LA starting from LA1, LA2, LA3 to LA8 and 11 items in LM from LM1, LM2, LM3 to LM11. The end user response to the closed ended statements would be on Likert scale for performance evaluation where rating would be done as follows: Not done = 1, Partially done = 2, Reasonably done = 3, Well done = 4, Outstandingly done = 5.

Phase 5: Expert Validation

Modified Delphi technique in 2 rounds from experienced content specialists was used.[6]

Delphi Round 1

Twenty medical education experts through non-probability purposive sampling were individually requested to rate each item according to its relevancy to the construct that it measured, where 5= Extremely Relevant (ER), 4= Moderately Relevant (MR), 3= Averagely Relevant (AR), 2= Slightly Relevant (SR), 1= Not Relevant (NR). An option after each item was also given where experts could give their opinion regarding changing/improving or deleting the statement. Responses from 15 experts were received. Calculations were performed on the percentage replies, median, content validity scale, and content validity index of the items. The remarks regarding each item statements were also analyzed. According to the rule, items having I-CVI ≤0.70 are removed, items with I-CVI range 0.70 - 0.90 are revised and item with I-CVI ≥0.90 are included.6 Since I-CVI values of all the items was > 0.90 and thus all were included. Additionally, median response values and percentage agreement of item relevance were computed. Following item statement revisions, a second round of the instrument containing 30 items was delivered.

Delphi Round 2

The revised tool was sent to those 15 experts who responded in the Delphi round 1. In this round each expert could also see the responses of the fellow Delphi subjects. The experts were informed to indicate each item based on how necessary it was in measuring the given construct. To make sure that the instrument had accurate and crucial content, the content validity ratio (CVR) was calculated. The content validity ratio (CVR) was measured to ensure that correct and most important content was included in the instrument.[7,8] The panelists were asked to rate the items of tools on the following scale: Necessary, Useful but not much necessary or Unnecessary. Items having a CVR range from 0.60 to 0.80 were amended after the responses were examined. Not one had a value that was below 0.60 that allowed it to be taken out of the tool.

Phase 6: Cognitive Pretesting

For response process validity evidence, 5 participants were individually interviewed through verbal probing method10. Appropriate response from predefined coding criteria from 1 to 7, 1 being clear and requiring no change, to 7 requiring removal was recorded.[9]

Phase 7: Pilot testing

To ensure construct validity and tool reliability a sample size of 348 medical teachers in basic and preclinical sciences was calculated at 95% confidence

interval with 5% margin of error using Raosoft sample size calculator.[10] Tool was emailed to participants with a set of instructions. Each participant was requested to arrange appropriate time to observe lab teaching practice of one of their junior colleagues in a collaborative and non-threatening environment while maintaining confidentiality of the teacher being observed.

Confirmatory factor analysis was done with AMOS. To establish internal consistency reliability of tool Cronbach alpha value was calculated on SPSS. Pearson's correlation was applied determining correlation among study constructs

RESULTS

Phase 1-4:

At the end of phase 4, a set of 30 items was generated. The broad areas identified for teacher observation included Lab instruction method (LI), lab activities (LA) and lab management (LM). These were heavily aligned with the concepts of Charlotte Danielson's Enhancing Professional Practice: A Framework for Teaching, and University of Virginia's Classroom Assessment Scoring System (CLASS) scales.[11-12] The general guidelines to provide effective feedback to teachers were in alignment with the US non-profit organization 'The New Teacher Project (TNTP)' before developing themes and items for tool.

Phase 5:

Delphi Round 1

Based on 15 experts' comments, item LI1, LI3, LI4, LI8, LI10, LA2, LA3, LM1, LM3, LM7 were revised. Since I-CVI values of all the items was > 0.90 therefore all were included. Percentage agreement of items relevance and median values of responses were also calculated. Excellent cut-off values for S-CVI/UA and S-CVI/Ave are deemed to be ≥ 0.80 and ≥ 0.90, respectively.[13] The scale content validity index/universal agreement (S-CVI/UA) was 0.984, and the scale content validity index/average (S-CVI/Ave) was 0.833.

Delphi Round 2

Out of 15 experts, 12 responded for round 2. LM10 was shifted to LA as per respondents' suggestions. Revised scale had 11 Items in LI, 9 in LA and 10 in LM domains. According to the rule, items having I-CVI ≤0.70 are removed, items with I-CVI range 0.70 - 0.90 are revised and item with I-CVI ≥ 0.90 are included.13 All of the items' I-CVI values were greater than 0.90, so they were included. Also computed were the median values of the responses and the percentage agreement of the relevance of the items. 0.80 and 0.90 are considered excellent cut-off values for S-CVI/UA and S-CVI/Ave, respectively.[14] The results of the scale content validity index/universal agreement (S-CVI/UA) and scale content validity index/average (S-CVI/Ave) were 0.989 and 0.867, respectively

Phase 6:

Five medical teachers participated and responded according to the predefined codes. For statements where 3 or more out of 5 participants required explanation were rephrased. 4 LI8 was removed because none of the participants found it relevant. LI3, LI10, LA4, LA5, LA7, LA9, LM2, LM&, LM8, LM10 were rephrased. A total of 29 items were finalized. (Table I)

Phase 7:

In pilot testing 310 participants responded out of 348. Datasheets generated on excel were imported on SPSS and AMOS to determine tool reliability and construct validity via Cronbach alpha value and confirmatory factor analysis. Items with a factor loading of 0.5 are retained, according to the rule.14 Since factor loading of all 29 items was > 0.5 therefore all were retained in the tool (Table II). All the fit indices of the measurement model were in acceptable range (Table III) after one modification in error terms. "The researcher can impose a variety of constraints on the model to improve fit because of SEM's flexibility".[15] The correlation of error terms in measurement model represented that the items of the construct measure something in common. The modification is applied for two items of "Lab Activities" (see Figure I).

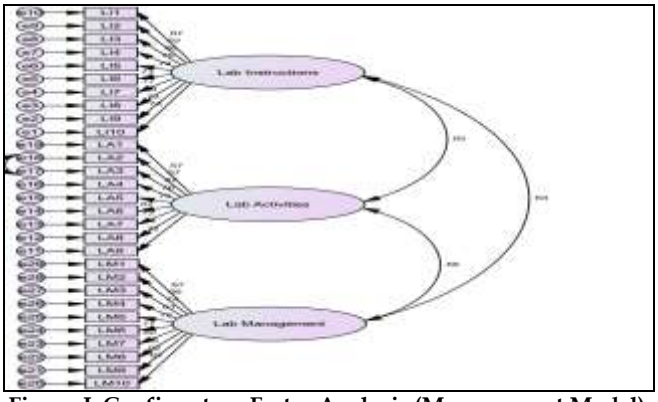

Figure-I: Confirmatory Factor Analysis (Measurement Model)

Table I: Instrument: Direct observation of teaching skills

Theme	Item Code	Statements	1= Not done	2= Partially done	3= Reasonably done	4= Well done	5= Outstandingly done
Lab Instructions	LI1	Defines specified lab objectives.					
	LI2	Explains the learning outcomes clearly.					
	LI3	Correlates the basic principle of specified practical with theory.					
	LI4	Activates students' prior knowledge regarding the specified practical.					
	LI5	Communicates key concepts accurately.					
	LI6	Communicates relevant content.					
	LI7	Delivers procedural steps in a systematic manner.					
	LI8	Ensures student learning by repeating and questioning.					
	LI9	Encourages students active participation.					
	LI10	Speaks in loud, clear, and polite tone.					
Lab Activities/Strategies	LA1	Takes lab safety measures e.g gloves, lab coat etc.					
	LA2	Demonstrates the working and parts of apparatus/ equipment/specimen to students.					
	LA3	Demonstrates procedure accurately.					
	LA4	Demonstrates precautionary measures while performing procedure					
	LA5	Highlights key findings and relates them to theory.					
	LA6	Intermittently checks for students understanding by asking questions.					
	LA7	Interprets the results obtained					
	LA8	Summarizes the laboratory procedure.					
	LA9	Identifies and clarifies misconceptions related to lab procedure					
Lab Management	LM1	Confirms the availability and functioning of the relevant equipment/apparatus before demonstration					
	LM2	Collaborates well with co workers					
	LM3	Appropriately utilizes lab class time.					
	LM4	Ensures that students take safety measures during lab.					
	LM5	Reinforces positive behavior.					
	LM6	Maintains lab discipline in non-threatening collaborative environment.					
	LM7	Observes and guides students during lab work/performance					
	LM8	Checks students learning by walking around and asking relevant questions					
	LM9	Discusses student's lab results with them.					
	LM10	Reviews lab objectives with the students to ensure learning					

The correlation among study constructs including Lab Instructions, Lab activities and Lab Management skills using Pearson's Correlation was statistically significant with *p*-values of < 0.01(Table IV)

Reliability of the Tool

With SPSS version 21, the tool's internal consistency was assessed. When Cronbach's alpha is greater than 0.8, it indicates that a tool has good internal consistency. The values calculated are shown in Table-V

DISCUSSION

Using AMEE Guide No. 87 as a guide, a methodical seven-step process was used to create and validate a questionnaire designed to evaluate lab teacher performance in medical education. A 29-item tool was refined through focus groups, the Delphi method, literature review, and cognitive pretesting; the results showed high content validity (S-CVI/Ave = 0.867, S-CVI/UA = 0.989). The construct validity of the tool was confirmed through pilot testing involving 310 observations. All items demonstrated good factor loading (>0.5) and excellent internal consistency

(α=0.909–0.912) across three dimensions: management, activities, and lab instruction.

Table-II: Item's Factor Loadings and Construct's Reliability and Validity

Constructs	Items	Factor Loadings	Composite Reliability	Convergent Validity (AVE)*
Lab Instructions	LI1	0.669	0.967	0.745
	LI2	0.618		
	LI3	0.664		
	LI4	0.751		
	LI5	0.735		
	LI6	0.726		
	LI7	0.773		
	LI8	0.682		
	LI9	0.748		
	LI10	0.731		
Lab Activities	LA1	0.573	0.956	0.706
	LA2	0.665		
	LA3	0.674		
	LA4	0.7		
	LA5	0.746		
	LA6	0.669		
	LA7	0.725		
	LA8	0.712		
	LA9	0.716		
Lab Management	LM1	0.669	0.97	0.768
	LM2	0.657		
	LM3	0.74		
	LM4	0.614		
	LM5	0.722		
	LM6	0.767		
	LM7	0.799		
	LM8	0.729		
	LM9	0.816		
	LM10	0.669		

**AVE – Average Variance Extracted*

An objective validated assessment tool can help novice teachers in identifying his/her strengths and weaknesses.[16,17] Three distinct approaches were used in a study by Urrutia et al. to assess teaching effectiveness during the first two years of medical school. In their study, one of the methods they used to assess teaching performance was a validated instrument developed by the university Educational Evaluation Department (Dirección General de Evaluación Educativa), consisting of 30 items. Like current tool, their questionnaire also had a Cronbach's α of 0.93; every statement had five possible Likert-type responses (1=never, 2=seldom, 3=sometimes, 4=often, and 5=always). The tool took into account three aspects of teaching: teaching strategies, student respect, and learning assessment. This tool is comparable to our current tool in terms of item count, though responses for teaching performance are also collected on a 5-point Likert scale, but participants were students.

Table-III: Measurement Model Fit Indices

Fit Indices		Recommended Cut-off value	Measurement Model
Absolute Fit Measures	Observed Normed χ2 (CMIN/df)	>= 5 The smaller, the better	2.99
	Goodness of Fit Index (GFI)	> = 0.8 Near to 1	0.791
	Root Mean Square Error of Approximation (RMSEA)	> = 0.1 < = 0.08	0.080
Incremental Fit Measures	normed fit index (NFI)	> = 0.08 Near to 1	0.805
	Relative Fit Index (RFI)	Near to 1 (Higher the better)	0.788
	Incremental Fit Index (IFI)	Near to 1 (Higher the better)	0.861
	Tucker-Lewis Index (TLI)	Near to 1 (Higher the better)	0.848
	Comparative Fit Index (CFI)	Near to 1 (Higher the better)	0.860
	Adjusted Goodness of Fit (AGFI)	> = 0.8 Near to 1	0.756
Parsimonious Fit Measures	Parsimonious Normed Fit Index (PNFI)	The higher the better	0.740

Table-IV: Correlation among Study Constructs

Correlations				
		Lab Instructions	Lab Activities	Lab Management
Lab Instructions	Pearson Correlation	1	.583**	.572**
	Sig. (2-tailed)		.000	.000
	N	310	310	310
Lab Activities	Pearson Correlation	.583**	1	.803**
	Sig. (2-tailed)	.000		.000
	N	310	310	310
Lab Management	Pearson Correlation	.572**	.803**	1
	Sig. (2-tailed)	.000	.000	
	N	310	310	310

***Correlation is significant at the 0.01 level (2-tailed)*

There is no exact or strict number of Delphi judges claimed in literature for tool validation studies, however six to ten judges often suffice, but to achieve unanimous agreement, up to twenty experts have been recommended.[4] Therefore, to ensure clearer consensus we enrolled 20 experts for round 1, although 15 responded.

According to published research, in order to assess the measurement model's goodness of fit, each

construct should have three to four items. After one adjustment in error terms, all of the measurement model's fit indices fell within an acceptable range. In their study, Ahmed A. A. and colleagues evaluated the validity of the modified System for Evaluation of Teaching Qualities (mSETQ) for assessing clinical teachers in Bahrain. They also used these fit indices. The factor loadings, or standardized regression weights, were displayed in their findings, and every factor loading value was greater than 0.40.[18] This value was even better in our study—it was > 0.5 for every item.

Table-V: Reliability Analysis

Constructs	Cronbach Alpha
Lab Instructions	0.909
Lab Activities	0.889
Lab Management	0.912

This instrument is novel because it takes a comprehensive approach to assessing laboratory teaching performance, including components that are essential and advised by national and international educational institutions. Along with teaching strategies, it places a strong emphasis on student participation, safety precautions, and the demonstration of laboratory techniques. Its validation procedure also guarantees validity and reliability, which makes it a useful tool for enhancing the efficacy of instruction in medical science labs.

Novice teachers in medical science labs can benefit from the validated instrument if used by senior teachers, supervisors, or peers to assess laboratory teaching performance. By offering unbiased feedback, its collaborative application would promote skill development and improve the efficacy of instruction for aspiring medical educators.

LIMITATION OF STUDY

Standard setting using cut-off scores could have been done but since the AMEE guide 87 followed for this study did not include it therefore it was left.

CONCLUSION

Following the seven-step process recommended in AMEE Guide 87, our study results concluded DOTS to be reliable and valid in content, construct and response process.

Conflict of Interest: None.

Funding Source: None.

Authors' Contribution

Following authors have made substantial contributions to the manuscript as under:

SA & MI: Data acquisition, data analysis, critical review, approval of the final version to be published.

HA: Study design, data interpretation, drafting the manuscript, critical review, approval of the final version to be published.

Authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

1. Pastore S. Teacher assessment literacy: a systematic review. Front Educ 2023; 8: 1217167. https://doi.org/10.3389/feduc.2023.1217167
2. Wiggs NB, Reddy LA, Kettler R, Hua A, Dudek C, Lekwa A, et al. Convergence between teacher self-report and school administrator observation ratings using the Classroom Strategies Assessment System. J Psychoeduc Assess 2023; 41(1): 162–177. http://doi.org/10.1177/15345084221112858
3. Unissa R, Alhasan NA. Investigating educators' perspectives on classroom observation and the impact of feedback on professional development. Innovare J Educ 2024; 12(6): 1–6. Available from: https://journals.innovareacademics.in/index.php/ijoe/article/view/52896
4. Artino Jr AR, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. Med Teach 2014; 36(6): 463-474. https://doi.org/10.3109/0142159x.2014.889814
5. Gibbs G. The nature of qualitative analysis. Analyzing qualitative data. 2007: 1-9.
6. Morgado FF, Meireles JF, Neves CM, Amaral A, Ferreira ME. Scale development: ten main limitations and recommendations to improve future research practices. Psicologia: Reflexão e Crítica 2017; 30. https://doi.org/10.1186/s41155-016-0057-1
7. Hsu, C. and Sandford, B. The Delphi Technique: Making Sense of Consensus. Prac Asses Res Eval 2007; 12: 1-8.
8. Suryadi T, Alfiya F, Yusuf M, Indah R, Hidayat T, Kulsum K. Content validity for the research instrument regarding teaching methods of the basic principles of bioethics. J Pendidik Kedokter Indones 2023; 12(2): 186–202. http://doi.org/10.22146/jpki.77062
9. Linneberg MS, Korsgaard S. Coding qualitative data: A synthesis guiding the novice. Qualitative Res J 2019. https://doi.org/10.1108/QRJ-12-2018-0012
10. Moreno-Murcia J, Torregrosa YS, Pedreo NB. Questionnaire evaluating teaching competencies in the university environment. Evaluation of teaching competencies in the university. J New Approaches Edu Res 2015; 4(1): 54-61. https://doi.org/10.7821/naer.2015.1.106
11. Felland N. Charlotte Danielson's enhancing professional practice: A framework for teaching. Perspect Learn 2001; 2(1): 16.
12. Couchenour D, Chrisman JK. The SAGE encyclopedia of contemporary early childhood education: SAGE Publications; 2016. http://doi.org/10.4135/9781483340333.n138
13. Yusoff MSB. ABC of content validation and content validity index calculation. Edu Med J 2019; 11(2): 49-54. https://doi.org/10.21315/eimj2019.11.2.6

14. Deniz MS, Alsaffar AA. Assessing the validity and reliability of a questionnaire on dietary fibre-related knowledge in a Turkish student population. J Health Popul Nutr 2013; 31(4): 497. https://doi.org/10.3329/jhpn.v31i4.20048
15. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods Psych Res Online 2003; 8(2): 23-74. https://doi.org/10.23668/psycharchives.12784
16. Smith J. An objective structured teaching exercise (OSTE) for physicians employing multi-source feedback. MedEd PORTAL 2015; 11. https://doi.org/10.15766/mep_2374-8265.10157
17. Cladera M. An application of importance-performance analysis to students' evaluation of teaching. Educ Asses Eval Acc 2021; 33(4): 701-715. https://link.springer.com/article/10.1007/s11092-020-09338-4
18. Al Ansari A, Arekat MR, Salem AH. Validating the modified System for Evaluation of Teaching Qualities: a teaching quality assessment instrument. Adv Med Educ Pract 2018; 9: 881–886. https://doi:10.2147/AMEP.S181094