Team Objective Structured Bedside Assessment (TOSBA) in Paediatric Undergraduate Students

Heena Rais[1], Azam Saeed Afzal[2], Amber Shamim Sultan[3] and Farhana Zafar[1]

[1]*Department of Paediatrics, Ziauddin Medical University Hospital, Karachi, Pakistan*
[2]*Department of Community Health Sciences, The Aga Khan University, Karachi, Pakistan*
[3]*Department for Educational Development and Department of Surgery, The Aga Khan University, Karachi, Pakistan*

ABSTRACT

Objective: To develop a team objective structured bedside assessment (TOSBA) tool for assessing the paediatric undergraduate students' clinical skills at the bedside.

Study Design: A validation, cross-sectional study.

Place and Duration of the Study: Department of Paediatrics, Ziauddin Medical University Hospital, Karachi, Pakistan, from March to June, 2023.

Methodology: Three groups of 4th-year students (n = 61) at Ziauddin Medical University underwent this formative assessment process 4-5 times during a 4-week rotation. Psychometric analysis for reliability was done by calculating the internal consistency and item-total correlation of TOSBA scores. At the same time, validity was determined by correlating TOSBA scores with the end-of-rotation objective structured clinical evaluation (OSCE) scores.

Results: The TOSBA tool's overall reliability was good, with Cronbach's alpha values >0.7 for all clinical skills assessed. Spearman's correlation revealed r^2 >0.4 (p <0.001) for history taking and clinical reasoning. Item-total correlation varied across stations, occasionally falling below 0.7. Strong correlations (0.46 to 0.73, p <0.001) were observed between similar constructs. The multi-trait-multi-method matrix highlighted divergent validity, showing no or negative correlations within the same method (TOSBA or OSCE), except for physical examination, which differed from OSCE.

Conclusion: The TOSBA tool developed for formative assessment of undergraduate paediatric students is a reliable and moderately valid tool for formative assessment of undergraduate students.

Key Words: *Objective structured clinical evaluation, Reliability, Team objective structured bedside assessment, Undergraduate students, Validity*.

How to cite this article: Rais H, Afzal AS, Sultan AS, Zafar F. Team Objective Structured Bedside Assessment (TOSBA) in Paediatric Undergraduate Students. *J Coll Physicians Surg Pak* 2025; **35(07)**:897-903.

INTRODUCTION

Undergraduate medical education is anchored on a medical clerkship. It is the desire of incoming students to obtain as much clinical experience as possible.[1] Effective bedside teaching (BST) is important to teach about physical examination, communication skills, critical thinking, problem-solving, decision-making, ethics, and professionalism.[2] Teachers are adopting evidence-based methods for clinical teaching with effective feedback, for example, a one-minute preceptor, SNAPPS (Summarise/Narrow/Analyse/Probe/Plan/Select) in the outpatient department and MiPLAN in the ward setting.[3]

Correspondence to: Dr. Heena Rais, Department of Paediatrics, Ziauddin Medical University Hospital, Karachi, Pakistan
E-mail: heena.rais@zu.edu.pk

Received: September 28, 2024; Revised: February 20, 2025; Accepted: May 08, 2025
DOI: https://doi.org/10.29271/jcpsp.2025.07.897

The formative assessment facilitates learning through a continuous feedback process and, at the same time, provides the opportunity to improve.[4] Workplace-based formative assessment in the form of mini clinical evaluation exercise (Mini-CEX), a single student at a time,[5] and team-based assessment in the form of team objective structured bedside assessment (TOSBA) are utilised as both teaching and assessment tools.[6]

TOSBA is a modified team objective structured clinical evaluation (TOSCE) in which students perform on an actual patient, not a standardised one. The TOSBA evaluation focuses on team diagnostic and patient management abilities, history taking, examination, and clinical reasoning and feedback after each activity.[6]Undergraduate students are getting low scores in the end-of-clerkship Objective Structured Clinical Evaluation (OSCE) of paediatrics at University despite scheduled bedside teaching. Stakeholders, including students, faculty, and academic coordinators, are dissatisfied with their clinical exam performance. According to student comments, despite having regular clinical teaching sessions in inpatient and outpatient settings, students are unable to perform under observation and

obtain personalised feedback because of congested clinics and a lack of faculty time. This factor contributed to the decline of BST.[7]

Currently, high-stakes tests are used to evaluate student learning, but whether a student has truly learned and can apply what they have learned does not receive the importance it deserves in Pakistani medical education. Formative assessment is recommended as an essential component of the curriculum to ensure that the goals of the evaluation are in line with the outcomes of the medical education that the students are receiving. The moment has come to give medical students a chance, where faculty members may watch them perform and give them input on how to get better at clinical skills.

Many of the problems associated with clinical education are addressed by TOSBA, such as the shifting of clinical instruction away from the bedside, the focus on symptoms, and the omission of communication, and problem-solving techniques, which result in students' subpar performance. Its validity and reliability, however, have been the subject of relatively little research. One was conducted by Miller *et al*.[6] in 2007 in the particular context of medicine and surgery, and the other was conducted more recently by Khalil and Yasmin[8] to evaluate its validity and reliability in terms of team score rather than individual score in the medical field. It was hypothesised that TOSBA is a valid and reliable tool to assess the clinical skills of undergraduate students at bedside in the paediatric clerkship. This study was planned with the objective of developing a TOSBA tool for assessing paediatric undergraduate students' clinical skills at the bedside.

METHODOLOGY

This psychometric validation, cross-sectional study was performed at the Department of Paediatrics, Ziauddin Medical University Hospital, Karachi, Pakistan, from March to June 2023. Fourth-year medical students of either gender during their paediatric clerkship were included in the study. The Ethical Review Committee approved the study prior to its commencement (Reference Code: 6301222HRPED, Dated: January 20, 2023). The data were collected from groups coming for rotation from March to June. However, this formative evaluation was carried out throughout the duration of the academic year 2023, following the study's conclusion. A non-probability purposive homogeneous sampling technique was used. Three clinical groups comprising 67 students were recruited. The inclusion criteria were 4th-year paediatric clerkship students who agreed to receive a briefing about TOSBA and its outcomes. The exclusion criteria were students with attendance below 70% and who were unable to participate in the formative assessment. Six irregular students were excluded, so 61 students participated in this study. Informed written consent was sought from all participants.

Three groups of students coming for a four-week paediatrics clerkship were recruited for piloting. In a detailed orientation session on the first day of their rotation, TOSBA was briefed to the participants by paediatric faculty. TOSBA is a structured, team-based formative assessment tool designed to evaluate undergraduate students' clinical skills at bedside. TOSBA emphasises collaborative learning, allowing small groups of students to be assessed on core competencies such as history-taking, physical examination, clinical reasoning, communication skills, and decision-making in real patient encounters. TOSBA integrates multiple assessment stations per rotation, ensuring repeated evaluations for each student. Depending on the number of students, TOSBA was organised four to five times per rotation: Twice in the second and third weeks. Those who did not match the exclusion criteria were given regular informal observations once in the last week. A full 75-minute TOSBA session included 10 minutes for each activity, plus an additional 5 minutes for feedback. Students who refrained from participating got equal-length self-study periods in parallel. Five teams worked on five typical paediatric cases—pneumonia, diarrhoea, anaemia, malnutrition, and measles—during each TOSBA session. After completing four or five cases in the rotation, each group conducted one case at every interaction.

A team of five students completed a separate task on each patient: Taking a history, performing a physical examination, formulating a differential diagnosis / investigation, creating a management plan, and counselling. The fundamental idea behind each task's assignment was to build upon prior task-based knowledge and feedback to provide each student with a unique task that would not be replicated the following time. At the conclusion of each assignment, the facilitator designated to lead the TOSBA session observed the participants' performance and promptly gave official feedback based on the education incorporated. The paediatric faculty members attended a half-day session on how to give effective feedback and getting formal instruction prior to the tool's pilot project. The workshop component includes hands-on activities regarding reflection, facilitator response, and commitment to a joint action plan. The facilitator gave detailed written feedback on each student's performance, including global ratings, accurate answers, and errors. Following the facilitator's in-person observations, each student received comments on their performance. The abilities, ideas, and feedback of every group member were visible to the other members of the group. For instance, during their TOSBA exchanges, student A gave one presentation on history but heard the stories and criticism of four other people. Out of the total 67 participants, 61 completed four to five rounds of the TOSBA circuit throughout their rotation period, and 10 OSCE stations concluded the rotation. To gather validation data for the TOSBA tool in accordance with Messick's validity approach, psychometric analysis of both TOSBA and OSCE scores was conducted. Figure 1 shows a CONSORT diagram explaining the study flow.

The data analysis was conducted using IBM-SPSS Statistics version 26. Cronbach's alpha (c-alpha) was calculated to check the internal consistency taking a value ≥0.70 considered acceptable for reliability. Spearman's correlation was utilised, with scores above 0.40 meeting the reliability criteria. The corre-

lation between TOSBA scores for each clinical skill and the corresponding OSCE score was calculated, with the coefficient of determination (r^2) used to assess convergent validity. Divergent validity was evaluated by calculating r^2 between TOSBA scores for each clinical skill and the OSCE score of a different skill. Additionally, the end-of-rotation OSCE global rating and TOSBA grades based on the global rating were correlated. A p-value of less than 0.05 was considered significant.

RESULTS

All five clinical skills evaluated by the TOSBA had Cronbach's alpha values >0.7. History taking, counselling skills, physical examination, clinical reasoning, and decision-making returned Cronbach's alpha scores of 0.826, 0.833, 0.928, 0.895, and 0.868, respectively. Internal consistency was again determined to be >0.7 in the majority of cases, with a few outliers, upon closer examination of the TOSBA station-wise. After examining the TOSBA station-wise data in more detail, internal consistency was redetermined to be >0.7 in the majority of cases, with very few outliers. The three stations in the decision-making skills component (diarrhoea = 0.47, measles = 0.36, and pneumonia = 0.55) revealed lower values of Cronbach's alpha, demonstrating less reliability and validity of that station.

Since the study's data were ordinal and non-normally distributed, Spearman's correlation test was employed to determine correlation. For every item pertaining to history-taking and clinical reasoning skills, the r^2 value was determined to be >0.4 ($p < 0.001$). The r^2 values for subsets of the TOSBA tool were generally above 0.4 with few exceptions. Item-total correlation was calculated for clinical skills on a station-wise basis, resulting in a c-alpha of less than 0.7. A strong correlation was observed between the scores of the two constructs, ranging from 0.46 to 0.73 (p-value <0.001) across four different skills. These results suggest that the two tests are similar in terms of convergent validity. Additionally, the scores of the physical examination subset showed a negative correlation, indicating that as one set of scores increases, the other decreases, and vice versa (Table I). Within each clinical skill, score correlations were also conducted station-wise to ascertain construct similarity, which reflected the varying degrees of correlations between the scores.

History-taking evaluation exhibited no strong correlation with clinical reasoning or physical examination but showed a negative correlation with counselling and decision-making skills. Physical examination scores were found to have a negative correlation with counselling skills as well as decision-making, while it did not have any significant correlation with history-taking and a weak correlation with clinical reasoning. Clinical reasoning scores only correlated with physical examination and did not relate to other skills. Decision-making scores were negatively correlated with history-taking and physical examination, showing no correlation with counselling skills or clinical reasoning. These findings indicate substantial validity evidence and highlight the variability among the constructs of the different subgroups. The divergent validity of the instrument is shown in Table II.

Table I: Correlation analysis of TOSBA and OSCE scores.

Scores	Skills	TOSBA	OSCE	Correlation coefficient	p-value
Total scores (%)	History taking	55 (43-62)	64 (60-70)	0.73	<0.001
	Counselling skills	51 (40-58)	60 (55-70)	0.57	<0.001
	Physical examination	45 (34-55)	65 (58-74)	-0.07	0.596
	Critical reasoning	58 (37-63)	60 (50-75)	0.46	<0.001
	Decision-making	55 (46-60)	60 (50-70)	0.73	<0.001
Global rating scores (%)	History taking	70 (60-70)	80 (70-80)	0.43	0.001
	Counselling skills	59 (50-60)	70 (70-80)	0.30	0.030
	Physical examination	60 (50-68)	70 (70-80)	0.26	0.058
	Clinical reasoning	70 (50-70)	70 (70-80)	0.50	<0.001
	Decision-making	70 (60-70)	70 (70-80)	0.32	0.012
Total TOSBA scores *versus* global rating scores	History taking	55 (43-62)	70 (60-70)	0.72	<0.001
	Counselling skills	51 (40-58)	59 (50-60)	0.65	<0.001
	Physical examination	45 (34-55)	60 (50-68)	0.66	<0.001
	Critical reasoning	58 (37-63)	70 (50-70)	0.89	<0.001
	Decision-making	55 (46-60)	70 (60-70)	0.79	<0.001

*TOSBA and OSCE values are presented in the median and inter-quartile ranges. *Spearman's correlation applied.*

Table II: Correlation matrix showing divergent validity of TOSBA.

Parameters		History taking	Counselling skills	Physical examination	Critical reasoning	Decision making
History taking	r	1.00	-0.05	0.11	0.08	-0.02
	p-value	-	0.728	0.424	0.563	0.90
Counselling skills	r	-0.05	1.00	-0.03	0.04	0.01
	p-value	0.728	-	0.836	0.796	0.96
Physical examination	r	0.11	-0.03	1.00	0.23	-0.031
	p-value	0.424	0.836	-	0.091	0.825
Critical reasoning	r	0.08	0.04	0.23	1.00	0.04
	p-value	0.563	0.796	0.091	-	0.768
Decision-making	r	-0.02	0.01	-0.03	0.04	1.00
	p-value	0.902	0.960	0.825	0.768	-

Spearman's correlation applied.

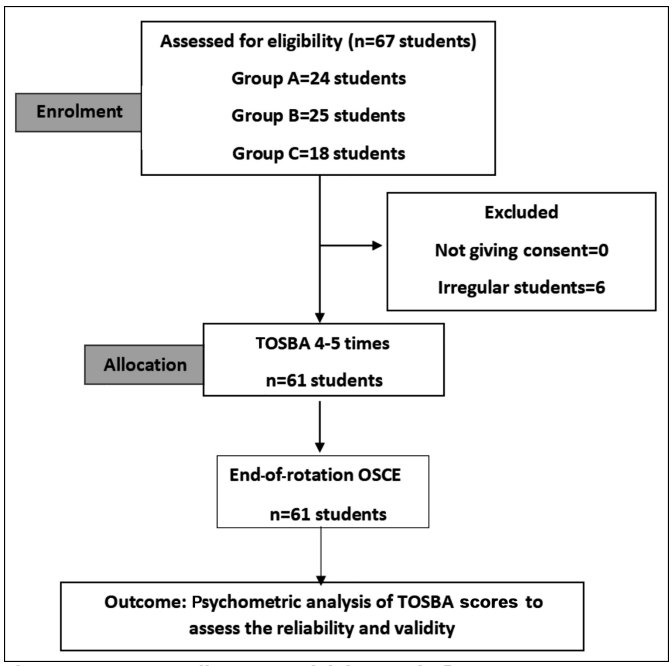

Figure 1: CONSORT diagram explaining study flow.

The heterotrait monomethod demonstrated divergent validity by showing no or a negative correlation between scores obtained for various abilities within the same method, either TOSBA or OSCE. This clearly defined the divergent validity of the TOSBA tool and again showed no or negative association. Reliability was diagonal, representing the internal structure of TOSBA and OSCE in terms of c-alpha of all skill subsets, *i.e.*, >0.7. On the contrary, the validity diagonal was the correlation of scores secured in similar skills/constructs of the two methods. TOSBA and OSCE indicated a strong correlation, except for the physical examination. Figure 2 shows the overall validity and reliability of the TOSBA tool depicted in a multitrait-multimethod matrix (MMM).

DISCUSSION

The present findings demonstrated the general reliability of the TOSBA assessment method, which was created for formative evaluation of paediatric undergraduate students' clinical skills. The validity evidence varies for each of the five clinical abilities, however, with physical examination shows a negative association, while history-taking and decision-making skills show a positive link. The acceptability, validity, cost-effectiveness, and educational impact of an evaluation tool all affect its utility.[9,10] Although utility frameworks have been analysed to determine the effectiveness of assessment tools, there are still a lot of instruments for watching clinical skills in action that do not have any valid data or outline the learning objectives. A systematic review by Kogan *et al*. suggested that only 11 of the 55 formative assessment instruments published in the literature had evidence of their validity based on their internal organisation and relationships to other factors.[11]

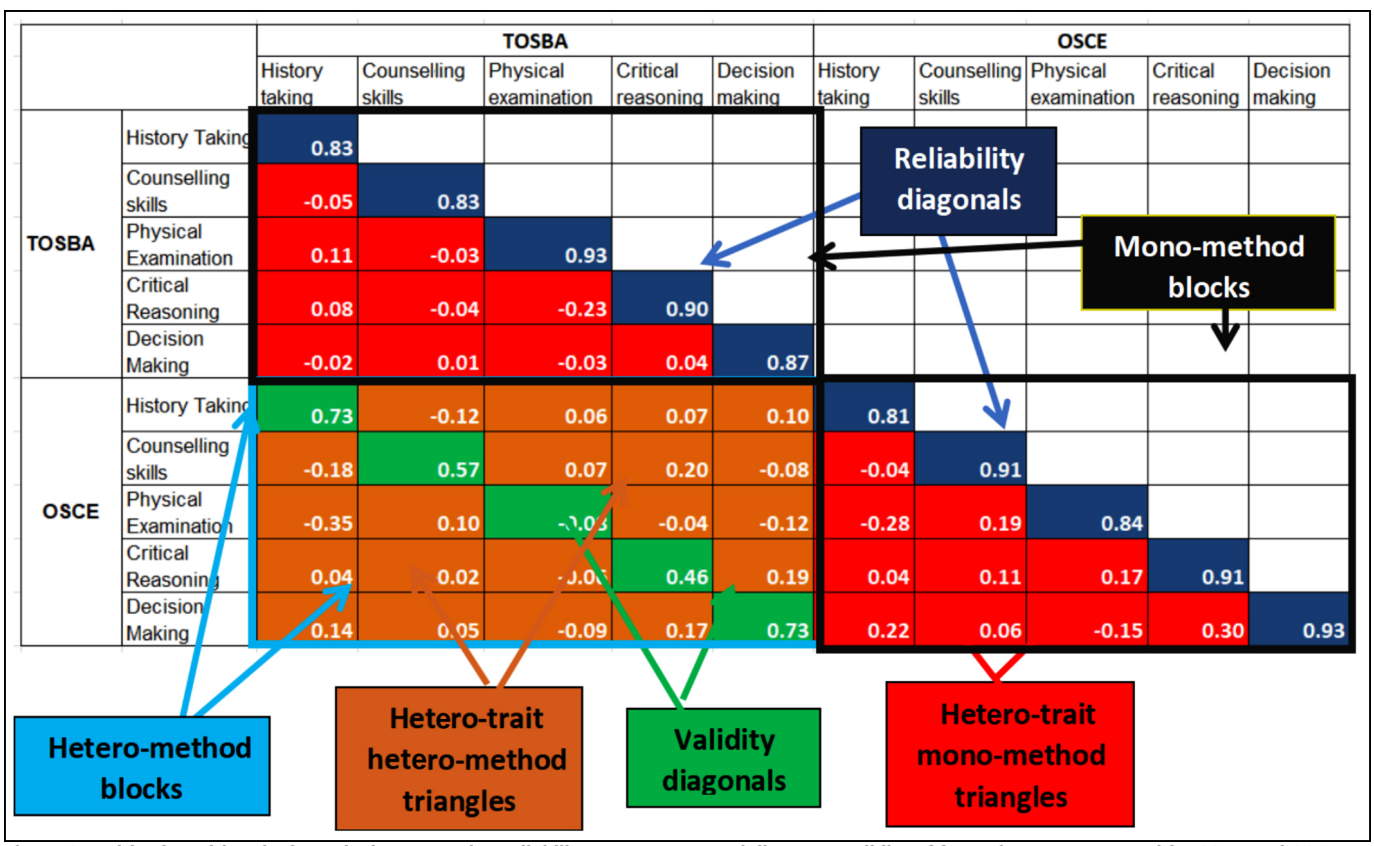

Figure 2: Multitrait-multimethod matrix demonstrating reliability, convergent, and divergent validity of formative assessment with TOSBA and OSCE.

The most likely explanation is that the first TOSBA was established at Royal College Ireland in 2007 and the 1st validity report was released in 2009.[12] It has now been used in other therapeutic disciplines for formative assessment purposes in a modified version, offering some indication of its usefulness in terms of acceptance, consequences, reliability, and validity.

In this study, the c-alpha was >0.7 for all five clinical skills in subsets of TOSBA, exhibiting an excellent reliability index. A systematic assessment of 58 studies on the applicability of Mini-CEX reported a Cronbach's alpha of 0.58 to 0.97. Yet there were also significant differences in the estimated number of encounters required for a desired G coefficient.[13] The item-total correlation was computed as an additional internal structure metric. The majority of the TOSBA tool's items can distinguish between high and low achievers, according to the results. Items with poor correlation should be changed or removed, if at all possible, in order to improve the TOSBA and obtain a higher c-alpha. The TOSBA was utilised in the present study to grade students on two scales: The global rating scale, which uses the P- to P+ grading method often employed by researchers,[14] and the numerical rating scale. Research findings indicate that overall scores on the OSCE have strong psychometric qualities, regardless of whether they are utilised in tandem with a checklist or alone.[15] A good correlation (r^2 = 0.6-0.8) between the two measures within TOSBA indicated agreement, according to this research's findings.

The present results showed a strong association between the TOSBA scores and the OSCE scores for each of the four clinical skills (r^2 = 0.4-0.7) both on the global rating scale (r^2 = 0.262-0.501) and on the numerical rating scale, comparable to the observation of Meagher *et al*.[12] Meagher *et al*. also reported that the patient-based aspect of the OSCE, the clinical long case, and the communication skills, all showed strong correlations with TOSBA results. A moderate association (r^2 = 0.35) was found in their study between the performance on the final exam and TOSBA scores. Meagher *et al*. added that TOSBA outperforms Object Structured Long Examination Record (OSLER) in terms of predicting final exam performance (r^2 = 0.35) and OSLER (r^2 = 0.15).[12] These findings also align with those of Mortez *et al*., who found moderate-to-significant correlations between other clinical exams and mini-CEX scores.[13] In this study, the divergent validity of TOSBA through both hetero-trait, hetero-method, and hetero-trait monomethod showed reliable values. Researchers in the past have also correlated TOSBA and MCQ essay components, and their findings corresponded less well with an OSCE data component and less well with fact-oriented, knowledge- and memory-testing MCQs, essay papers, and brief note assessments.[12,13] The TOSBA physical exam subgroup in this study shows a very weak negative correlation (r^2 = -0.08) or no association at all with the OSCE score on the numerical rating scale, while the global rating shows a weak correlation (r^2 = 0.232) that indicates weak agreement between the two constructs. The weak negative correlation between TOSBA physical examination scores and OSCE scores, along with the weak global rating correlation, highlights a potential discrepancy between the two assessments. These findings are critical, as physical examination skills are often poor in exit examinations, raising concerns about skill acquisition in undergraduate training. The disparity may stem from differences in assessment context as TOSBA evaluates students in real bedside settings with actual patients, requiring adaptability, whereas OSCE stations offer standardised, structured scenarios. OSCEs may focus on checklist-based, task-specific performance, while TOSBA emphasises clinical integration and teamwork. This weak agreement underscores the need for enhanced physical examination training, ensuring that formative assessments such as TOSBA effectively reinforce core competencies required for summative evaluations and clinical practice. However, the present findings are inconsistent with the finding of Durning *et al*., revealing a correlation coefficient of 0.6 (p = 0.002) between Mini-CEX physical examination component scores and ABIM MEF scores in emergency medicine.[16] Since the physical examination is the only psychomotor domain of learning that is evaluated in TOSBA, students' scores at the end of the rotation (who get lower scores in TOSBA) may be improved by the experiential learning cycle, which consists of actual patient encounters during TOSBA, feedback during rotation, and strategic implementation of the improvement plan, which includes watching videos of examinations, practising with peers and junior residents. It is supported by the findings of Jain *et al*.,[17] and Durning *et al*.,[16] in general, not specific to physical examination skills, showing that TOSBA, as a formative assessment method, improves all three major learner domains and reinforces team communication skills. According to Jain *et al*.,[17] the study group's post-test mean score (7.51 ± 0.67) was higher than the control group's (6.34 ± 0.12), which did not receive feedback (p <0.0001). However, Deane *et al*.[18] did not reveal any data regarding the correlation between TOSBA and summative exam scores because they perceived that a perfect correlation, much less a desired one, between formative and summative exam scores was not anticipated because the purpose of formative assessment was to assist learning. They stated that most students, who did not do well on the TOSBA, went on to pass the summative exams (n = 20/21, 95%). This could be due to the satisfaction of students with their performance and the lenient marking of the facilitator in the formative assessment of physical examination. An additional factor that prevents any link between the scores is the lack of standardisation in the clinical findings of patients, a well-known drawback of bedside assessment, as evident in long cases.[19] For TOSBA, the authors used actual admitted patients; for OSCE, a combination of actual and standardised patients were used. By comparing the formative OSCE and TOSBA scores at the same rotation period with similar sets of patients, one may

more accurately determine the degree of agreement between the physical examination constructs.[20,21]

The study has some limitations, despite the fact that it produced a formative evaluation instrument for undergraduate paediatric students' clinical skills that provides evidence of validity and reliability. Because of the small sample size, additional validity checks, such as exploratory factor analysis, were not carried out to further improve the TOSBA tool. If exploratory factor analysis is used on a small sample size, it could yield a biased conclusion. A further limitation of Messick's paradigm was the lack of evidence about the consequences of using TOSBA as a formative assessment tool.

CONCLUSION

The TOSBA tool was developed to assess the clinical skills of undergraduate paediatric students within a formative framework, grounded in evidence-based practices. This innovative tool has demonstrated both validity and reliability in its specific context. Beyond cumulative validity evidence, this study uniquely examined skill-wise and station-wise correlations, contributing new insights into the construct validity of TOSBA, which have not been previously documented in the literature.

ETHICAL APPROVAL:

Approval for the study was obtained from the Ethical Review Committee of the Ziauddin Medical University Hospital, Karachi, Pakistan (Reference Code: 6301222HRPED, Dated: January 20, 2023). The study was conducted in accordance with the Declaration of Helsinki.

COMPETING INTEREST:

The authors declared no conflict of interest.

AUTHORS' CONTRIBUTION:

HR: Designed the study, conducted the analysis, and drafted and revised the manuscript for important intellectual content.

ASA: Conceived and designed the manuscript, and critical revision.

ASS, FZ: Conducted the analysis, interpreted the results, and drafted the manuscript.

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

REFERENCES

1. Teunissen PW, Westerman M. Opportunity or threat: The ambiguity of the consequences of transitions in medical education. *Med Educ* 2011; **45(1)**:51-9. doi: 10.1111/j.1365-2923.2010.03755.x.
2. LaCombe MA. On bedside teaching. *Ann Intern Med* 1997; **126(3)**:217-20. doi: 10.7326/0003-4819-126-3-199702010-00007.
3. Stickrath C, Aagaard E, Anderson M. MiPLAN: A learner-centered model for bedside teaching in today's academic medical centers. *Acad Med* 2013; **88(3)**:322-7. doi: 10.1097/ACM.0b013e318280d8f7.
4. Rauf A, Shamim MS, Aly SM, Chundrigar T, Alam SN. Formative assessment in undergraduate medical education: concept, implementation and hurdles. *J Pak Med Assoc* 2014; **64(1)**:72-5.
5. Yusuf L, Ahmed A, Yasmin R. Educational impact of mini-clinical evaluation exercise: A game changer. *Pak J Med Sci* 2018; **34(2)**:405-11. doi: 10.12669/pjms.342.14667.
6. Miller SD, Butler MW, Meagher F, Costello RW, McElvaney NG. Team objective structured bedside assessment (TOSBA): A novel and feasible way of providing formative teaching and assessment. *Med Teach* 2007; **29(2-3)**:156-9. doi: 10.1080/01421590701287889.
7. Sultan AS. Bedside teaching: An indispensible tool for enhancing the clinical skills of undergraduate medical students. *J Pak Med Assoc* 2019; **69(2)**:235-40.
8. Khalil KF, Yasmin R. Assessing team in team objectively structured bedside assessment (TOSBA). *J Pak Med Assoc* 2020; **70(8)**:1436-8. doi: 10.5455/JPMA.11480.
9. Arekat M, Shehata MH, Deifalla A, Al-Ansari A, Kumar A, Alsenbesy M, *et al*. Evaluation of the utility of online objective structured clinical examination conducted during the COVID-19 pandemic. *Adv Med Educ Pract* 2022; **13**:407-18. doi: 10.2147/AMEP.S357229.
10. Tabriz FD, Bahramian A, Asdagh S, Ghadimi F, Katebi K. The effect of direct observation of procedural skills on learning and satisfaction of students in community dentistry course. *J Adv Med Educ Prof* 2024; **12(2)**:111-7. doi: 10.30476/JAMP.2023.100431.1890.
11. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. *J Am Med Assoc* 2009; **302(12)**:1316-26. doi: 10.1001/jama.2009.1365.
12. Meagher FM, Butler MW, Miller SD, Costello RW, Conroy RM, McElvaney NG. Predictive validity of measurements of clinical competence using the team objective structured bedside assessment (TOSBA): Assessing the clinical competence of final year medical students. *Med Teach* 2009; **31(11)**:e545-0. doi: 10.3109/01421590903095494.
13. Mortaz HS, Jalili M, Masoomi R, Shirazi M, Nedjat S, Norcini J. The utility of mini-clinical evaluation exercise in undergraduate and postgraduate medical education: A BEME review: BEME Guide No. 59. *Med Teach* 2020; **42(2)**:125-42. doi: 10.1080/0142159X.2019.1652732.
14. Bobos P, Ziebart C, Furtado R, Lu Z, MacDermid JC. Psychometric properties of the global rating of change scales in patients with low back pain, upper and lower extremity disorders. A systematic review with meta-analysis. *J Orthop* 2020; **21**:40-8. doi: 10.1016/j.jor.2020.01.047.
15. Elabd K, Abdul-Kadir H, Alkhenizan A, Alkhalifa MK. A comparison of the checklist scoring systems, global rating systems, and borderline regression method for an objective structured clinical examination for a small cohort in a Saudi medical school. *Cureus* 2023; **15(6)**:e39968. doi: 10.7759/cureus.39968.

16. Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med* 2002; **77(9)**:900-4. doi: 10.1097/00001888-200209000-00020.

17. Jain S, Kumar S, Tankhiwale N. Team based objective formative assessment method for postgraduate students: Evaluation of students' clinical performance and perceptions. *J Med Res* 2016; **2(2)**:47-52. doi: 10.31254/jmr.2016.2209.

18. Deane RP, Joyce P, Murphy DJ. Team objective structured bedside assessment (TOSBA) as formative assessment in undergraduate obstetrics and gynaecology: A cohort study. *BMC Med Educ* 2015; **15**:172. doi: 10.1186/s12909-015-0456-5.

19. Grannis SJ, Xu H, Vest JR, Kasthurirathne S, Bo N, Moscovitch B, *et al*. Evaluating the effect of data standardization and validation on patient matching accuracy. *J Am Med Inform Assoc* 2019; **26(5)**:447-56. doi: 10.1093/jamia/ocy191.

20. Sulaiman ND, Shorbagi SI, Abdalla NY, Daghistani MT, Mahmoud IE, Al-Moslih AM. Group OSCE (GOSCE) as a formative clinical assessment tool for pre-clerkship medical students at the University of Sharjah. *J Taibah Univ Med Sci* 2018; **13(5)**:409-14. doi: 10.1016/j.jtumed.2018.06.003.

21. Al-Hashimi K, Said UN, Khan TN. Formative objective structured clinical examinations (OSCEs) as an assessment tool in UK undergraduate medical education: A review of its utility. *Cureus* 2023; **15(5)**:e38519. doi: 10.7759/cureus.38519.