INTERNATIONAL MULTI DISCIPLINARY JOURNAL FOR RESEARCH & DEVELOPMENT
eISSN 2394-6334 Impact factor: 8.854 Volume 13, issue 04 (2026)

INNOVATIVE APPROACHES TO ASSESSING MEDICAL STUDENTS' COMPETENCIES IN SIMULATION-BASED LEARNING: DEVELOPMENT AND INTEGRATION OF A TESTING SYSTEM

Khudoyberdiev A.A., Abdullaeva U.U.

Tashkent State Medical University, Tashkent, Uzbekistan

Abstract. The article presents the results of a study examining the effectiveness of an innovative competency assessment system for medical students in simulation-based learning environments. The developed testing system has been integrated into the educational process and enables comprehensive evaluation of both theoretical knowledge and practical skills. Analysis of student survey data and test results demonstrates improved objectivity of assessment and enhanced quality of future physicians' training. The proposed approach contributes to the development of clinical thinking and increases student motivation for learning.

Keywords: simulation-based learning, competency-based approach, medical education, testing system, skills assessment, objective structured clinical examination, digitalization of education.

Introduction (Relevance)

The transformation of medical education in the twenty-first century is inextricably linked to the widespread adoption of simulation technologies, which provide a safe, reproducible, and ethically sound environment for the acquisition and consolidation of clinical competencies [1, 8]. Globally, simulation-based medical education (SBME) has evolved from a supplementary pedagogical tool into a cornerstone of undergraduate and postgraduate clinical training, underpinned by the fundamental principle that procedural and diagnostic skills must be mastered before their application on real patients [3, 12]. However, the mere introduction of simulation equipment and standardized scenarios into the curriculum does not automatically guarantee the attainment of the desired educational outcomes unless it is accompanied by a robust, valid, and reliable system for assessing the competencies acquired [5, 14].

Traditional assessment instruments — written examinations, oral questioning, and direct observation by a single examiner — have well-documented limitations when applied to simulation-based contexts: subjectivity of evaluation, insufficient granularity in measuring procedural performance, and an inability to capture the dynamic, integrative nature of clinical decision-making [2, 9]. The Objective Structured Clinical Examination (OSCE), introduced by Harden and Gleeson in 1975, represented a landmark advance in standardizing clinical skills assessment; nevertheless, its implementation in resource-limited settings remains challenging, and its integration with digital testing platforms is still at an early stage in many medical universities of Central Asia [6, 15].

The Republic of Uzbekistan has embarked on a comprehensive reform of its higher medical education system, with the State Programme for the Development of Medical Education 2022-2030 explicitly mandating the transition to competency-based curricula and the adoption of modern assessment technologies [10]. Within this context, Tashkent State Medical University has prioritized the development of simulation centres and the modernization of assessment practices. Despite these institutional efforts, a significant gap remains between the availability of simulation infrastructure and the quality of competency measurement tools applied within it. Specifically, there is an absence of validated, digitally integrated testing systems capable of simultaneously evaluating theoretical knowledge, clinical reasoning, and procedural performance within a unified simulation cycle [4, 11].

The relevance of the present study is therefore determined by the convergence of three factors: (1) the growing global evidence base supporting the superiority of competency-based, technology-enhanced assessment over traditional methods in medical education; (2) the national policy imperative to modernize medical training in Uzbekistan; and (3) the practical necessity of providing faculty with objective, data-driven tools for monitoring student progress throughout simulation cycles. Addressing this gap through the development and empirical evaluation of an integrated testing system constitutes both a scientific contribution and a practical response to an urgent educational challenge.

Aim of the Study: The aim of the present study was to develop an integrated testing system for the assessment of medical students' competencies within simulation-based learning cycles and to evaluate its effectiveness in improving the quality of theoretical knowledge acquisition, practical skill performance, and overall clinical preparedness among fifth-year medical students of the Faculty of General Medicine at Tashkent State Medical University.

The specific objectives were as follows:

1. To design a multi-component digital testing system incorporating pre-simulation knowledge assessment, intra-simulation performance checklists, and post-simulation reflective evaluation modules.

2. To compare academic performance outcomes, skill proficiency scores, and subjective satisfaction indicators between students who underwent simulation cycles with integrated testing support and those who followed the conventional simulation curriculum without such support.

3. To assess the statistical significance of inter-group differences and to formulate evidence-based recommendations for the broader implementation of the proposed system.

Materials and Methods

A prospective controlled educational study was conducted at the Simulation Training Centre of Tashkent State Medical University during the 2023-2024 academic year. The study population comprised fifth-year students of the Faculty of General Medicine who were enrolled in the mandatory simulation cycle as part of their clinical skills training programme.

A total of 113 students were enrolled in the study and allocated into two groups according to the academic subgroup assignment schedule established at the beginning of the academic year, thereby ensuring allocation that was independent of individual academic performance:

- Group 1 (Experimental Group): 67 students who underwent the simulation cycle with the full integration of the newly developed testing system.

- Group 2 (Control Group): 46 students who completed the identical simulation cycle under the conventional curriculum, without access to the integrated testing system.

The two groups were comparable in terms of gender distribution, prior academic performance (cumulative GPA for years 1-4), and age (Table 1). Ethical approval for the study was obtained from the Institutional Review Board of Tashkent State Medical University (Protocol No. 14, dated 05.09.2023). All participants provided written informed consent.

Inclusion criteria: full-time fifth-year students enrolled in the simulation cycle; absence of prior participation in a pilot testing of the system; provision of informed consent.

Exclusion criteria: absence from more than 20% of simulation sessions; withdrawal of consent; concurrent enrolment in an external clinical skills programme.

The integrated testing system developed for this study comprised three sequential and interrelated modules:

Module 1 — Pre-Simulation Knowledge Assessment (PSKA). Before each simulation session, students in Group 1 completed a 20-item computer-based test covering the theoretical foundations relevant to the forthcoming clinical scenario (e.g., pathophysiology, pharmacology, diagnostic algorithms). Items were constructed in multiple-choice (single best answer) and clinical vignette formats. The test was administered via the university's Learning Management System (LMS) and was time-limited to 25 minutes. Immediate automated feedback with explanatory comments was provided upon completion.

Module 2 — Intra-Simulation Performance Checklist (ISPC). During simulation sessions, faculty assessors used a standardized digital checklist to evaluate procedural accuracy, adherence to clinical protocols, communication skills, and teamwork. The checklist was developed on the basis of internationally validated tools, including the Ottawa Crisis Resource Management Global Rating Scale and procedure-specific checklists from the Royal College of Physicians and Surgeons of Canada. Data were entered in real time via tablet devices and automatically uploaded to the student's individual performance profile.

Module 3 — Post-Simulation Reflective Evaluation (PSRE). Following each simulation session, students completed a structured self-assessment questionnaire (15 items, 5-point Likert scale) addressing perceived competence, identification of learning gaps, and satisfaction with the educational experience. Simultaneously, faculty completed a brief summative evaluation form. Both datasets were integrated into the student's cumulative competency profile.

The control group (Group 2) underwent the same simulation scenarios and received standard faculty feedback at the end of each session, without access to pre-session digital testing, standardized intra-session checklists, or structured post-session self-assessment tools.

The following primary and secondary outcome measures were assessed at the end of the simulation cycle (8 weeks):

1. Theoretical Knowledge Score (TKS): mean score on a standardized 40-item written examination covering all topics addressed during the simulation cycle (maximum 100 points).

2. Practical Skills Score (PSS): mean score on an OSCE-format practical examination comprising 6 stations (maximum 100 points per station; composite score reported as a percentage).

3. Clinical Reasoning Score (CRS): mean score on a 10-item extended matching question (EMQ) examination assessing diagnostic and management decision-making (maximum 100 points).

4. Student Satisfaction Index (SSI): mean score on a validated 20-item questionnaire assessing overall satisfaction with the simulation learning experience (maximum 100 points).

5. Rate of Competency Achievement (RCA): proportion of students achieving a passing threshold of ≥75 points across all three objective assessment components simultaneously.

Statistical analysis was performed using IBM SPSS Statistics, Version 26.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean ± standard deviation (M ± SD). The normality of distribution was verified using the Shapiro-Wilk test. Between-group comparisons of normally distributed continuous variables were performed using the independent samples Student's *t*-test. For non-normally distributed variables, the Mann-Whitney U test was applied. Categorical variables were compared using the chi-square (χ^2) test or Fisher's exact test, as appropriate. The level of statistical significance was set at *p* < 0.05 for all analyses. Effect sizes were calculated using Cohen's *d* for continuous variables, with values of 0.2, 0.5, and 0.8 interpreted as small, medium, and large effects, respectively.

Results: The baseline characteristics of the two study groups are presented in Table 1. No statistically significant differences were identified between Group 1 and Group 2 with respect to age, gender distribution, or prior cumulative GPA, confirming the comparability of the groups at the outset of the study.

Table 1.

Baseline Characteristics of Study Participants

Characteristic	Group 1 (n = 67)	Group 2 (n = 46)	p-value
Age, years (M ± SD)	22.4 ± 0.8	22.6 ± 0.9	0.214
Gender: male, n (%)	31 (46.3%)	22 (47.8%)	0.871
Gender: female, n (%)	36 (53.7%)	24 (52.2%)	0.871
Cumulative GPA, years 1-4 (M ± SD)	3.74 ± 0.31	3.71 ± 0.34	0.618
Prior simulation experience, n (%)	12 (17.9%)	9 (19.6%)	0.818

Note: GPA expressed on a 5.0 scale. p-values derived from independent samples t-test (continuous variables) and χ^2 test (categorical variables)

The results of the end-of-cycle objective assessments are presented in Table 2. Students in Group 1 demonstrated statistically significantly higher scores across all three objective performance measures compared with students in Group 2.

The mean Theoretical Knowledge Score in Group 1 was 84.7 ± 6.2 points, compared with 74.3 ± 7.8 points in Group 2 (*t* = 8.14; *p* < 0.001; Cohen's *d* = 1.51, large effect). The mean Practical Skills Score was 81.9 ± 5.7 in Group 1 versus 70.6 ± 8.1 in Group 2 (*t* = 8.73; *p* < 0.001; Cohen's *d* = 1.63, large effect). The mean Clinical Reasoning Score was 79.4 ± 6.9 in Group 1 compared with 68.2 ± 8.4 in Group 2 (*t* = 7.89; *p* < 0.001; Cohen's *d* = 1.47, large effect).

Table 2.

Comparison of Academic Performance Outcomes Between Groups

Outcome Measure	Group 1 (n = 67) M ± SD	Group 2 (n = 46) M ± SD	t	p	Cohen's d
Theoretical Knowledge Score (TKS)	84.7 ± 6.2	74.3 ± 7.8	8.14	< 0.001	1.51
Practical Skills Score (PSS)	81.9 ± 5.7	70.6 ± 8.1	8.73	< 0.001	1.63
Clinical Reasoning Score (CRS)	79.4 ± 6.9	68.2 ± 8.4	7.89	< 0.001	1.47
Composite Score (mean of TKS+PSS+CRS)	82.0 ± 5.6	71.0 ± 7.4	9.12	< 0.001	1.70

Note: All scores expressed on a 100-point scale. p-values derived from independent samples t-test. Effect size interpretation: d ≥ 0.8 = large.*

The Rate of Competency Achievement — defined as the proportion of students attaining a score of ≥75 points simultaneously on all three objective assessments — was markedly higher in Group 1 than in Group 2 (Table 3). In Group 1, 52 out of 67 students (77.6%) achieved the competency threshold, compared with 18 out of 46 students (39.1%) in Group 2. The difference was statistically significant ($\chi^2 = 17.84$; *p* < 0.001; odds ratio [OR] = 5.54; 95% confidence interval [CI]: 2.51-12.22).

Table 3.

Rate of Competency Achievement (≥75 points on all three assessments)

Indicator	Group 1 (n = 67)	Group 2 (n = 46)	χ^2	p	OR (95% CI)
Achieved competency threshold, n (%)	52 (77.6%)	18 (39.1%)	17.84	< 0.001	5.54 (2.51-12.22)
Did not achieve threshold, n (%)	15 (22.4%)	28 (60.9%)	—	—	—

A more granular analysis of the distribution of students across performance levels (Table 4) further illustrates the superiority of outcomes in Group 1. The proportion of students achieving high performance (score ≥85 points on the composite measure) was 38.8% in Group 1 versus 8.7% in Group 2. Conversely, the proportion of students demonstrating low performance (score <65 points) was 3.0% in Group 1 compared with 23.9% in Group 2.

Table 4.

Distribution of Students by Composite Performance Level

INTERNATIONAL MULTI DISCIPLINARY JOURNAL FOR RESEARCH & DEVELOPMENT
eISSN 2394-6334 Impact factor: 8.854 Volume 13, issue 04 (2026)

Performance Level	Score Range	Group 1 (n = 67) n (%)	Group 2 (n = 46) n (%)
High	≥ 85 points	26 (38.8%)	4 (8.7%)
Satisfactory	75-84 points	26 (38.8%)	14 (30.4%)
Borderline	65-74 points	13 (19.4%)	17 (37.0%)
Low	< 65 points	2 (3.0%)	11 (23.9%)

Note: $\chi^2 = 24.17$; $p < 0.001$ (overall distribution comparison).*

The Student Satisfaction Index scores are presented in Table 5. Students in Group 1 reported significantly higher overall satisfaction with the simulation learning experience compared with Group 2 (83.6 ± 7.1 vs. 71.4 ± 9.3; *t* = 7.96; *p* < 0.001; Cohen's *d* = 1.49). Notably, the subscale scores for "clarity of assessment criteria" (87.2 ± 6.8 vs. 65.3 ± 10.1; *p* < 0.001) and "perceived usefulness of feedback" (85.9 ± 7.4 vs. 68.7 ± 9.8; *p* < 0.001) showed the largest inter-group differences, suggesting that the transparency and immediacy of feedback provided by the integrated testing system were particularly valued by students.

Table 5.

Student Satisfaction Index: Comparison Between Groups

Satisfaction Subscale	Group 1 (n = 67) M ± SD	Group 2 (n = 46) M ± SD	*t*-statistic	p
Overall satisfaction with simulation cycle	83.6 ± 7.1	71.4 ± 9.3	7.96	< 0.001
Clarity of assessment criteria	87.2 ± 6.8	65.3 ± 10.1	13.52	< 0.001
Perceived usefulness of feedback	85.9 ± 7.4	68.7 ± 9.8	10.81	< 0.001
Motivation to self-directed learning	81.3 ± 8.2	70.8 ± 10.4	6.10	< 0.001
Perceived relevance to clinical practice	84.7 ± 7.9	73.6 ± 9.7	6.83	< 0.001

Note: All subscale scores expressed on a 100-point scale.*

To assess the trajectory of skill acquisition over time, mean Practical Skills Scores were recorded at three time points: at the beginning (Week 1), midpoint (Week 4), and end (Week 8) of the simulation cycle (Table 6). In Group 1, a consistent and statistically significant improvement in PSS was observed across all three time points (Week 1: 62.4 ± 8.3; Week 4: 74.8 ± 7.1; Week 8: 81.9 ± 5.7; *F* = 94.3; *p* < 0.001, repeated measures ANOVA). In Group 2, improvement was also observed but was significantly less pronounced (Week 1: 61.9 ± 9.1; Week 4: 66.3 ± 8.4; Week 8: 70.6 ± 8.1; *F* = 21.7; *p* < 0.001). The between-group difference in the rate of improvement (i.e., the group × time interaction) was statistically significant (*F* = 18.6; *p* < 0.001), indicating that the integrated testing system was associated with a steeper and more sustained trajectory of practical skill development.

Table 6.

Dynamics of Practical Skills Score Across the Simulation Cycle

Time Point	Group 1 (n = 67) M ± SD	Group 2 (n = 46) M ± SD	t-statistic	p
Week 1 (baseline)	62.4 ± 8.3	61.9 ± 9.1	0.31	0.757
Week 4 (midpoint)	74.8 ± 7.1	66.3 ± 8.4	5.88	< 0.001
Week 8 (end of cycle)	81.9 ± 5.7	70.6 ± 8.1	8.73	< 0.001
Intra-group improvement (Week 1→8)	+19.5 points	+8.7 points	—	—

Discussion. The results of the present study provide compelling evidence that the integration of a structured, multi-component digital testing system into simulation-based medical education cycles significantly enhances the quality of competency acquisition among fifth-year medical students. The large effect sizes observed across all primary outcome measures (Cohen's *d* ranging from 1.47 to 1.70) are particularly noteworthy and indicate that the observed differences are not only statistically significant but also educationally meaningful.

The superiority of Group 1 outcomes across theoretical, practical, and clinical reasoning domains is consistent with the broader literature on technology-enhanced assessment in medical education. Cook et al. [3] demonstrated in a systematic review that deliberate practice with immediate feedback — a core feature of the Module 1 and Module 2 components of our system — is among the most powerful determinants of skill acquisition in simulation-based contexts. Similarly, Issenberg et al. [8] identified feedback as the single most important feature of effective simulation-based education, a finding that is reflected in the particularly large inter-group differences observed in the "perceived usefulness of feedback" satisfaction subscale in the present study.

The finding that the Rate of Competency Achievement was 5.54 times higher in Group 1 than in Group 2 (OR = 5.54; 95% CI: 2.51-12.22) has direct implications for curriculum design. It suggests that the provision of structured pre-session knowledge activation, real-time performance monitoring, and post-session reflective evaluation creates a self-reinforcing learning cycle that substantially increases the proportion of students achieving clinically meaningful competency thresholds. This is particularly relevant in the context of Uzbekistan's medical education reform, where ensuring a minimum standard of clinical competency across all graduating physicians is a stated policy priority [10].

The trajectory analysis (Table 6) further reveals that the benefits of the integrated testing system were not confined to the end-of-cycle assessment but were observable as early as Week 4 of the simulation cycle, with the between-group gap widening progressively over time. This pattern is consistent with the concept of "learning curves" in simulation-based education described by McGaghie et al. [12] and suggests that the testing system functions not merely as an assessment tool but as an active driver of learning — a distinction that is central to the competency-based education paradigm [5].

The high student satisfaction scores in Group 1, particularly regarding the clarity of assessment criteria and the usefulness of feedback, align with self-determination theory [13], which posits that perceived competence and autonomy — both of which are supported by transparent, timely, and actionable feedback — are fundamental drivers of intrinsic motivation in educational settings. The significantly higher motivation for self-directed learning reported by Group 1 students suggests that the testing system may have long-term benefits extending beyond the immediate simulation cycle, potentially fostering habits of reflective practice that are essential for lifelong professional development [7].

Several limitations of the present study merit acknowledgement. First, the non-randomized allocation of students to groups, while based on pre-existing administrative subgroup assignments and controlled for baseline characteristics, does not entirely exclude the possibility of selection bias. Second, the study was conducted at a single institution, which may limit the generalizability of findings to other medical universities with different resource profiles and curricular structures. Third, the follow-up period was limited to a single simulation cycle (8 weeks), and the long-term impact of the integrated testing system on clinical performance during hospital rotations and beyond remains to be investigated. Future research should address these limitations through multi-centre randomized controlled trials with extended follow-up periods.

Conclusions. The results of the present study demonstrate that the integration of a structured, digitally supported, multi-component testing system into simulation-based medical education cycles leads to a statistically and educationally significant improvement in the quality of competency acquisition among fifth-year medical students. Students who underwent the simulation cycle with the integrated testing system achieved substantially higher scores in theoretical knowledge, practical skills, and clinical reasoning, with large effect sizes confirming the practical significance of these differences. The proportion of students achieving the established competency threshold was nearly twice as high in the experimental group as in the control group, and the trajectory of skill development was steeper and more sustained throughout the cycle. High levels of student satisfaction, particularly regarding the transparency of assessment criteria and the quality of feedback, indicate that the proposed system not only improves objective outcomes but also enhances the subjective learning experience and intrinsic motivation. The developed testing system is feasible for implementation within the existing infrastructure of Tashkent State Medical University and represents a replicable model for other medical institutions in Uzbekistan and the broader Central Asian region seeking to modernize their simulation-based competency assessment practices in alignment with contemporary international standards of medical education.

References

1. Aggarwal R., Mytton O.T., Derbrew M. et al. Training and simulation for patient safety // Quality and Safety in Health Care. — 2010. — Vol. 19 (Suppl. 2). — P. i34-i43.

2. Boulet J.R., Murray D.J. Simulation-based assessment in anesthesiology: requirements for practical implementation // Anesthesiology. — 2010. — Vol. 112, № 4. — P. 1041-1052.

3. Cook D.A., Hamstra S.J., Brydges R. et al. Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis // Medical Teacher. — 2013. — Vol. 35, № 1. — P. e867-e898.

4. Dent J.A., Harden R.M., Hunt D. A Practical Guide for Medical Teachers. — 5th ed. — Edinburgh: Elsevier, 2017. — 424 p.

5. Frank J.R., Snell L., Sherbino J. (eds.) CanMEDS 2015 Physician Competency Framework. — Ottawa: Royal College of Physicians and Surgeons of Canada, 2015. — 30 p.

6. Harden R.M., Gleeson F.A. Assessment of clinical competence using an objective structured clinical examination (OSCE) // Medical Education. — 1979. — Vol. 13, № 1. — P. 41-54.

7. Holmboe E.S., Sherbino J., Long D.M. et al. The role of assessment in competency-based medical education // Medical Teacher. — 2010. — Vol. 32, № 8. — P. 676-682.

8. Issenberg S.B., McGaghie W.C., Petrusa E.R. et al. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review // Medical Teacher. — 2005. — Vol. 27, № 1. — P. 10-28.

9. Kogan J.R., Holmboe E.S., Hauer K.E. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review // JAMA. — 2009. — Vol. 302, № 12. — P. 1316-1326.

10. Ministry of Health of the Republic of Uzbekistan. State Programme for the Development of Medical Education of the Republic of Uzbekistan for 2022-2030. — Tashkent: Ministry of Health, 2022. — 48 p.

11. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement // Academic Medicine. — 2001. — Vol. 76, № 4. — P. 390-393.

12. McGaghie W.C., Issenberg S.B., Cohen E.R. et al. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence // Academic Medicine. — 2011. — Vol. 86, № 6. — P. 706-711.

13. Ryan R.M., Deci E.L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being // American Psychologist. — 2000. — Vol. 55, № 1. — P. 68-78.

14. Scalese R.J., Obeso V.T., Issenberg S.B. Simulation technology for skills training and competency assessment in medical education // Journal of General Internal Medicine. — 2008. — Vol. 23 (Suppl. 1). — P. 46-49.

15. Zayapragassarazan Z., Kumar S. Active learning methods // Online Submission. — 2012. — Vol. 19, № 1. — P. 3-5.