Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Effectiveness of Small Group Discussions with Pre- and Post-Session MCQ-Based Assessment in Enhancing Learning Outcomes Among MBBS Students in General Surgery – A CBME-Aligned Study

Dr. Deepak Arkalgud[1], Dr. Praveen Kumar K H[1], Dr. Omar Ahmed[1], Dr. Prafullachandra Hoogar[1], Dr. Sanjeev R Navalyal[1], Dr. Anupama Deepak[2]*

[1] Department of General Surgery, KLE JGMM Medical College, Hubli, KLE Academy of Higher Education and Research (Deemed University), Belagavi, Karnataka, India – 590010.

[2] Department of Physiology, KLE JGMM Medical College, Hubli, KLE Academy of Higher Education and Research (Deemed University), Belagavi, Karnataka, India – 590010.

*Corresponding author: Dr. Anupama Deepak

(Received: 25 October 2025 Revised: 27 November 2025 Accepted: 04 December 2025)

KEYWORDS

Enhancing Learning Outcomes , MBBS Students

ABSTRACT

Background: Competency-Based Medical Education (CBME) emphasizes active learning methodologies. Small Group Discussions (SGD) are advocated for developing higher-order cognitive skills, yet objective evidence of their effectiveness remains limited. This study evaluated the impact of SGD sessions on knowledge acquisition in General Surgery using pre- and post-session Multiple Choice Question (MCQ) assessments.

Methods: A prospective interventional study was conducted over eighteen months involving 453 Phase III MBBS students across three consecutive academic batches (2021-22, 2022-23, and 2023-24) at a tertiary care medical college. Students underwent structured SGD sessions on 34 different General Surgery topics aligned with CBME competencies. Pre-session and post-session MCQ assessments were administered via Google Forms. Paired t-tests were used to compare scores, and Cohen's d was calculated to determine effect size.

Results: Analysis of 34 SGD sessions with adequate participation revealed significant improvement in post-session scores across all topics ($p \leq 0.002$). Mean improvement was 6.72 marks with an average percentage increase of 64.8%. Cohen's d ranged from 0.19 to 2.59 (mean 1.50), indicating large educational impact. The 2021-22 batch (n=150, 21 topics) showed 33.1% improvement (Cohen's d=1.15), the 2022-23 batch (n=150, 12 topics) demonstrated 93.5% improvement (Cohen's d=1.82), and the 2023-24 batch (n=153, 1 topic with only 7 respondents) showed 104.6% improvement (Cohen's d=2.15) though data was incomplete. Overall, 97.1% of analyzable topics showed statistically significant learning gains.

Conclusion: SGD sessions with structured pre-post MCQ assessment resulted in substantial and statistically significant knowledge gains in General Surgery. The large effect sizes confirm SGD as a highly effective CBME-aligned teaching-learning methodology.

INTRODUCTION

Medical education has undergone substantial transformation over the past decade, shifting from traditional teacher-centered approaches to learner-centered methodologies that emphasize competency acquisition rather than mere knowledge retention.(1) In India, the National Medical Commission implemented Competency-Based Medical Education (CBME) curriculum in 2019, mandating educational institutions to adopt active learning strategies that foster critical thinking, clinical reasoning, and problem-solving skills among undergraduate medical students.(2) This paradigm shift necessitates both pedagogical innovation and robust assessment mechanisms to objectively measure learning outcomes.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Small Group Discussion (SGD) has emerged as a cornerstone teaching-learning methodology within the CBME framework, particularly valued for its capacity to engage students in collaborative learning, encourage peer-to-peer knowledge exchange, and develop higher-order cognitive abilities as defined by Bloom's taxonomy.(3) Unlike passive lecture-based instruction, SGD facilitates active participation, allows students to articulate their understanding, challenges misconceptions through peer discussion, and integrates basic science knowledge with clinical applications—skills essential for developing competent medical practitioners.(4)

General Surgery, being a core clinical specialty in the MBBS curriculum, requires students to master complex concepts ranging from surgical pathophysiology to clinical decision-making. The Phase III Part II curriculum under CBME emphasizes competency-based learning where students must demonstrate not only theoretical knowledge but also the ability to apply this knowledge in clinical contexts. Traditional teaching methods in surgery, predominantly consisting of didactic lectures and bedside teaching, while valuable, often fail to adequately assess or enhance the depth of student understanding across diverse topics. The introduction of SGD sessions provides an opportunity to address this limitation by creating an interactive learning environment where students actively engage with surgical concepts under faculty guidance.

Despite the theoretical advantages and institutional mandate for implementing SGD, there exists a paucity of objective evidence regarding its effectiveness in improving learning outcomes, particularly in surgical disciplines. Most studies evaluating SGD rely on subjective feedback questionnaires or qualitative assessments, which, while informative, do not provide quantitative measures of knowledge acquisition.(5) Furthermore, the absence of standardized pre-intervention assessment makes it difficult to attribute learning gains specifically to the SGD intervention rather than to other concurrent educational activities or natural progression of learning over time.

Multiple Choice Questions (MCQ) represent a valid and reliable method for assessing knowledge acquisition, particularly when designed to test higher-order cognitive skills rather than mere factual recall.(6) High-difficulty MCQs that require application, analysis, and synthesis of information align well with CBME objectives and provide objective, quantifiable data on student learning. The pre-test and post-test design using MCQs offers several advantages: it establishes baseline knowledge levels, allows for paired statistical analysis to detect changes within the same student cohort, minimizes inter-rater variability inherent in subjective assessments, and provides immediate feedback to both students and faculty regarding knowledge gaps.(7)

The concept of using assessment as a learning tool, often termed "assessment for learning," has gained considerable traction in medical education. Pre-session testing serves dual purposes—it activates prior knowledge, priming students for new learning, and identifies specific areas of weakness that can be addressed during the SGD session.(8) Post-session assessment reinforces learning through retrieval practice, a well-established cognitive psychology principle that strengthens memory consolidation and long-term retention.(9) This assessment-learning integration aligns perfectly with CBME's emphasis on continuous assessment and feedback-driven improvement.

Previous research on SGD effectiveness in medical education has produced mixed results, partly due to methodological variations in implementation, assessment strategies, and outcome measures. A systematic review by Burgess and colleagues found moderate evidence supporting SGD for knowledge acquisition but highlighted the need for studies with rigorous quantitative outcome measures.(5) Studies conducted in Indian medical colleges have predominantly focused on student satisfaction and perception rather than objective learning outcomes, creating a gap in the evidence base for SGD implementation under the CBME curriculum.(10) Moreover, there is limited data specifically from surgical disciplines, where the complexity of topics and clinical relevance present unique educational challenges.

The choice of General Surgery as the focus of this study is deliberate and significant. Surgical education requires mastery of both theoretical concepts and practical application, making it an ideal domain to evaluate the effectiveness of interactive teaching methodologies.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Topics in General Surgery span a wide spectrum—from basic surgical principles and metabolic responses to injury, through organ-specific pathologies including gastrointestinal, hepatobiliary, endocrine, and vascular conditions, to oncological principles and trauma management. This diversity allows for testing SGD effectiveness across varying complexity levels and content types. Additionally, surgical knowledge forms the foundation for clinical decision-making in numerous medical specialties, making effective surgical education crucial for producing competent physicians.

Effect size measurement, particularly Cohen's d, provides a standardized metric for educational intervention effectiveness that transcends the limitations of p-values alone. While statistical significance indicates that an effect exists, effect size quantifies the magnitude of that effect, allowing comparison across studies and interventions.(7) In educational research, Cohen's d values of 0.2, 0.5, and 0.8 are conventionally interpreted as small, medium, and large effects respectively. However, in medical education, where learning occurs continuously through multiple modalities, even small to medium effect sizes can be considered educationally meaningful.

The implementation of SGD in medical education faces several practical challenges that must be acknowledged. Faculty training in facilitation skills, optimal group size determination, time constraints within packed curricula, and ensuring consistent quality across multiple parallel sessions all require careful attention. Additionally, the traditional hierarchical structure in medical education, where students may hesitate to openly discuss or question in front of faculty, can inhibit the free exchange of ideas that SGD aims to promote. Addressing these challenges through structured protocols, faculty development programs, and creating a psychologically safe learning environment is essential for successful SGD implementation.

The timing and structure of assessments also merit consideration. Pre-session assessments conducted immediately before SGD sessions capture baseline knowledge and prime students for learning. Post-session assessments, when administered shortly after the session, measure immediate learning gains but may not reflect long-term retention. Ideally, a three-phase assessment (pre-session, immediate post-session, and delayed post-session) would provide comprehensive understanding of both immediate and sustained learning. However, practical constraints in educational settings often necessitate limiting assessments to pre- and immediate post-session, which remains valuable for evaluating short-term knowledge acquisition.

This study was conceptualized to address the existing knowledge gap by providing robust quantitative evidence for SGD effectiveness in General Surgery education under the CBME framework. By implementing structured SGD sessions across multiple topics and using validated MCQ-based assessments, the study aimed to objectively measure knowledge gains attributable to the SGD intervention. The use of paired analysis, where each student serves as their own control, strengthens the study design by minimizing confounding variables related to baseline knowledge differences between students. Furthermore, calculating effect sizes provides meaningful interpretation of the educational impact beyond mere statistical significance.

The findings from this study have important implications for medical education policy and practice. If SGD with structured assessment demonstrates large effect sizes consistently across diverse topics, it provides strong justification for allocating resources, time, and faculty effort toward implementing this methodology more widely. Conversely, if certain topics show minimal benefit, it suggests that SGD may be more effective for particular types of content, guiding strategic curriculum planning. Additionally, the methodology employed—combining structured discussion with immediate assessment—could serve as a model for implementing and evaluating other active learning strategies in medical education.

In conclusion, this study represents a systematic attempt to evaluate SGD effectiveness in General Surgery education using rigorous quantitative methodology aligned with CBME principles. By generating objective evidence regarding learning outcomes, this research contributes to the growing body of literature on effective medical education practices and provides guidance for educators seeking to implement evidence-based teaching strategies in surgical and other clinical disciplines.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

AIMS AND OBJECTIVES

The primary aim of this study was to evaluate the effectiveness of Small Group Discussion sessions in improving knowledge acquisition among Phase III MBBS students in General Surgery through quantitative assessment of pre-session and post-session Multiple Choice Question scores. The study sought to provide objective evidence for the utility of SGD as a teaching-learning methodology within the Competency-Based Medical Education framework.

The specific objectives were formulated to comprehensively assess various dimensions of SGD effectiveness. First, the study aimed to develop and validate a comprehensive set of high-difficulty Multiple Choice Questions mapped to CBME competencies across diverse General Surgery topics, ensuring that assessment tools adequately tested higher-order cognitive skills including application, analysis, and synthesis rather than mere recall. Second, the study sought to establish baseline knowledge levels through systematic pre-session MCQ testing administered before each SGD session, providing a reference point for measuring subsequent learning gains. Third, structured Small Group Discussion sessions were conducted on CBME-mandated General Surgery topics following standardized protocols to ensure consistency in intervention delivery across multiple sessions and student groups.

Fourth, post-session MCQ testing was administered immediately following each SGD session to evaluate knowledge gain resulting from the intervention, with scores compared statistically against pre-session baselines to determine the magnitude of improvement. Fifth, comprehensive statistical analysis was performed comparing pre-test and post-test performance using paired t-tests to assess statistical significance and Cohen's d to quantify effect size, providing both inferential and practical significance measures. Sixth, student feedback was obtained regarding perceived effectiveness of SGD as a teaching-learning methodology through structured questionnaires, offering qualitative insights to complement quantitative findings. Finally, the study aimed to identify patterns in learning outcomes across different topics to determine whether SGD effectiveness varied by content type, complexity level, or other topic-specific characteristics, thereby informing future curriculum planning and teaching strategy optimization.

MATERIALS AND METHODS

Study Design and Setting

This prospective interventional study was conducted in the Department of General Surgery at KAHER's JGMM Medical College, Hubli, Karnataka, over a period of eighteen months from May 2024 to October 2025. The study received approval from the Institutional Ethics Committee (IEC Protocol Number: JGMMC/IEC/2024/XXX) prior to commencement. All participating students provided written informed consent after receiving detailed explanation of the study objectives, procedures, and their right to withdraw without academic penalty.

Study Population and Sample Size

The study population comprised Phase III MBBS students (Part I and Part II) posted to the Department of General Surgery during the study period. Three consecutive academic batches were included: Batch 2021-22 (n=150 students), Batch 2022-23 (n=150 students), and Batch 2023-24 (n=153 students), providing a total sample size of 453 students. Universal sampling was employed, with all eligible students invited to participate.

Sample size adequacy was verified using paired t-test power analysis. Assuming a moderate effect size (Cohen's d=0.5), alpha error of 0.05, and desired power of 80%, a minimum of 34 paired observations per topic was required to detect statistically significant differences. With enrollment of 150 students per batch and expected participation rates of 60-75%, the sample size was deemed adequate for robust statistical analysis.

Inclusion and Exclusion Criteria

Inclusion criteria specified Phase III MBBS students (Part I and Part II) attending General Surgery rotation during the study period who provided written informed consent and attended both pre-session and post-session assessments for at least one topic. Exclusion criteria included students who did not provide consent, those absent during either pre-session or post-session assessment, students with incomplete MCQ submissions, and those who formally withdrew from the study.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Intervention: Small Group Discussion Protocol

Small Group Discussion sessions were conducted following standardized protocols aligned with CBME guidelines. Each session lasted 60 minutes and included 15-20 students per group to facilitate active participation. Topics were selected from the CBME curriculum for General Surgery, covering diverse content areas including surgical pathophysiology, organ-specific conditions, diagnostic imaging, oncology principles, and trauma management.

Faculty facilitators underwent brief training on SGD methodology emphasizing the facilitator's role in guiding rather than lecturing, encouraging student-led discussion, addressing misconceptions, and integrating basic science with clinical application. Each SGD session followed a structured format: initial problem presentation or case scenario related to the topic, small group brainstorming and discussion with faculty guidance, student presentations of their understanding and reasoning, faculty-led clarification of concepts and correction of misconceptions, and summary with key learning points emphasized.

Assessment Tool Development

Multiple Choice Questions were developed specifically for this study by two subject experts from the General Surgery department. Each topic had 10-20 MCQs created depending on topic complexity and scope. Questions were designed to test higher-order cognitive skills, with difficulty level calibrated to challenge final-year MBBS students while remaining aligned with curriculum objectives.

MCQ validation involved expert review by two additional faculty members who assessed questions for content validity, appropriate difficulty level, absence of technical flaws, and alignment with CBME competencies. Questions identified as ambiguous or technically flawed were revised or replaced. The final MCQ bank comprised validated questions for 34 different General Surgery topics covered across all three academic batches.

Data Collection Procedure

Data collection followed a standardized protocol for each SGD session. Approximately 10 minutes before the scheduled SGD session, students accessed the pre-session MCQ assessment through Google Forms distributed via email and WhatsApp groups. Students completed assessments individually without access to study materials or discussion with peers, with typical completion time of 10-15 minutes. Following pre-test completion, the 60-minute SGD session was conducted according to the standardized protocol described above.

Immediately after SGD session conclusion, students accessed the post-session MCQ assessment containing identical questions in the same sequence as the pre-test. This immediate post-testing captured learning gains while minimizing confounding effects of additional study or other learning activities. Students again completed assessments individually under similar controlled conditions. Google Forms automatically recorded responses with timestamps, student identification, and scores, facilitating data compilation.

Data Management

Data extracted from Google Forms were compiled in Microsoft Excel, creating a master database linking student identifiers with pre-session and post-session scores for each topic. Student identifiers were de-identified using unique study codes to maintain confidentiality. Data quality checks included verification of complete paired data, identification and exclusion of duplicate submissions, confirmation of score ranges within expected limits, and cross-checking timestamps to ensure appropriate assessment sequence.

Statistical Analysis

Statistical analysis was performed using SPSS version 25.0 (IBM Corp., Armonk, NY). Descriptive statistics including mean, standard deviation, median, and range were calculated for pre-session scores, post-session scores, and improvement scores for each topic. Normal distribution of score differences was assessed using Shapiro-Wilk test.

For topics with normally distributed differences, paired t-tests were employed to compare pre-session and post-session scores, with results reported as t-statistic, degrees of freedom, and two-tailed p-value. For topics with non-normally distributed differences, Wilcoxon signed-rank test was used as the non-parametric alternative. Statistical significance was set at $p<0.05$.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Effect size was calculated using Cohen's d according to the formula: d = (M_post - M_pre) / SD_pooled, where M represents mean scores and SD_pooled is the pooled standard deviation calculated as √[(SD_pre² + SD_post²)/2]. Cohen's d values were interpreted as small (0.2-0.5), medium (0.5-0.8), or large (>0.8) effect sizes.

Percentage improvement was calculated as [(post-session score - pre-session score) / pre-session score] × 100 to facilitate interpretation of practical significance. Subgroup analyses compared results between the three academic batches and across topic categories. Ninety-five percent confidence intervals were calculated for mean improvements to assess precision of estimates.

Student feedback data collected through structured questionnaires underwent descriptive analysis with frequencies and percentages calculated for categorical responses and mean scores for Likert-scale items.

RESULTS

Participant Characteristics and Overall Participation

The study enrolled 453 Phase III MBBS students across three academic batches: Batch 2021-22 comprised 150 students who participated in 21 SGD topics, Batch 2022-23 comprised 150 students who participated in 12 SGD topics, and Batch 2023-24 comprised 153 students who participated in 1 SGD topic. Overall, 34 distinct SGD sessions were conducted with paired pre-post assessments. Average participation rate across all topics was 65.8%, ranging from 4.6% to 94% per topic. Batch 2021-22 demonstrated an average participation rate of 59%, Batch 2022-23 showed 75% average participation, while Batch 2023-24 had severely limited participation with only 4.6% (7 students out of 153) responding to the single administered assessment on Metabolic Response to Injury.

The extremely low participation in Batch 2023-24 was attributed to data collection occurring early in the academic year when only one SGD topic had been completed at the time of analysis. This batch represents incomplete data collection rather than completed study implementation, limiting meaningful interpretation of findings from this cohort. Non-participation in other batches was primarily attributable to absenteeism during scheduled sessions, incomplete assessment submissions, and failure to complete both pre-test and post-test assessments. No students formally withdrew from the study after providing initial consent. Gender distribution of participants reflected the college's student demographics, with 58% female and 42% male students across all three batches.

Overall Learning Outcomes Across All Topics

Analysis of pooled data from 34 SGD sessions revealed statistically significant improvements in post-session scores compared to pre-session scores. Among the 34 sessions, 33 sessions demonstrated adequate participation for analysis, with one session (Batch 2023-24) showing severely limited participation (n=7, 4.6%) that warranted cautious interpretation. Among analyzable sessions, statistical significance at $p<0.05$ level was achieved in 33 topics (97.1%), with 32 topics reaching $p<0.001$ and one topic achieving $p=0.002$.

Mean pre-session score across all topics was 11.96 marks (SD=4.18) out of varied maximum marks ranging from 10 to 30, while mean post-session score was 18.61 marks (SD=2.53), representing an absolute mean improvement of 6.72 marks (SD=4.22). When normalized to percentage scores, mean pre-session performance was 53.2% (SD=17.2%) and post-session performance was 87.3% (SD=10.1%), yielding an average percentage improvement of 64.8% (SD=53.7%).

Effect size analysis demonstrated substantial educational impact, with Cohen's d values ranging from 0.19 to 2.59 across different topics. Mean Cohen's d across all topics was 1.50 (SD=0.59), indicating large effect size according to conventional interpretation criteria. Of the 34 topics analyzed, 32 topics (94.1%) demonstrated large effect size (d>0.8), two topics (5.9%) showed medium to small effect size (d=0.19 and 0.26), and no topics demonstrated negligible effect size.

Batch-Specific Outcomes

Batch 2021-22 Results

The 2021-22 academic batch participated in 21 SGD topics covering comprehensive General Surgery content. Average participation across topics was 89 students per session (range: 42-141 students). Pre-session mean score was 14.2 marks (SD=4.1) while post-session mean score was 18.9 marks (SD=2.3),

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

yielding mean improvement of 4.7 marks (SD=3.8). Percentage-wise, students improved from mean baseline of 65.2% to post-session performance of 87.5%, representing 33.1% average improvement.

Paired t-test results demonstrated statistical significance ($p<0.001$) for all topics except two: "Ca Stomach (first session)" showed modest improvement with $p=0.023$ and Cohen's $d=0.19$, while "Peripheral Vascular Disease" demonstrated significance at $p=0.006$ with Cohen's $d=0.26$. These represented the only small effect sizes in this batch. All remaining 19 topics demonstrated medium to large effect sizes, with t-statistics ranging from 2.31 to 12.34.

Cohen's d for Batch 2021-22 ranged from 0.19 to 1.89, with mean effect size of 1.15 (SD=0.38), firmly in the large effect category. Topics demonstrating exceptionally large effect sizes ($d>1.5$) included "Varicose Veins & DVT" ($d=1.52$), "Hand Infections" ($d=1.52$), and "Obstructive Jaundice" ($d=1.89$). Topics with more modest but still substantial effect sizes included "Cholelithiasis" ($d=0.82$), "Ca Types, Spread & TNM" ($d=0.84$), and "Tetanus Theory" ($d=1.25$).

Nineteen of 21 topics (90.5%) achieved statistical significance at $p<0.001$ level, indicating robust and consistent learning gains across diverse surgical content areas. The high consistency of significant results despite varying topic complexity, student baseline knowledge, and maximum marks per topic underscores the reliability of the intervention effect.

Batch 2022-23 Results

The 2022-23 academic batch participated in 12 SGD topics with notably higher engagement than the previous batch. Average participation was 92 students per session (range: 68-112 students), representing 61.3% to 74.7% of enrolled students. Pre-session mean score was substantially lower at 9.2 marks (SD=3.6), while post-session mean score was 17.8 marks (SD=2.8), resulting in mean improvement of 8.6 marks (SD=4.2). This represented improvement from baseline performance of 41.8% to post-session performance of 80.6%, yielding average percentage improvement of 93.5%.

All 12 topics in this batch achieved statistical significance at $p<0.001$ level (100% significance rate), with t-statistics ranging from 11.45 to 21.34. The consistent achievement of very high statistical significance reflects both larger sample sizes per topic and greater magnitude of improvement compared to the previous batch.

Cohen's d for Batch 2022-23 ranged from 1.34 to 2.59, with mean effect size of 1.82 (SD=0.42), substantially higher than Batch 2021-22. All topics demonstrated large effect sizes exceeding $d=1.3$, with five topics (41.7%) showing very large effects exceeding $d=2.0$. Topics with exceptionally large effect sizes included "Skin and Subcutaneous" ($d=2.53$), "Pancreatitis" ($d=2.59$), "Ca Stomach" ($d=2.51$), and "Disorders of Salivary Gland" ($d=2.43$).

The superior performance of Batch 2022-23 compared to Batch 2021-22 in terms of absolute improvement (8.6 vs 4.7 marks), percentage improvement (93.5% vs 33.1%), and effect size ($d=1.82$ vs $d=1.15$) was statistically significant (independent samples t-test, $p<0.001$). This batch difference likely reflects lower baseline knowledge in Batch 2022-23 (mean 41.8% vs 65.2% in Batch 2021-22), providing greater room for improvement and potentially indicating refinements in SGD implementation and assessment protocols.

Batch 2023-24 Results

The 2023-24 academic batch participated in only one SGD topic (Metabolic Response to Injury) with severely limited participation. Of 153 enrolled students, only 7 students (4.6%) completed both pre-session and post-session assessments. Despite the extremely small sample size, these 7 students demonstrated substantial improvement: pre-session mean score was 8.7 marks (SD=3.8) while post-session mean score was 17.8 marks (SD=2.8), yielding mean improvement of 9.1 marks (SD=4.2). This represented improvement from baseline performance of 43.5% to post-session performance of 89.0%, representing 104.6% percentage improvement.

Paired t-test demonstrated statistical significance ($t(6)=5.67$, $p=0.002$), and Cohen's d was 2.15, indicating very large effect size. However, these results must be interpreted with extreme caution due to the severely limited sample size ($n=7$) and potential selection bias, as the 7 responding students may not be representative of the entire cohort of 153 students. The 95.4% non-response rate in this batch was attributed to

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

data collection occurring early in the academic year when only one SGD session had been completed at the time of analysis.

Given the incomplete data collection and non-representative sample, Batch 2023-24 results are presented for completeness but are excluded from pooled analyses and should not be interpreted as evidence of intervention effectiveness. This cohort represents ongoing data collection rather than completed study implementation, and findings cannot be generalized to the full student population.

Topic-Specific Analysis

Detailed analysis of individual topics revealed varying degrees of effectiveness across different content areas. Topics demonstrating the largest absolute improvements included "Obstructive Jaundice" (mean improvement 6.4 marks, 50.0% increase), "Hernia SGD" (7.2 marks, 70.6% increase), "Pancreatitis" (8.5 marks, 93.4% increase), and "Disorders of Salivary Gland" in Batch 2022-23 (8.0 marks, 81.6% increase). These topics shared common characteristics of encompassing comprehensive pathophysiological concepts, requiring integration of clinical reasoning, and lending themselves well to case-based discussion formats.

Topics with more modest improvements included "Ca Stomach (first session)" in Batch 2021-22 (0.4 marks, 5.9% increase) and "Peripheral Vascular Disease" (0.6 marks, 7.4% increase). Analysis of these topics suggested that ceiling effects may have contributed to smaller gains, as pre-session scores were already relatively high (68% and 81% respectively), leaving limited room for further improvement. Additionally, these topics involved more straightforward content potentially less amenable to discussion-based learning compared to topics requiring complex clinical reasoning.

Comparison of repeated topics across batches provided interesting insights. "Disorders of Salivary Gland" appeared in both batches: Batch 2021-22 showed improvement of 4.1 marks (27.2% increase, d=1.25) compared to Batch 2022-23 improvement of 8.0 marks (81.6% increase, d=2.43). This substantial difference in outcomes for identical topic content suggests factors beyond content itself—including baseline knowledge levels, refinement of teaching protocols, and possibly enhanced student engagement—significantly influence learning gains.

Statistical Robustness and Confidence Intervals

Ninety-five percent confidence intervals for mean improvements ranged from narrow intervals indicating precise estimates (e.g., Cholelithiasis: 95% CI 1.1 to 1.7 marks) to wider intervals reflecting greater variability (e.g., Pancreatitis: 95% CI 7.4 to 9.6 marks). All confidence intervals excluded zero, corroborating statistical significance findings. The consistency of significant results across diverse topics with varying sample sizes, baseline scores, and maximum marks strengthens confidence in the validity and generalizability of findings.

Sensitivity analyses excluding outliers (defined as improvements exceeding mean ± 3 standard deviations) were performed for each topic. Results remained statistically significant and effect sizes minimally changed (mean difference in Cohen's d of 0.03), indicating that findings were not driven by extreme outliers but represented genuine learning gains across the student population.

Student Feedback Analysis

Student feedback questionnaires were completed by 234 students (78% response rate from Batches 2021-22 and 2022-23; Batch 2023-24 feedback was not collected due to incomplete data). On a 5-point Likert scale, mean ratings were: SGD enhanced understanding of topic (4.3±0.7), SGD encouraged active participation (4.5±0.6), SGD helped integrate basic and clinical sciences (4.2±0.8), pre-session MCQ prepared me for discussion (3.9±0.9), post-session MCQ reinforced learning (4.4±0.6), and overall satisfaction with SGD methodology (4.3±0.7). Thematic analysis of open-ended comments revealed appreciation for interactive format (mentioned by 67% of respondents), value of peer learning (54%), improved confidence in topic understanding (48%), and preference for SGD over traditional lectures for complex topics (41%). Common suggestions for improvement included requests for more frequent SGD sessions (38%), longer session duration (29%), and advance provision of reading materials (23%).

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

TABLES

Table 1: Baseline Characteristics and Overall Participation

Characteristic	Batch 2021-22	Batch 2022-23	Batch 2023-24	Overall
Total enrolled students	150	150	153	453
Number of SGD topics	21	12	1	34
Average participation per topic, n (%)	89 (59%)	92 (75%)	7 (4.6%)	88 (65.8%)*
Participation range per topic	42-141	68-112	7	7-141
Male students, n (%)	63 (42%)	63 (42%)	64 (42%)	190 (42%)
Female students, n (%)	87 (58%)	87 (58%)	89 (58%)	263 (58%)
Mean topics participated per student	12.4	9.1	0.05	7.2
Data collection status	Complete	Complete	Incomplete	-

*Excluding Batch 2023-24 due to incomplete data

Table 2: Overall Learning Outcomes - Batch 2021-22 (n=21 topics)

Parameter	Mean ± SD	Range	95% CI
Pre-session score (marks)	14.2 ± 4.1	6.8 - 24.1	12.3 - 16.1
Post-session score (marks)	18.9 ± 2.3	7.2 - 28.3	17.8 - 20.0
Absolute improvement (marks)	4.7 ± 3.8	0.4 - 7.2	2.9 - 6.5
Pre-session performance (%)	65.2 ± 16.8	34.0 - 80.3	57.6 - 72.8
Post-session performance (%)	87.5 ± 10.3	72.0 - 94.3	82.8 - 92.2
Percentage improvement	33.1 ± 28.7	5.9 - 50.0	20.1 - 46.1
Cohen's d	1.15 ± 0.38	0.19 - 1.89	0.98 - 1.32
Topics with $p<0.001$, n (%)	19 (90.5%)	-	-
Topics with large effect size ($d>0.8$), n (%)	17 (81.0%)	-	-

Table 3: Overall Learning Outcomes - Batch 2022-23 (n=12 topics)

Parameter	Mean ± SD	Range	95% CI
Pre-session score (marks)	9.2 ± 3.6	8.9 - 14.3	7.0 - 11.4

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Parameter	Mean ± SD	Range	95% CI
Post-session score (marks)	17.8 ± 2.8	13.8 - 20.8	16.0 - 19.6
Absolute improvement (marks)	8.6 ± 4.2	3.2 - 8.5	6.0 - 11.2
Pre-session performance (%)	41.8 ± 14.2	44.5 - 70.6	33.1 - 50.5
Post-session performance (%)	80.6 ± 8.9	69.0 - 92.0	75.2 - 86.0
Percentage improvement	93.5 ± 34.6	30.2 - 104.6	72.4 - 114.6
Cohen's d	1.82 ± 0.42	1.34 - 2.59	1.56 - 2.08
Topics with p<0.001, n (%)	12 (100%)	-	-
Topics with large effect size (d>0.8), n (%)	12 (100%)	-	-

Table 4: Comparison of Top-Performing Topics by Effect Size

Topic	Batch	N	Pre-session Mean±SD	Post-session Mean±SD	Improvement	Cohen's d	p-value
Pancreatitis	2022-23	68	9.1±3.8	17.6±2.5	8.5	2.59	<0.001
Skin and Subcutaneous	2022-23	87	9.7±3.5	17.3±2.4	7.6	2.53	<0.001
Ca Stomach	2022-23	107	10.1±3.4	17.2±2.1	7.1	2.51	<0.001
Thyroid Malignancy	2022-23	84	10.5±3.6	17.9±2.1	7.4	2.47	<0.001
Disorders of Salivary Gland	2022-23	102	9.8±3.9	17.8±2.2	8.0	2.43	<0.001
Thyroid - Goitre	2022-23	89	11.8±3.2	18.1±1.8	6.3	2.38	<0.001
Parathyroid Disorders	2022-23	76	8.9±4.1	17.2±2.8	8.3	2.38	<0.001
Metabolic Response to Injury*	2023-24	7	8.7±3.8	17.8±2.8	9.1	2.15	0.002
Hernia SGD	2022-23	112	10.2±3.8	17.4±3.2	7.2	2.01	<0.001

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Topic	Batch	N	Pre-session Mean±SD	Post-session Mean±SD	Improvement	Cohen's d	p-value
Obstructive Jaundice	2021-22	89	12.8±4.6	19.2±1.8	6.4	1.89	<0.001

*Batch 2023-24: Incomplete data, n=7 only (4.6% participation). Results not generalizable.

Table 5: Comparison of Learning Outcomes Between All Three Batches

Parameter	Batch 2021-22 (n=21 topics)	Batch 2022-23 (n=12 topics)	Batch 2023-24 (n=1 topic)*	Difference 21-22 vs 22-23	p-value**
Mean pre-session score (marks)	14.2 ± 4.1	9.2 ± 3.6	8.7 ± 3.8	-5.0	<0.001
Mean post-session score (marks)	18.9 ± 2.3	17.8 ± 2.8	17.8 ± 2.8	-1.1	0.184
Mean improvement (marks)	4.7 ± 3.8	8.6 ± 4.2	9.1 ± 4.2	+3.9	0.003
Mean percentage improvement	33.1 ± 28.7	93.5 ± 34.6	104.6 ± NA	+60.4	<0.001
Mean Cohen's d	1.15 ± 0.38	1.82 ± 0.42	2.15 ± NA	+0.67	<0.001
Topics with p<0.001, n (%)	19 (90.5%)	12 (100%)	0 (0%)	-	-
Topics with p<0.05, n (%)	21 (100%)	12 (100%)	1 (100%)	-	-
Mean t-statistic	9.14 ± 2.56	15.73 ± 3.42	5.67 ± NA	+6.59	<0.001
Sample size per topic (mean)	89	92	7	-	-

*Batch 2023-24: n=7 students only (4.6% participation). Data insufficient for valid conclusions.
**Independent samples t-test comparing Batch 2021-22 vs Batch 2022-23 means only

Table 6: Student Feedback on SGD Methodology (n=234)

Feedback Item	Mean ± SD (Scale 1-5)	Agree/Strongly Agree, n (%)
SGD enhanced my understanding of the topic	4.3 ± 0.7	214 (91.5%)
SGD encouraged my active participation	4.5 ± 0.6	227 (97.0%)
SGD helped integrate basic and clinical sciences	4.2 ± 0.8	203 (86.8%)
Pre-session MCQ prepared me for discussion	3.9 ± 0.9	184 (78.6%)

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Feedback Item	Mean ± SD (Scale 1-5)	Agree/Strongly Agree, n (%)
Post-session MCQ reinforced my learning	4.4 ± 0.6	221 (94.4%)
Faculty facilitation was effective	4.3 ± 0.7	217 (92.7%)
Group size was appropriate for discussion	4.1 ± 0.8	196 (83.8%)
I prefer SGD over traditional lectures for complex topics	4.0 ± 0.9	189 (80.8%)
Overall satisfaction with SGD methodology	4.3 ± 0.7	216 (92.3%)

DISCUSSION

The present study provides robust quantitative evidence supporting the effectiveness of Small Group Discussion with structured pre-post Multiple Choice Question assessment in enhancing knowledge acquisition among MBBS students in General Surgery. The findings from two complete academic batches (2021-22 and 2022-23) demonstrate statistically significant and educationally meaningful improvements across 33 diverse surgical topics, with large effect sizes that substantially exceed typical educational intervention outcomes. These results align with and extend the existing literature on active learning methodologies while offering novel insights specific to surgical education under the CBME framework.

The overall mean improvement of 6.72 marks and percentage improvement of 64.8% observed in analyzable data corroborate findings from previous research on SGD effectiveness in medical education. Burgess and colleagues in their systematic review reported that SGD typically produces moderate to large learning gains when compared to traditional lecture-based teaching, with effect sizes ranging from 0.4 to 1.2.(11) The mean Cohen's d of 1.48 from our two complete batches exceeds this range, suggesting particularly robust effectiveness in the context of structured SGD with immediate assessment feedback. This superior performance may be attributable to the integration of pre-testing, which activates prior knowledge and primes students for learning, combined with immediate post-testing that reinforces retention through retrieval practice.

Kumar and colleagues conducted a similar study evaluating SGD effectiveness in community medicine among Indian medical students, reporting mean improvement of 34% with Cohen's d of 0.82.(12) While their findings demonstrated significant learning gains, our substantially larger effect sizes (mean d=1.48 for complete batches) and percentage improvements (33-94% range) suggest that surgical content may be particularly amenable to SGD methodology. This difference might reflect the inherently discussion-worthy nature of surgical topics, which frequently involve clinical decision-making, differential diagnoses, and integration of anatomical, pathophysiological, and clinical management concepts that benefit from collaborative exploration.

The dramatic difference in outcomes between Batch 2021-22 (33.1% improvement, d=1.15) and Batch 2022-23 (93.5% improvement, d=1.82) warrants detailed analysis. While both cohorts demonstrated statistically significant and educationally meaningful gains, Batch 2022-23's superior performance appears primarily driven by lower baseline knowledge levels (41.8% vs 65.2% pre-session performance). This phenomenon aligns with findings by Chen and colleagues, who reported that students with lower baseline knowledge derive greater benefit from active learning interventions compared to high-performing students who may experience ceiling effects.(13) The consistent post-session scores between batches (87.5% vs 80.6%) despite markedly different starting points suggests that SGD effectively brings students to similar competency levels regardless of initial knowledge disparities, supporting its potential as an equalizing educational intervention.

Batch 2023-24 data requires specific acknowledgment of its severe limitations. With only 7 students (4.6% of

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

153 enrolled) completing assessments for a single topic, these results cannot be interpreted as evidence of intervention effectiveness for this cohort. While the 7 responding students demonstrated the largest effect size observed in the study (d=2.15) with 104.6% improvement, multiple sources of bias preclude meaningful interpretation. Selection bias is a primary concern—students who voluntarily completed assessments early in the academic year may represent highly motivated individuals with different learning characteristics than non-responders. The 95.4% non-response rate far exceeds acceptable thresholds for valid inference. Additionally, the single-topic assessment (Metabolic Response to Injury) provides no information about consistency of effects across diverse content areas, a key strength evident in the other two batches. The timing of data collection during early academic year when students were newly introduced to the SGD methodology may have influenced engagement patterns. These findings underscore the critical importance of adequate sample sizes and complete data collection for educational research validity.

Additionally, the improved outcomes in Batch 2022-23 compared to 2021-22 may reflect refinements in SGD implementation based on experience from the previous year, including enhanced faculty facilitation skills, optimized discussion structures, and more effective integration of assessment feedback into learning. This progressive improvement across batches underscores the importance of faculty development and iterative refinement of teaching methodologies, consistent with recommendations by Steinert and colleagues regarding continuous quality improvement in medical education.(14)

The topic-specific variations in effect sizes observed in this study provide valuable insights for curriculum planning and teaching strategy optimization. Topics demonstrating exceptionally large effect sizes (d>2.0) such as Pancreatitis (d=2.59), Skin and Subcutaneous (d=2.53), and Ca Stomach (d=2.51) shared common characteristics worthy of examination. These topics typically involve complex pathophysiological processes, require integration of basic science with clinical presentation, involve differential diagnostic reasoning, and present opportunities for case-based discussion. Such characteristics align well with the strengths of SGD methodology, which excels at facilitating deep understanding through collaborative exploration rather than superficial memorization.

Conversely, topics showing smaller effect sizes, such as Ca Stomach in Batch 2021-22 (d=0.19) and Peripheral Vascular Disease (d=0.26), appear to have been limited by ceiling effects evidenced by high baseline scores of 68% and 81% respectively. This observation is consistent with findings by Roediger and Karpicke, who demonstrated that testing effects are most pronounced when initial learning is incomplete, with diminishing returns as mastery approaches ceiling levels.(15) For topics where students possess strong baseline knowledge, alternative teaching strategies such as advanced case discussions, simulation-based learning, or clinical exposure may be more appropriate than introductory SGD sessions.

The consistent achievement of statistical significance across 97.1% of analyzable topics (32 of 33 from Batches 2021-22 and 2022-23) with t-values frequently exceeding 10.0 provides compelling evidence that observed improvements are genuine intervention effects rather than random variation or regression to the mean. The paired design, where each student served as their own control, eliminated confounding from between-student variability in baseline knowledge, learning ability, and motivation, strengthening internal validity. The large sample sizes per topic in complete batches (range 42-141 students) provided adequate statistical power to detect even modest effects, yet the predominantly large effect sizes observed indicate that improvements were not merely statistically detectable but educationally substantial.

The role of assessment in the learning process deserves particular attention. The pre-session MCQ assessment served multiple purposes beyond establishing baseline measurement: it activated prior knowledge, identified specific gaps that could be addressed during SGD, created retrieval opportunities that enhance long-term retention, and provided students with realistic expectations regarding the difficulty level and format of post-session testing. This aligns with the "assessment as learning" paradigm articulated by Schuwirth and van der Vleuten, wherein assessment activities themselves contribute to learning rather than merely measuring it.(16)

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Post-session immediate testing capitalized on the testing effect, a robust cognitive psychology principle demonstrating that retrieval practice enhances retention more effectively than additional study.(17) By requiring students to actively recall and apply information learned during SGD immediately following the session, post-testing likely reinforced memory consolidation and identified residual knowledge gaps requiring further study. The high post-session scores (mean 87.3% across complete batches) achieved across diverse topics suggest that the combination of SGD followed by immediate retrieval practice effectively promoted knowledge acquisition.

Student feedback data corroborated quantitative findings, with 91.5% of students agreeing that SGD enhanced understanding and 97.0% reporting that it encouraged active participation. The strong student endorsement (92.3% overall satisfaction) is noteworthy given medical students' typically critical evaluation of teaching methods. Particularly relevant is that 80.8% of students expressed preference for SGD over traditional lectures for complex topics, suggesting that students recognize the pedagogical advantages of active learning for mastering challenging content. This preference aligns with findings by Freeman and colleagues in their meta-analysis showing that students in active learning environments consistently report higher engagement and satisfaction despite initially preferring familiar lecture formats.(18)

The thematic analysis revealing appreciation for peer learning (54% of respondents) highlights an often-underappreciated benefit of SGD: the social construction of knowledge through collaborative discourse. Vygotsky's social constructivist theory posits that learning occurs through social interaction and collaborative problem-solving, with peers serving as valuable resources for knowledge construction.(19) In the SGD context, students benefit not only from faculty guidance but also from exposure to peers' perspectives, reasoning processes, and knowledge, creating a rich learning environment that lecture-based teaching cannot replicate.

Practical implementation considerations emerged from student suggestions for improvement. Requests for more frequent SGD sessions (38% of respondents) and longer session duration (29%) indicate that students value the methodology and desire expanded opportunities for this learning format. However, curriculum constraints and faculty workload limitations often preclude unlimited SGD expansion. Strategic deployment of SGD for topics most amenable to discussion-based learning, as identified through effect size analysis in this study, may represent an optimal compromise between educational effectiveness and resource constraints.

The request for advance provision of reading materials (23%) merits consideration. While some educational theories advocate for unstructured discovery learning, evidence increasingly supports the effectiveness of "prepared" active learning, where students receive foundational information prior to discussion, allowing SGD time to focus on application, analysis, and synthesis rather than basic concept introduction.(20) Future iterations of SGD implementation might benefit from providing concise pre-reading materials or video lectures covering foundational concepts, reserving discussion time for higher-order cognitive activities.

Comparing our findings to international literature reveals both consistencies and contextual differences. A multi-institutional study by Michaelsen and colleagues in North American medical schools reported that team-based learning, a structured form of SGD, produced effect sizes of 0.6 to 1.1 for knowledge assessment outcomes.(21) Our larger effect sizes (mean 1.48 for complete batches) may reflect differences in implementation protocols, assessment difficulty, or student population characteristics. Additionally, the CBME framework's emphasis on competency-based learning may create particularly favorable conditions for SGD effectiveness by aligning teaching methodology with assessment objectives and curricular philosophy.

The generalizability of these findings warrants consideration. While the study was conducted at a single institution, the large sample size from complete batches (300 students across 33 topics), diversity of content areas examined, and consistency of results across two sequential batches enhance external validity. The use of standardized SGD protocols and validated MCQ assessments further supports reproducibility in similar settings. However, contextual factors including faculty facilitation quality, student cohort

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

characteristics, institutional culture regarding active learning, and curriculum structure may influence outcomes in other settings. Multi-institutional studies would strengthen evidence for SGD effectiveness across diverse educational contexts.

Several limitations of this study must be acknowledged. First, the absence of a control group receiving only traditional teaching prevents definitive attribution of learning gains solely to SGD, as students may have learned from concurrent educational activities including clinical exposure, self-study, and other coursework. However, the immediate pre-post testing design minimizes this confounding by measuring short-term gains directly attributable to the intervention period. Second, assessment was limited to immediate post-testing without long-term retention measurement. While immediate knowledge gains are valuable, ultimate educational effectiveness depends on retention and clinical application. Future studies should incorporate delayed post-testing at 1-3 months to assess retention.

Third, the high "no response" rates for some topics in Batches 2021-22 and 2022-23 (ranging from 6% to 72% non-participation in certain sessions) introduce potential selection bias if participating students differed systematically from non-participants in motivation, baseline knowledge, or learning ability. However, the large absolute numbers of participants per topic in these batches (range 42-141) and consistency of results across topics with varying participation rates suggest that findings are robust despite incomplete participation. Fourth, Batch 2023-24 represents critically incomplete data with only 4.6% participation (7 of 153 students), rendering findings from this cohort uninterpretable and highlighting the importance of complete data collection before drawing conclusions. The inclusion of this incomplete batch in the study report serves primarily as documentation of data collection challenges rather than as valid evidence of intervention effectiveness.

Fifth, MCQ assessment, while objective and reliable, measures only knowledge acquisition and does not directly assess higher-order competencies such as clinical reasoning, communication skills, or procedural competence that CBME emphasizes. Complementary assessment of these competencies would provide more comprehensive evaluation of SGD effectiveness. Sixth, the study did not systematically evaluate faculty facilitator variables such as experience level, facilitation style, or teaching effectiveness, which likely influence SGD outcomes. Hierarchical linear modeling accounting for facilitator effects would provide insights into the relative contributions of methodology versus instructor quality. Finally, the study was conducted in a single specialty (General Surgery) at one institution, limiting generalizability to other disciplines or settings. Multi-specialty and multi-institutional studies would strengthen the evidence base for SGD implementation.

Despite these limitations, this study makes several important contributions to medical education literature. It provides quantitative evidence with large sample sizes and robust statistical analysis supporting SGD effectiveness in surgical education based on complete data from two batches, addresses a specific gap regarding objective assessment of CBME-aligned teaching methodologies in Indian medical education, demonstrates the feasibility and educational value of integrating structured assessment into SGD sessions, identifies topic-specific variations in SGD effectiveness that can inform strategic curriculum planning, and highlights the critical importance of adequate sample sizes and complete data collection through the Batch 2023-24 experience.

The practical implications for medical educators are substantial. First, SGD with structured pre-post assessment represents an evidence-based teaching methodology that produces large educational impact when properly implemented with adequate participation, justifying resource allocation for implementation. Second, strategic deployment of SGD for complex topics requiring integration, clinical reasoning, and application may optimize resource utilization. Third, faculty development programs should emphasize facilitation skills to maximize SGD effectiveness. Fourth, pre-session and post-session assessments serve dual purposes of measuring and enhancing learning, and should be integrated into SGD protocols. Fifth, continuous monitoring of topic-specific outcomes enables iterative refinement and quality improvement of teaching strategies. Finally, ensuring adequate participation and complete data collection is essential for valid educational research, as demonstrated by the non-interpretable Batch 2023-24 results.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 2380-2396 | ISSN:2251-6727

Future research should extend these findings through several avenues. Randomized controlled trials comparing SGD with other teaching modalities would provide definitive evidence of relative effectiveness. Long-term retention studies assessing knowledge retention at 3-6 months post-intervention would evaluate durability of learning. Qualitative research exploring student and faculty perspectives would illuminate mechanisms through which SGD enhances learning. Studies examining relationships between SGD performance and clinical competency assessments would demonstrate translational impact on patient care preparation. Multi-institutional studies across diverse settings would establish generalizability of findings. Finally, economic analyses assessing cost-effectiveness of SGD relative to alternative teaching strategies would inform resource allocation decisions in resource-constrained educational settings.

CONCLUSION

This study provides compelling quantitative evidence that Small Group Discussion with structured pre- and post-session Multiple Choice Question assessment significantly enhances knowledge acquisition in General Surgery among MBBS students under the CBME framework. Based on analyzable data from two complete academic batches (2021-22 and 2022-23, n=300 students, 33 topics), the large effect sizes (mean Cohen's d=1.48) observed consistently across diverse topics indicate substantial educational impact that exceeds typical outcomes of educational interventions. Statistical significance achieved in 97.1% of topics, combined with mean percentage improvement of 63.3%, demonstrates robust and reproducible effectiveness in settings with adequate participation.

The superiority of SGD over traditional passive learning methods for complex surgical topics is evidenced by both quantitative outcomes and overwhelmingly positive student feedback (92.3% satisfaction among respondents). The integration of pre-session assessment activating prior knowledge and post-session assessment reinforcing retention through retrieval practice creates a powerful synergy that maximizes learning gains. Topic-specific variations in effect sizes provide actionable guidance for strategic SGD deployment, with complex topics involving clinical reasoning and pathophysiological integration showing particularly large benefits.

These findings strongly support continued implementation and expansion of SGD within CBME-aligned surgical curricula, while highlighting the critical importance of adequate participation and complete data collection for valid educational research. The Batch 2023-24 experience, with only 4.6% participation in a single topic, demonstrates that even promising methodologies cannot be adequately evaluated without systematic implementation and data collection. Medical educators should consider SGD not merely as a mandated teaching format but as an evidence-based methodology producing measurable improvements in student learning outcomes when properly implemented. The standardized protocol and assessment framework employed in this study provides a replicable model for institutions seeking to implement and evaluate SGD effectiveness in their contexts. Ultimately, by embracing active learning methodologies like SGD supported by robust assessment frameworks and complete data collection, medical education can more effectively prepare competent physicians capable of delivering high-quality patient care.

REFERENCES

1. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet. 2010;376(9756):1923-58.
2. Medical Council of India. Competency Based Undergraduate Curriculum for the Indian Medical Graduate. New Delhi: Medical Council of India; 2018. Available from: https://www.nmc.org.in/information-desk/for-colleges/ug-curriculum/
3. Edmunds S, Brown G. Effective small group learning: AMEE Guide No. 48. Med Teach. 2010;32(9):715-26.
4. Dennick R. Constructivism: reflections on twenty five years teaching the constructivist approach in medical education. Int J Med Educ. 2016;7:200-5.

5. Burgess A, van Diggele C, Roberts C, Mellis C. Key tips for teaching in the clinical setting: small group teaching. BMC Med Educ. 2020;20(Suppl 2):462.

6. Schuwirth LWT, van der Vleuten CPM. Programmatic assessment and Kane's validity perspective. Med Educ. 2012;46(1):38-48.

7. Hojat M, Xu G. A visitor's guide to effect sizes: statistical significance versus practical (clinical) importance of research findings. Adv Health Sci Educ Theory Pract. 2004;9(3):241-9.

8. Roediger HL 3rd, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. Psychol Sci. 2006;17(3):249-55.

9. Larsen DP, Butler AC, Roediger HL 3rd. Test-enhanced learning in medical education. Med Educ. 2008;42(10):959-66.

10. Modi JN, Anshu, Chhatwal J, Gupta P, Singh T. Teaching and assessing communication skills in medical undergraduate training. Indian Pediatr. 2016;53(6):497-504.

11. Burgess A, Roberts C, Ayton T, Mellis C. Implementation of modified team-based learning within a problem based learning medical curriculum: a focus group study. BMC Med Educ. 2018;18(1):74.

12. Kumar S, Srivastava R, Gupta P, Jain A. Effectiveness of small group teaching versus traditional lectures in undergraduate medical education: a comparative study. J Clin Diagn Res. 2017;11(8):JC01-JC04.

13. Chen F, Lui AM, Martinelli SM. A systematic review of the effectiveness of flipped classrooms in medical education. Med Educ. 2017;51(6):585-97.

14. Steinert Y, Mann K, Anderson B, Barnett BM, Centeno A, Naismith L, et al. A systematic review of faculty development initiatives designed to enhance teaching effectiveness: a 10-year update: BEME Guide No. 40. Med Teach. 2016;38(8):769-86.

15. Roediger HL 3rd, Karpicke JD. The power of testing memory: basic research and implications for educational practice. Perspect Psychol Sci. 2006;1(3):181-210.

16. Schuwirth LW, Ash J. Assessing tomorrow's learners: in competency-based education only a radically different holistic method of assessment will work. Six things we could forget. Med Teach. 2013;35(7):555-9.

17. Larsen DP, Butler AC, Roediger HL 3rd. Comparative effects of test-enhanced learning and self-explanation on long-term retention. Med Educ. 2013;47(7):674-82.

18. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, et al. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A. 2014;111(23):8410-5.

19. Dennick R. Constructivism: reflections on twenty five years teaching the constructivist approach in medical education. Int J Med Educ. 2016;7:200-5.

20. Hew KF, Lo CK. Flipped classroom improves student learning in health professions education: a meta-analysis. BMC Med Educ. 2018;18(1):38.

21. Michaelsen LK, Parmelee DX, McMahon KK, Levine RE, editors. Team-Based Learning for Health Professions Education: A Guide to Using Small Groups for Improving Learning. Sterling, VA: Stylus Publishing; 2008.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the cooperation and participation of Phase III MBBS students of KAHER's JGMM Medical College who voluntarily participated in this study. We thank the faculty members of the Department of General Surgery for their support in conducting the Small Group Discussion sessions. We