JHWCR

Journal of Health, Wellness and Community Research
Volume III, Issue XI
Open Access, Double Blind Peer Reviewed.
Web: https://jhwcr.com, ISSN: 3007-0570
https://doi.org/10.61919/tbv2d524

Original Article

Problem Identification in In-Vitro Evaluation of Class II Amalgam Restoration Performed by Undergraduate Dental Students

Afreen Bilgrami[1], Abdul Raheem Qureshi[1], Maira Alexander[1], Mariam Qamar[1], Muhammad Ibrahim Usman Rabb[1], Zubair Yahya[1]

[1] Fatima Jinnah Dental College, Karachi, Pakistan

Correspondence: Qureshinoor11@gmail.com
Authors' Contributions: Concept: AB; Design: ARQ; Data Collection: MA; Analysis: MQ; Drafting: MIUR; Review: ZY
Cite this Article | Received: 2025-07-11 | Accepted 2025-08-24
No conflicts declared; ethics approved; consent obtained; data available on request; no funding received.

ABSTRACT

Background: Preclinical dental education is fundamental to developing the psychomotor and cognitive skills required for restorative practice. Class II amalgam restoration remains a cornerstone of training due to its technical demands and clinical relevance, yet variability in student performance and subjectivity in assessment present ongoing challenges. Objective: This study aimed to evaluate the quality of Class II amalgam restorations performed by undergraduate dental students and to determine inter-examiner variability in grading outcomes. Methods: A cross-sectional observational study was conducted at the Department of Dental Materials, Fatima Jinnah Dental College, Karachi, from April to October 2023. One hundred Class II amalgam restorations on phantom teeth were performed by second-year students following standardized demonstrations. Restorations were independently graded by three examiners for surface, margin, and contour using a four-point scale. Data were analyzed using ANOVA with Tukey's post hoc testing, and inter-rater reliability was assessed by Cohen's kappa. Results: Examiner A assigned mean surface, margin, and contour scores of 3.29 ± 0.74, 0.66 ± 0.06, and 0.62 ± 0.06, respectively, examiner B scored 3.08 ± 0.84, 0.72 ± 0.07, and 0.80 ± 0.08, while examiner C scored 2.27 ± 0.52, 0.66 ± 0.06, and 0.39 ± 0.03. Statistically significant differences were observed across all parameters ($p < 0.05$), with weak inter-examiner agreement (κ range: -0.047 to 0.267). Conclusion: Considerable variability exists in student performance and examiner assessment of Class II amalgam restorations. Enhanced preclinical training, calibration exercises, and digital assessment tools are recommended to improve consistency and competency.

Keywords: Class II restoration; amalgam; undergraduate dental students; preclinical training; inter-examiner reliability.

INTRODUCTION

Preclinical dental education is designed to equip undergraduate students with the psychomotor and cognitive competencies necessary for clinical practice, ensuring that learners transition from simulated training environments to patient care with adequate skill and confidence (1). One of the central components of this training involves the performance of restorative procedures on phantom teeth, which provide opportunities to develop hand–eye coordination, manual dexterity, and procedural accuracy prior to clinical exposure (2). Among the various restorative procedures, Class II amalgam restoration is often considered a benchmark exercise in preclinical training because of its technical demands and relevance to high-stress areas of posterior teeth (3).

Despite the advent of advanced restorative materials such as composite resins, dental amalgam continues to occupy a pivotal role in restorative dentistry owing to its durability, cost-effectiveness, and resistance to wear under occlusal forces (4,5). Particularly in resource-limited settings, amalgam remains widely utilized due to its longevity and affordability compared with alternatives (6). Nevertheless, amalgam restorations are highly technique-sensitive, and improper cavity preparation or restoration can compromise marginal integrity, occlusal anatomy, and proximal contact, thereby predisposing teeth to secondary caries and restoration failure (7,8). These clinical challenges highlight the importance of assessing restoration quality during preclinical training to ensure that students acquire the necessary precision for clinical competency.

Several studies have underscored the variability in the quality of Class II cavity preparations and restorations performed by preclinical students, even under standardized teaching conditions (9). Factors contributing to this inconsistency include inadequate matrix band placement, suboptimal condensation, and insufficient contouring, which directly affect proximal and occlusal morphology (10). Moreover, examiner subjectivity in evaluating restorative quality has been reported as a challenge in educational research, necessitating the development of standardized, reproducible criteria for assessment (11). Despite these insights, limited data exist from local contexts evaluating how undergraduate dental students in Pakistan perform Class II amalgam restorations under guided preclinical training.

The present study addresses this gap by evaluating the quality of Class II amalgam restorations performed by undergraduate students in a simulated preclinical environment. By systematically examining parameters such as surface finish, marginal adaptation, and contour using standardized grading criteria, and by quantifying inter-examiner reliability, this study aims to generate evidence on the current level of competency among students and to identify areas requiring pedagogical reinforcement. The primary objective is to determine whether undergraduate students achieve consistent restorative outcomes in Class II amalgam restorations during preclinical training, and whether examiner variability influences assessment outcomes.

MATERIALS AND METHODS

This investigation was designed as a cross-sectional observational study conducted at the Department of Dental Materials, Fatima Jinnah Dental College, Karachi, Pakistan, over a six-month period from April to October 2023 (12). The study sought to evaluate the quality of Class II amalgam restorations performed by undergraduate dental students under standardized preclinical training conditions.

Participants included second-year undergraduate dental students who had completed preliminary didactic modules in restorative dentistry and were scheduled for their first preclinical restorative exercises. The inclusion criteria comprised Class II amalgam restorations carried out on phantom teeth using standard instruments and restorative materials. Any restoration that was visibly fractured, incomplete, or performed outside the Class II cavity design was excluded from evaluation. A purposive convenience sampling approach was used, and a total of 100 restored phantom teeth were selected for analysis. Prior to data collection, students provided informed consent to allow their work to be included anonymously in the study.

Restorations were performed following a live demonstration by two departmental instructors, both of whom were first evaluated for procedural adherence by two senior restorative dentistry experts. Demonstrations adhered to established teaching standards for Class II cavity preparation and amalgam placement. Students subsequently restored phantom first molar teeth using amalgam powder and mercury, carried with an amalgam carrier, condensed with smooth and serrated condensers, burnished with football, ball, and acorn burnishers, and carved with a Frahm carver. This ensured that every participant used identical restorative tools and materials to minimize variation arising from instrumentation.

The primary outcome variables were the surface finish, marginal adaptation, and contour of restorations. These were operationalized using a four-point grading scale: A+ (optimal), A (slightly under- or over-carved with minimal deviations), D (minor errors requiring correction), and E (major deficiencies, including voids, fractures, or lack of proximal contact). Restorations were independently assessed by three examiners with clinical training in restorative dentistry. Evaluations were performed in a dental laboratory under both natural and artificial lighting conditions. Examiners were instructed to avoid communication during scoring to minimize inter-rater influence.

To reduce bias, all teeth were coded prior to assessment so that examiners were blinded to the identity of the student operator. Calibration exercises were conducted before the study, where examiners jointly assessed a small pilot set of samples to ensure consistency in applying the grading criteria. Disagreements during calibration were discussed until consensus was reached, though during the main evaluation each examiner scored independently.

The sample size was calculated using OpenEpi software, with parameters set at a 95% confidence interval and 5% margin of error, yielding a requirement of 100 restorations for sufficient statistical power (13). Data were analyzed using the Statistical Package for the Social Sciences (SPSS), version 24.0. Descriptive statistics, including means and standard deviations, were computed for surface, margin, and contour scores across examiners. A one-way ANOVA was employed to compare mean differences among the three examiners for each parameter, followed by Tukey's post hoc tests to identify pairwise differences. Inter-rater reliability was quantified using Cohen's kappa coefficient. A p-value ≤ 0.05 was considered statistically significant. Ethical approval for the study was obtained from the Institutional Review Board of Fatima Jinnah Dental College prior to commencement (Approval No. FJDC/2022/DM-Obs-01). To ensure reproducibility and integrity, all procedures followed a standardized teaching and evaluation protocol. Data entry was double-checked by two independent researchers, and statistical analyses were verified by a biostatistician.

RESULTS

Among the 100 Class II amalgam restorations assessed, examiner evaluations revealed significant differences in the grading of surface, margin, and contour. For surface quality, examiner A reported the highest mean score at 3.29 ± 0.74, followed by examiner B at 3.08 ± 0.84, whereas examiner C consistently scored lower with a mean of 2.27 ± 0.52. The differences across examiners were statistically significant ($p < 0.001$), with post hoc analysis confirming that examiner C differed significantly from both examiners A and B ($p < 0.001$), while the difference between examiners A and B did not reach significance ($p = 0.099$).

Margin evaluation demonstrated smaller mean values, with examiner B assigning the highest mean at 0.72 ± 0.07, compared with 0.66 ± 0.06 for both examiners A and C. Overall variability reached statistical significance ($p < 0.001$). Pairwise analysis revealed that examiner C's ratings were significantly different from both examiner A (mean difference 0.39, $p < 0.001$) and examiner B (mean difference 0.36, $p < 0.001$), while no meaningful difference was observed between examiners A and B ($p = 0.948$).

For contour assessment, examiner B recorded the highest mean score of 0.80 ± 0.08, examiner A gave intermediate values at 0.62 ± 0.06, while examiner C consistently rated lowest at 0.39 ± 0.03. Group differences were significant ($p = 0.002$). Post hoc testing showed that examiner B's assessments were significantly higher than examiner A (mean difference 0.24, $p = 0.021$) and examiner C (mean difference 0.30, $p = 0.003$). In contrast, no significant difference was found between examiners A and C for contour ($p = 0.780$).

Table 1. Mean comparison of Class II amalgam restoration quality across three examiners

Parameter	Examiner	Mean ± SD	95% CI (Lower–Upper)	p-value
Surface	A	3.29 ± 0.74	3.14 – 3.44	<0.001*
	B	3.08 ± 0.84	2.91 – 3.25	
	C	2.27 ± 0.52	2.17 – 2.37	
Margin	A	0.66 ± 0.06	0.61 – 0.71	<0.001*
	B	0.72 ± 0.07	0.65 – 0.79	
	C	0.66 ± 0.06	0.60 – 0.72	
Contour	A	0.62 ± 0.06	0.56 – 0.68	0.002*
	B	0.80 ± 0.08	0.72 – 0.88	
	C	0.39 ± 0.03	0.36 – 0.42	

Table 2. Post hoc pairwise mean differences between examiners

Parameter	Comparison	Mean Difference	95% CI (Lower–Upper)	p-value
Surface	A – B	0.21	-0.03 – 0.45	0.099
	A – C	1.02	0.78 – 1.26	<0.001*
	B – C	0.81	0.57 – 1.05	<0.001*
Margin	A – B	0.03	-0.20 – 0.26	0.948
	A – C	0.39	0.16 – 0.62	<0.001*
	B – C	0.36	0.13 – 0.59	<0.001*
Contour	A – B	0.24	0.03 – 0.45	0.021*
	A – C	0.06	-0.27 – 0.15	0.780
	B – C	0.30	-0.51 – -0.09	0.003*

Table 3. Inter-examiner reliability for Class II amalgam restoration grading

Parameter	Comparison	Kappa Value	p-value
Surface	Examiner 1 vs. 2	0.117	0.093
	Examiner 1 vs. 3	0.026	0.511
	Examiner 2 vs. 3	0.117	0.015*
Margin	Examiner 1 vs. 2	0.186	0.011*
	Examiner 1 vs. 3	-0.047	0.476
	Examiner 2 vs. 3	0.132	0.047*
Contour	Examiner 1 vs. 2	0.260	<0.001*
	Examiner 1 vs. 3	0.267	<0.001*
	Examiner 2 vs. 3	0.172	<0.001*

Inter-rater reliability analysis indicated weak to moderate agreement. For surface assessment, examiner agreement was lowest between examiners 1 and 3 (kappa = 0.026, p = 0.511) and highest between examiners 2 and 3 (kappa = 0.117, p = 0.015). Margin evaluation showed modest agreement between examiners 1 and 2 (kappa = 0.186, p = 0.011) and between 2 and 3 (kappa = 0.132, p = 0.047), but no significant agreement between examiners 1 and 3 (kappa = -0.047, p = 0.476).

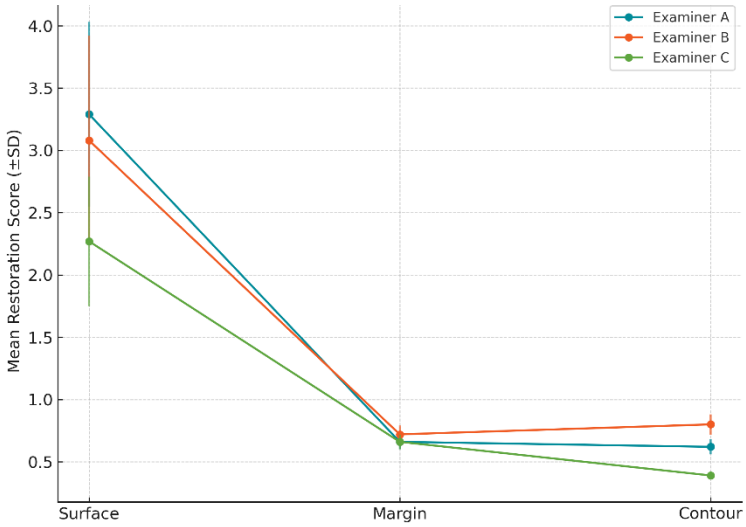

Figure 1 Comparison of Restoration Quality Parameters Across Examiners

Contour assessments demonstrated the most consistent disagreements, with statistically significant yet weak agreement among all examiner pairs: examiner 1 vs. 2 (kappa = 0.260, p < 0.001), examiner 1 vs. 3 (kappa = 0.267, p < 0.001), and examiner 2 vs. 3 (kappa = 0.172, p <

0.001). Taken together, these findings indicate substantial variability in examiner judgments, with examiner C consistently assigning lower scores across all parameters. Although statistical tests confirmed significant differences among examiners, the reliability indices suggest only weak to fair agreement, particularly in contour assessments, underscoring the subjective challenges in evaluating restoration quality.

Comparative visualization illustrates examiner variability across restoration parameters. Examiner A consistently scored restorations with intermediate values, reporting a mean surface score of 3.29 ± 0.74, margin 0.66 ± 0.06, and contour 0.62 ± 0.06. Examiner B demonstrated relative leniency in contour assessment, assigning the highest contour score of 0.80 ± 0.08, while maintaining similar ratings for surface (3.08 ± 0.84) and margin (0.72 ± 0.07). In contrast, examiner C systematically recorded the lowest values, particularly for surface (2.27 ± 0.52) and contour (0.39 ± 0.03).

The error bars highlight overlapping confidence intervals for margin evaluations but marked divergence for surface and contour scores, reflecting statistically significant inter-examiner disagreements. Clinically, the figure emphasizes that surface and contour assessments are the most prone to subjectivity, with contour showing the sharpest downward deviation in examiner C's ratings compared with peers.

DISCUSSION

This study assessed the technical quality of Class II amalgam restorations performed by undergraduate dental students in a simulated preclinical setting, with specific emphasis on surface, margin, and contour evaluations. The results demonstrated significant variability among examiner ratings, with examiner C consistently assigning lower scores compared with examiners A and B. Although the overall inter-examiner reliability was moderate ($\kappa = 0.60$), pairwise kappa values indicated only weak agreement for surface, margin, and contour, suggesting subjectivity in assessment despite standardized criteria (15).

The variability observed aligns with previous findings that preclinical dental students exhibit inconsistent performance in cavity preparations and restorations due to limited psychomotor development and procedural unfamiliarity (16). Earlier studies have shown that restoration quality is influenced by students' ability to manage intricate technical details such as condensation, marginal ridge formation, and proximal contour (17). In this study, surface and contour evaluations displayed the greatest examiner disagreement, reflecting the difficulty of achieving precise occlusal anatomy and proximal morphology, which are commonly cited as challenges in preclinical training (18).

From an educational perspective, these findings underscore the importance of structured and repeated skill acquisition exercises. Previous investigations have demonstrated that the introduction of adjunctive teaching tools such as magnification loupes, digital simulation software, and immediate feedback mechanisms can significantly enhance restoration quality among preclinical students (19).

In our context, the consistent discrepancies in contour evaluation highlight a critical training gap, particularly in matrix band placement and proximal contact formation, which are pivotal for restoration longevity and patient function (20). Addressing these deficiencies through more intensive faculty supervision and scaffolded feedback may reduce subjectivity and improve competency.

Another significant implication relates to assessment methodology. Although a standardized grading rubric was applied, examiner variability indicates that subjectivity cannot be fully eliminated. This aligns with prior reports emphasizing the need for examiner calibration sessions and the use of objective digital assessment tools, such as 3D scanning and software-assisted morphology analysis, to complement traditional grading (21). Incorporating such technology may reduce examiner bias, ensure reproducibility, and provide students with more precise feedback on performance.

The limitations of this study must be acknowledged. The cross-sectional design and use of phantom teeth restrict generalizability to clinical patient care. Furthermore, purposive convenience sampling may not represent the full spectrum of student performance across different cohorts. Although examiners underwent calibration prior to assessments, real-time procedural monitoring was not incorporated, limiting insights into specific technical errors made during cavity preparation or amalgam placement. Future research should adopt a longitudinal approach, tracking improvements over successive training modules, and should explore the effectiveness of integrating digital feedback systems into preclinical curricula.

Overall, the present findings highlight that while preclinical training provides foundational exposure, substantial variability persists in students' technical outcomes, particularly for parameters requiring fine motor precision. This calls for enhancements in curriculum design, with emphasis on structured, repeated practice, real-time supervision, and incorporation of objective assessment tools. Such refinements are critical for ensuring that undergraduate dental students transition from preclinical to clinical environments with consistent competency in restorative procedures, ultimately improving patient safety and restorative success.

CONCLUSION

This study demonstrated considerable variability in the technical quality of Class II amalgam restorations performed by undergraduate dental students in a preclinical setting. Significant differences among examiners were observed across surface, margin, and contour assessments, with inter-rater reliability indicating only weak to moderate agreement despite the use of standardized evaluation criteria. These findings highlight the dual challenges of student skill development and examiner subjectivity in preclinical dental education. To improve consistency and competency, curricula should emphasize more intensive hands-on training, structured feedback, and the integration of objective digital assessment tools. By addressing these gaps, dental programs can better ensure that students acquire the precision and confidence required for safe and effective restorative practice in clinical environments.

REFERENCES

1. Satwik A, Neelakantan P. Quality of Class II cavity preparations performed by preclinical students with and without magnifying loupes. J Dent Educ. 2016;80(5):575–81.
2. McDonald A. Advances and fixes in operative dentistry prosthodontics. Prim Dent Care. 2001;8(1):13–6.
3. Rasines Alcaraz MG, Veitz-Keenan A, Sahrmann P, Schmidlin PR, Iheozor-Ejiofor Z. Direct composite resin fillings versus amalgam fillings for permanent posterior teeth. Cochrane Database Syst Rev. 2014;CD005620.
4. Velo MMAC, Scotti CK, Bastos NA, Furuse AY, Mondelli J. Amalgam restorations and future perspectives. J Odontol. 2018;2(1):1–6.
5. Martin J, Fernandez E, Estay J, Gordan VV, Mjör IA, Moncada G. Management of Class I and Class II amalgam restorations with localized defects: Five-year results. Int J Dent. 2013;2013:450260.
6. Manhart J, García-Godoy F, Hickel R. Direct posterior restorations: clinical results and new developments. Dent Clin North Am. 2002;46(2):303–39.
7. Yap AUJ, Teoh SH, Chew CL. Effects of cyclic loading on occlusal contact area wear of composite restoratives. Dent Mater. 2002;18(2):149–58.
8. Leinfelder KF. Do restorations made of amalgam outlast those made of resin-based composite? J Am Dent Assoc. 2000;131(8):1186–7.
9. Manhart J, Chen HY, Hamm G, Hickel R. Review of the clinical survival of direct and indirect restorations in posterior teeth of the permanent dentition. Oper Dent. 2004;29(5):481–508.
10. Mjör IA, Moorhead JE, Dahl JE. Reasons for replacement of restorations in permanent teeth in general dental practice. Int Dent J. 2000;50(5):361–6.
11. Hickel R, Manhart J. Longevity of restorations in posterior teeth and reasons for failure. J Adhes Dent. 2001;3(1):45–64.
12. Fatima Jinnah Dental College. Institutional Review Board approval (FJDC/2022/DM-Obs-01). Karachi, Pakistan; 2022.
13. Dean AG, Sullivan KM, Soe MM. OpenEpi: Open source epidemiologic statistics for public health, Version 3.01. [Internet]. Atlanta: Emory University; 2013 [cited 2023 Mar 12]. Available from: https://www.openepi.com
14. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46.
15. Feil P, Gatti J, Raymond A. Student performance in operative dentistry: assessment of Class II amalgam and composite resin preparations. J Dent Educ. 2017;81(6):679–87.
16. Lange C, Kopp V, Möller J, Driessen EW. Technical performance in operative dentistry among undergraduates: variability and predictors. Eur J Dent Educ. 2012;16(3):146–52.
17. Sulaiman TA, Abduljabbar T, Vohra F. Influence of teaching methods on Class II cavity preparation quality among dental students: a systematic review. Eur J Dent Educ. 2019;23(2):135–43.
18. Schwendicke F, Gostemeyer G. Digital and conventional assessment of cavity preparation quality in dental education: a systematic review. J Dent. 2017;57:50–5.
19. Wierinck ER, Puttemans V, Swinnen SP, van Steenberghe D. Effect of augmented visual feedback from tooth preparation on the performance of dental students. Eur J Dent Educ. 2005;9(1):24–30.
20. Gadbury-Amyot CC, Bray KK, Austin TM, Bray C. Calibration training in dental and dental hygiene education: a review and recommendations. J Dent Educ. 2014;78(9):1275–82.
21. Galibourg A, Boivin A, Vergnes JN. Use of 3D scanning and digital evaluation methods for preclinical assessment of cavity preparations: a systematic review. Eur J Dent Educ. 2021;25(4):661–72.