Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Digital vs. Expert Assessment of Preclinical Simulation Dental Cavity Preparations at the Faculty of Dentistry, University of Benghazi

Naeima Betamar

Naeima.betamar@uob.edu.ly

University of Benghazi

Hana Mahjoub

University of Benghazi

Zaid El Honi

University of Benghazi

Arig Abdullah

University of Benghazi

Research Article

Keywords: Digital assessment, Expert grading, Class I amalgam preparation, Exocad, Dental preclinic education, Intraoral scanning, Agreement analysis

Posted Date: July 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10146301/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Digital vs. Expert Assessment of Preclinical Simulation Dental Cavity Preparations at the Faculty of Dentistry, University of Benghazi

Naeima M. Betamar*[1], Hana E. Mahjoub[2], Zaid B. El Honi[1], Arig M. Abdullah[1]

1) **Naeima M. Betamar:** BDS, MSc, PhD**,** Department of Conservative Dentistry and Endodontics, Faculty of Dentistry, University of Benghazi, Benghazi, Libya
Orcid Number: 0009-0008-5025-4339
Email: Naeima.betamar@uob.edu.ly

2) **Hana E. Mahjoub:** BDS, MSc, PhD, Department of Prosthodontics, Faculty of Dentistry, University of Benghazi, Benghazi, Libya
Orcid Number: 0009-0003-1838-222
Email: Mahgobhana@gmail.com

3) **Zaid B. El Honi:** BDS, MSc, Department of Conservative Dentistry and Endodontics, Faculty of Dentistry, University of Benghazi, Benghazi, Libya
Orcid Number: 0009-0008-2183-625
Email: Zeidalhone@gmail.com

4) **Arig Mohammed Abdullah:** BDS, MSc, Department of Conservative Dentistry, Faculty of Dentistry, University of Benghazi, Benghazi, Libya
Orcid Number: 0009-0000-7471-1380
Email: Arij.abdullah@uob.edu.ly

Authors' affiliation addresses
Faculty of Dentistry, University of Benghazi, Benghazi, Libya

Correspondence:

* **Naeima Mohamed Betamar**, BDS, MSc, PhD**,** Department of Conservative Dentistry and Endodontics, Faculty of Dentistry, University of Benghazi, Benghazi, Libya

Phone: 00218913323718

Email: Naeima.betamar@uob.edu.ly

Abstract

Purpose: This study aimed to compare the agreement and reliability between digital assessment via Exocad and traditional expert consensus grading for Class I occlusal cavity preparations performed by second year dental students.

Materials and Methods: 115 Class I amalgam cavity preparations on typodont teeth were independently assessed by two pairs of calibrated examiners. One pair used Exocad (digital) with fixed millimetre thresholds for depth, isthmus width, marginal ridge thickness and pulpal floor form (0–5 scale, maximum total score 20 points). The other pair used traditional visual and tactile inspection with the same criteria. Agreement was analysed using intraclass correlation (ICC), Bland-Altman analysis and weighted kappa (quadratic).

Results: Expert scores (13.28 ± 2.45) were significantly higher than digital scores (11.75 ± 2.30), with a mean difference (expert – digital) of 1.53 points (95% CI: 1.25–1.81; paired t-test: $t(114) = 10.81$, $p < 0.001$). The ICC was 0.75 (95% CI: 0.66–0.82), indicating good reliability. Bland-Altman analysis showed a bias of 1.53 points and 95% limits of agreement from –1.45 to 4.51. Per-criterion weighted kappa values were highest for depth ($\kappa = 0.68$, substantial agreement) and lowest for marginal ridge thickness ($\kappa = 0.33$, fair agreement), where digital scores were consistently lower due to strict adherence to the ≥ 2.0 mm threshold.

Conclusions: Digital assessment using Exocad demonstrates moderate to good reliability compared to expert grading but produces systematically lower scores, particularly for marginal ridge thickness. Exocad is an accessible alternative for dental schools lacking dedicated grading software. A hybrid model; digital screening followed by expert review is recommended for preclinical education.

Keywords: Digital assessment; Expert grading; Class I amalgam preparation; Exocad; Dental preclinic education; Intraoral scanning; Agreement analysis.

Introduction

Assessment is a fundamental and crucial part of education across various fields because it sets standards and criteria, provides feedback, and helps judge quality [1]. Within the specific context of preclinical dentistry, a well structured evaluation system does more than just grade students. It identifies student strengths and weaknesses, improves curriculum development, and facilitates the transition of skills from preclinical laboratory simulation into clinical practice [2].

For years, traditional preclinical assessment in operative dentistry has relied on expert panels. In this model, teachers look at the prepared tooth, use checklist criteria, and assign a grade based on visual inspection and tactile feel. However, this approach has well known drawbacks because it can be subjective. Different examiners often disagree and even the same examiner may score differently on another day[3–5]. These inconsistencies frustrate students, make them lose interest, and can lead to worse learning outcomes. Students may start to focus more on just passing the exam rather than truly understanding the topic. While some researchers have suggested that improving the technical skills of faculty members might make grading more reliable [6], in practice that is difficult to achieve as it requires significant time, money, and constant effort across a whole department.

As dental student enrollment increases every year, dental schools require assessment methods that are both objective and efficient. This need has led to a growing interest in technology assisted systems[7–11] . These digital tools promise evaluations that are more precise, repeatable, and less biased than traditional panel assessments of restorative preparations[3]. Initially, digital technology was used mainly for the design of restorations (CAD/CAM), but it is currently used to evaluate student prepared cavities as well[12–14]. Modern software can measure depth, width, wall angle, and other features accurately and consistently, giving students the kind of reliable and objective feedback they need to improve their skills[12,15–17]. However, digital assessment is not a perfect solution, as its accuracy depends entirely on the quality of the scanning and careful alignment of the models [18].

Dental education typically progresses through theoretical courses and preclinical laboratory simulations before reaching clinical practice [19]. In the preclinical phase, phantom jaw models and simulated environments using artificial or extracted teeth allow students to develop foundational skills in operative dentistry. These include manual dexterity, good posture, and proper instrument handling[20–23]. The realism of preclinical training depends heavily on the models used. Stone and basic plastic teeth are useful for initial skill acquisition, but they do not fully replicate the tactile and functional challenges of clinical practice. Research has found that stone models require the removal of proximal contacts before cavity preparation, which does not reflect real clinical conditions, while plastic teeth are often too soft and may move during procedures [21]. More advanced phantom models with anatomically accurate teeth allow students to develop authentic manual dexterity[21].

Over the years, simulators have become valuable tools across many areas of dental training. They offer a safe and controlled space where students can repeat procedures as many times as needed, receive immediate feedback, and build confidence before treating real patients [24]. Whether

through virtual reality, haptic devices, or traditional manikins, these systems help bridge the gap between classroom theory and clinical reality. Yet, despite their benefits, simulators come with challenges. Effective assessment tools must balance reliability, validity, educational impact, acceptability, and cost[25]. Simulators can be expensive to install and maintain, and they require technical support as well as careful integration into existing curricula. Furthermore, some students and teachers feel that simulated environments never fully replicate the feel and unpredictability of working on a live patient[25].

Restorative dentistry, particularly precise cavity preparation, is a core skill that directly influences the success of a restoration. Through such practice, students gain competence in preparing carious or defective teeth using various preparation techniques [26,27]. Moreover, integrating patient management scenarios into preclinical education prepares students for direct clinical care[19].

Recent systematic reviews confirm that digital evaluation tools can match or even surpass expert reliability for certain preparation metrics [28,29]. However, most of those studies have focused on crown preparations, while only a few have looked at Class I cavities. One study used a laser scanning system on 38 preparations and found that digital evaluation was more objective than traditional grading[30]. Other researchers introduced a virtual reality simulator to second year students and found that those who trained with it performed significantly better in Class I cavity preparations [31]. While a 3D printed tooth model has been developed to objectively measure manual dexterity in cavity preparation [32], evidence remains limited for widely available CAD software such as Exocad with large sample sizes. Furthermore, no such agreement study has been conducted in Libya or the surrounding region.

The aim of this study was to compare the agreement and reliability between digital grading using Exocad and traditional expert grading for Class I occlusal cavity preparations performed by second year dental students at the Faculty of Dentistry, University of Benghazi. We sought to determine how well the two methods agree and whether digital evaluation could effectively complement or replace traditional grading in operative dentistry education.

Materials and Methods

Study design

This cross-sectional study was conducted at the Department of Conservative Dentistry, Faculty of Dentistry, University of Benghazi, during the 2025 academic year. The methodology was adapted from previously validated digital assessment protocols[12,13] and modified for Class I amalgam cavity preparations. Ethical approval was obtained from the Scientific Research Ethics Committee of the Faculty of Dentistry, University of Benghazi (Approval No. 0260).

The study sample consisted of 115 Class I cavity preparations performed by second year undergraduate dental students during their routine practical training. Each preparation was completed on a standardized mandibular left first molar typodont tooth from the Frasco Company, which was mounted on identical phantom head jaw models as shown in Fig. 1a. To ensure procedural consistency, every student used the same handpiece equipped with a No. 330 Komet bur and a HuFriedy Williams periodontal probe. Prior to the exercise, students received

standardized instructions based on current operative dentistry principles for amalgam cavity preparation as taught in the second year preclinical curriculum and supported by existing digital evaluation literature[13]. This sample size of 115 preparations provided the study with 80% power at an alpha level of 0.05 to detect a moderate agreement with an intraclass correlation coefficient of 0.70 or greater.

The benchmarks for an ideal preparation were defined as a cavity depth between 1.5 and 2.0 mm, an isthmus width representing one third of the intercuspal distance (approximately 1.25 to 1.5 mm), and a marginal ridge thickness of at least 2.0 mm for both mesial and distal aspects. Additionally, the pulpal floor form means that, it was required to be flat and smooth with rounded internal line angles (Table 1). To evaluate these preparations, two independent assessment methods were employed involving digital analysis via an intraoral scanner and traditional expert assessment by calibrated faculty members.

Assessment Methods

I) Digital assessment using an intraoral scanner (Medit i500) and Exocad DentalCAD software.

II) Expert assessment using a checklist-based grading system.

The evaluation process involved four calibrated examiners divided into two independent pairs. One pair managed the digital assessments while the other conducted the expert evaluations, with both groups blinded to each other's results.

Digital Assessment: for the digital workflow, the prepared teeth were cleaned and scanned using a Medit i500 intraoral scanner, which offers an accuracy of 21.0 µm ± 1.48 µm (powder-free optical scanning, 35 FPS) (Fig. 1b). These scans were exported in Standard Tessellation Language format and superimposed onto an unprepared reference model using Exocad's best fit alignment algorithm. To prevent potential distortion, the central occlusal cavity area was excluded from the alignment process, following the approach of Qutieshat and colleagues[13].

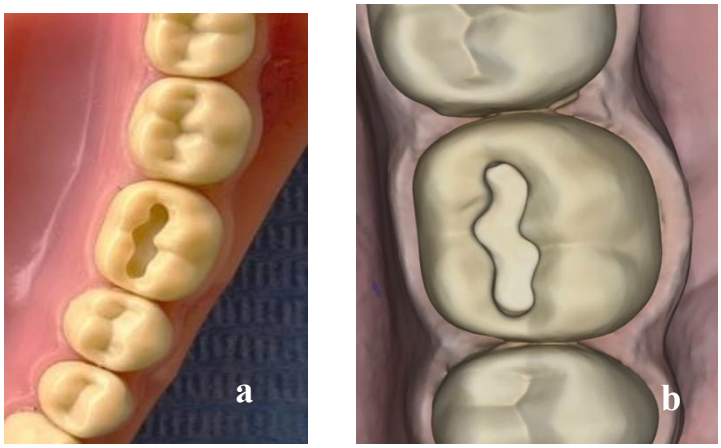


Fig. 1. (a) Class I amalgam cavity preparation on a mandibular left first molar typodont tooth. (b) Digitized image of the prepared tooth imported into Exocad following intraoral scanning (Medit i500).

Quantitative digital measurements were performed by two calibrated examiners using Exocad DentalCAD (3.2 Elefsina). They utilized the software's Measure Distance and Section tools to record values for isthmus width, cavity depth, and marginal ridge thickness. For each site, measurements were taken three times and the mean value was used for subsequent analysis. The isthmus width was measured at the mesial, central, and distal locations, while depth was assessed at the mesial pit, central fossa, and distal pit. Marginal ridge thickness was determined by identifying the narrowest point of the mesial and distal ridges and measuring a line perpendicular to the ridge from the internal margin to the external surface. The pulpal floor form was evaluated qualitatively for flatness and uniformity through cross sections to identify any deviations such as sloped or wavy surfaces, and also roundness of internal line angles as illustrated in Table 1 and Figures 2-4.

Table 1. Scoring criteria for amalgam cavity preparation on typodont teeth (0–5 points).

Score	Classification	Depth (mm)	Width/Isthmus (mm)	Marginal Ridge Thickness (mm)	Pulpal Floor Form
5	Ideal / clinically acceptable	1.50 – 2.00	1.25 – 1.50	≥ 2.00	**Flat, smooth, rounded internal line angles.** Cross-section shows a straight, horizontal line across the entire cavity.
4	Minor deviation (clinically manageable)	2.01–2.50 or 1.20–1.49	1.00–1.24 or 1.51–1.80	1.80 – 1.99	**Slight irregularity.** Cross-section shows a very minor, smooth curve but generally even depth.
3	Moderate deviation (may affect outcome)	2.51–3.00 or 0.90–1.19	0.80–0.99 or 1.81–2.20	1.50 – 1.79	**Moderate unevenness.** Cross-section shows a noticeable incline (sloped) or a distinct dip, but still smooth.
2	Major deviation (restoration likely compromised)	3.01–4.00 or 0.60–0.89	0.60–0.79 or 2.21–2.60	1.20 – 1.49	**Marked irregularity or step.** Cross-section shows a sharp, angular change in depth or a very steep uneven slope.
1	Severe deviation (restoration	> 4.00 or < 0.60	< 0.60 or > 2.60	< 1.20	**Grossly irregular.** Cross-section shows a wavy, jagged, or multi-leveled appearance.

	would probably fail)				
0	Critical error (unacceptable, requires re-preparation)	Extremely deep (>5.0)	Extremely wide (>4.0) or no defined isthmus	Absent marginal ridge	**No defined pulpal floor.** Cross-section shows no flat floor (too shallow) or perforation.

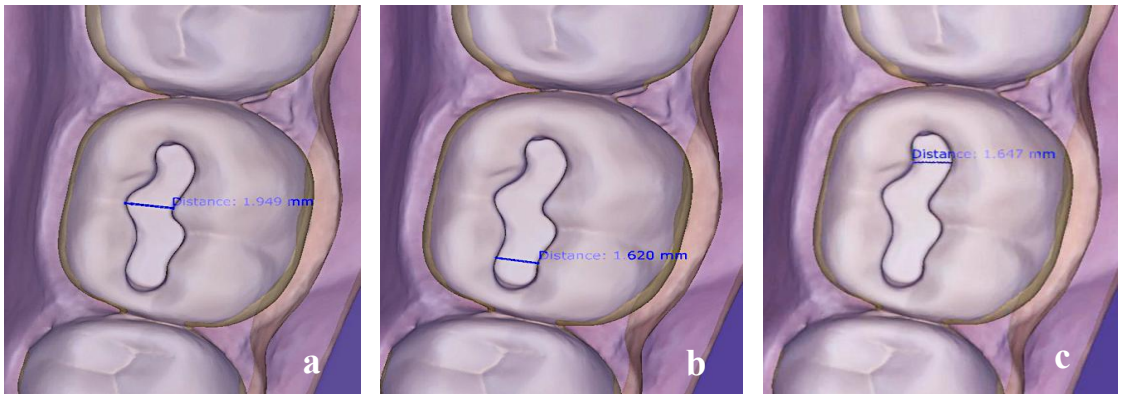


Fig. 2. Isthmus width measurement at three standardized locations: mesial (a), middle (central) (b), and distal (c) isthmus.

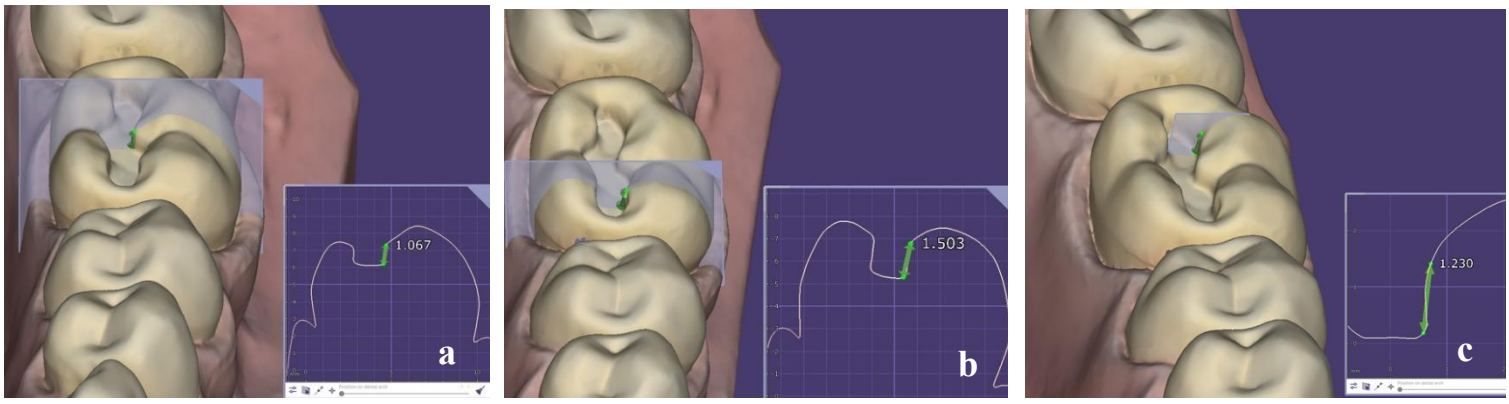


Fig. 3. Cavity depth measurement from the pulpal floor to the occlusal surface plane at the central fossa (a), mesial pit (b), and distal pit (c), using the height tool or measure Vertical distance.

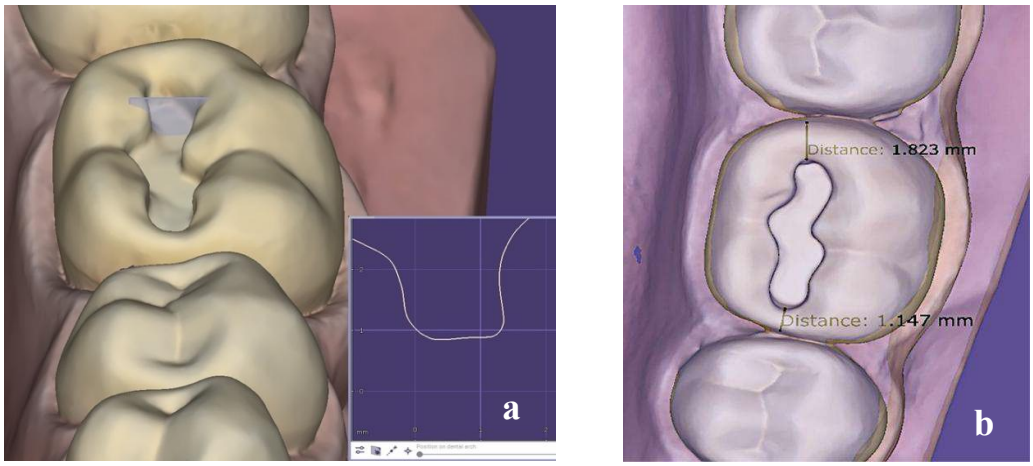


Fig. 4. (a) Pulpal floor form shown on a bucco-lingual cross-section. (b) Marginal ridge thickness measured at the narrowest point from the internal cavity margin to the external tooth surface.

The expert assessment was conducted by two experienced faculty members using a consensus-based checklist grading system. In line with the department's standard preclinical practice, the examiners evaluated each preparation together to reach a single consensus grade for each criterion rather than averaging independent scores. All physical measurements were taken using a periodontal probe with 3 mm markings. To maintain the integrity of the blinding process, each prepared tooth was identified only by a unique number from 1 to 115.

Calibration and Scoring Criteria

Before conducting the study, the two expert examiners participated in inter-examiner calibration sessions. They discussed and clarified the evaluation criteria based on the specific dimensions and characteristics of an ideal Class I amalgam cavity preparation (depth, isthmus width, marginal ridge thickness, and pulpal floor form), following the principles outlined in Sturdevant's Art and Science of Operative Dentistry[26]. The calibration process replicated a typical examination environment, and a unified scoring framework was established. When initial evaluations diverged, the panel members discussed discrepancies and reached a consensus based on their shared clinical and academic expertise.

The scoring framework for both methods was based on a 0 to 5 point scale where a score of 5 represented ideal or clinically acceptable compliance. Partial scores (4, 3, 2, 1, 0) were assigned according to the magnitude of deviation from the ideal, as detailed in Table 1. This system allowed for the differentiation of minor, moderate, and major errors based on their clinical significance, with a maximum possible total score of 20 points per tooth. Digital measurements in millimeters were converted into these 0 to 5 scores using criteria derived from Sturdevant's Art and Science of Operative Dentistry [26]and previously validated digital protocols [12,13].

2.3 Statistical Analysis

Statistical analyses were performed using GraphPad Prism software (version 11.0.1, San Diego, California, USA). Descriptive statistics (mean, standard deviation, median, minimum, maximum, range) were calculated for both assessment methods. The mean differences between expert and digital total scores were compared using a paired t-test. Agreement between the two methods was evaluated using the intraclass correlation coefficient (ICC, two-way random, absolute agreement) and Bland–Altman analysis with 95% limits of agreement. For individual criteria (depth, isthmus width, marginal ridge thickness, and pulpal floor form), weighted Cohen's kappa (quadratic weighting) was calculated using the Lensym online calculator[33] . Statistical significance was set at $p < 0.05$.

Results

Descriptive statistics and comparison of total scores

A total of 115 Class I cavity preparations were assessed by both expert and digital methods. Descriptive statistics for total scores (maximum 20) are summarised in Table 2**.** Digital scores (mean = 11.75, SD = 2.19) were consistently lower than expert scores (mean = 13.28, SD = 2.12).A paired t-test was conducted to compare the mean total scores. Digital scores were significantly lower than expert scores, with a mean difference (expert – digital) of 1.53 points (95% CI: 1.25 to 1.81; $t(114) = 10.81$, $p < 0.0001$). The effect size was large (partial $\eta^2 = 0.5063$), indicating that approximately 51% of the variance in the difference is explained by the assessment method (digital vs. expert). A strong positive correlation was observed between the two methods ($r = 0.75$, $p < 0.0001$), indicating that digital assessment ranks students similarly to expert evaluation despite the systematic difference in absolute scores.

Table 2. Descriptive statistics and paired t-test comparison of expert and digital total scores (n = 115).

Parameter	**Expert Assessment/score**	**Digital Assessment/score**
Total score (max 20)		
Mean (SD)	13.28 (2.12)	11.75 (2.19)
Median	13.5	11.5
Minimum	7.00	6.00
Maximum	17.00	17.00
Range	10.00	11.00
Standard error of mean	0.198	0.204
Paired t-test	**Value**	
Mean difference (expert – digital)	1.53	

95% confidence interval	1.25 to 1.81	
t ratio (df = 114)	10.81	
p value	< 0.0001	
Partial η^2 (effect size)	0.5063	
Pearson correlation (r)	0.7528	

Bland-Altman analysis for agreement between expert and digital total scores

A Bland-Altman plot was constructed to assess the agreement between expert and digital total scores (Fig. 5). The bias (mean difference, expert – digital) was 1.53 points, with a standard deviation of differences of 1.52. The 95% limits of agreement ranged from –1.45 to 4.51 points. This means that for 95% of the cavity preparations, the difference between expert and digital scores lies between –1.45 (digital higher) and 4.51 (expert higher). Although the average bias is small, the relatively wide limits of agreement suggest that individual agreement varies considerably. In some cases, digital assessment may over-score relative to expert evaluation (negative difference), while in others experts give higher scores. Overall, the Bland-Altman analysis confirms that digital and expert assessment methods show acceptable agreement at the group level; however, they are not interchangeable for individual student assessments. Figure 5 Bland-Altman plot showing agreement between expert and digital total scores (maximum 20 points). The solid red line represents the bias (mean difference = 1.53, expert higher). The dashed red lines indicate the 95% limits of agreement (4.51 and –1.45). For 95% of the 115 cavity preparations, the difference between the two methods lies within these limits.

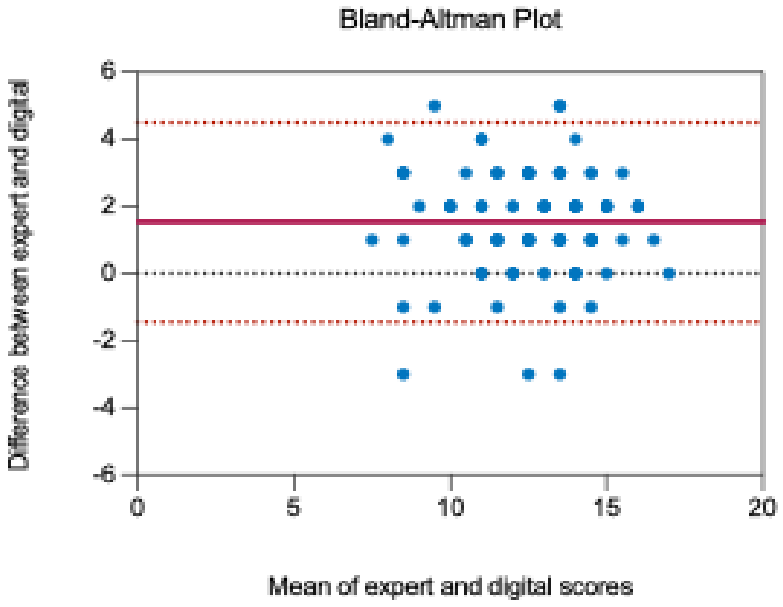


Fig. 5. Bland-Altman plot shows agreement between expert and digital total scores (maximum 20 points). The solid horizontal red line (bias) =1.53, upper dashed line (LoA) =4.51, Lower dashed line (LoA) = -1.45.

Agreement per criterion

Weighted kappa (quadratic) was calculated to assess agreement between expert and digital scores for each of the four cavity preparation criteria. Substantial agreement was observed for depth (κ = 0.68, 95% CI: 0.60–0.76), moderate agreement for isthmus width (κ = 0.52, 95% CI: 0.43–0.61) and pulpal floor form (κ = 0.48, 95% CI: 0.39–0.57), and fair agreement for marginal ridge thickness (κ = 0.33, 95% CI: 0.24–0.42) Table 3. These findings indicate that digital assessment most closely matches expert evaluation on depth measurements, while marginal ridge thickness shows the largest discrepancy, likely due to digital measurement of the absolute narrowest point versus expert visual estimation.

Table 3. Weighted kappa (quadratic) for agreement between expert and digital assessment per criterion (n = 115).

Criterion	Weighted κ	95% CI	Agreement level*
Depth	0.68	0.60–0.76	Substantial
Isthmus width	0.52	0.43–0.61	Moderate
Marginal ridge thickness	0.33	0.24–0.42	Fair
Pulpal floor form	0.48	0.39–0.57	Moderate

*Interpretation based on Landis & Koch [27] (0.41–0.60 = moderate, 0.61–0.80 = substantial, 0.21–0.40 = fair).

Discussion

The present study compared digital assessment via Exocad with traditional expert consensus grading for Class I amalgam cavity preparations performed by second year dental students. Three main findings emerged from this study. First, expert scores were significantly higher than digital scores by an average of 1.53 points out of 20. Second, the two methodologies demonstrated

moderate to good reliability with an intraclass correlation coefficient (ICC) of 0.73. Third, per criterion agreement varied considerably, reaching its highest level for cavity depth ($\kappa = 0.68$) and its lowest for marginal ridge thickness ($\kappa = 0.33$). Notably, this investigation represents the first such comparison conducted at the Faculty of Dentistry, University of Benghazi, providing original baseline data for both the institution and the wider region.

Regarding the systematic difference between the two assessment methods, the consistently lower scores provided by the digital system likely reflect the inherent precision of digital measurement. While Exocad records absolute linear dimensions to the nearest 0.001 mm, expert examiners utilizing a periodontal probe and visual inspection may tolerate minor deviations that do not fundamentally compromise a clinical outcome. This finding aligns with the work of Sly and colleagues [12], who found that E4D Compare software provided lower scores than faculty for Class I preparations, as well as observations by Qutieshat and colleagues [13], that automated algorithms tend to be more stringent than human evaluators. The simplest explanation for this trend is that faculty often apply clinical judgment and allow for minor variations, whereas digital assessment permits no such tolerance. While this lack of leniency is an advantage for standardization, it could present challenges for student passing rates if scoring thresholds are not appropriately adjusted.

In terms of reliability and agreement, the observed intraclass correlation coefficient of 0.73 indicates that digital and expert assessments are acceptably reliable for group comparisons, such as ranking student performance. However, the two methods are not yet interchangeable at the individual level. The relatively wide 95% limits of agreement ranging from -1.45 to 4.51 points, suggest that the two methods might differ by as much as 6 points for any given preparation. On a 20 point scale, such a gap represents a meaningful difference. The negative limit (-1.45) indicates that digital scores can occasionally exceed expert scores, though the majority of differences favour experts' examiners.

This pattern is consistent with earlier reports using prepCheck by Schepke and colleagues [7] and E4D Compare by Renne and colleagues [3], where individual agreement remained moderate despite strong group level correlations. Similar patterns have been observed in other digital platforms, where agreement values closely matched the results found in this study[10] .

When examining per criterion agreement, the highest level of alignment was found for cavity depth ($\kappa = 0.68$). This is logical, as depth is measured in a single vertical plane and experts can palpate it reliably using a periodontal probe. This substantial agreement (0.61-0.80) confirms that depth is the most objectively assessed parameter in both systems. Moderate agreement ($\kappa = 0.52$), was observed for isthmus width, probably due to slight variations in how experts visually estimate the one third intercuspal distance. Pulpal floor form landed in the fair to moderate range ($\kappa = 0.48$), reflecting the inherently qualitative nature of evaluating flatness and internal line angles, which are historically difficult to score with perfect consistency.

The most pronounced disagreement occurred regarding marginal ridge thickness, where agreement was classified as only fair ($\kappa = 0.33$). The obtained digital scores for this criterion were substantially lower than expert scores (mean 1.65 vs. 3.45 out of 5). First, digital measurement quantifies the absolute minimum linear distance from the internal cavity margin to the external tooth surface at the narrowest point of the marginal ridge, with precision to 0.001 mm. Expert examiners on the other hand, using a periodontal probe and visual inspection, tend to assess the marginal ridge more holistically. They may overlook localised under reduction if the ridge appears clinically intact. Second, experts often tolerate minor deviations (e.g., 1.8-1.9 mm) as clinically acceptable and give a score of 4, whereas the digital method strictly adheres to the ideal threshold of ≥ 2.0 mm and classifies anything below 2.0 mm as a deviation (scores 3 or lower). Third, the fixed scoring thresholds shown in Table 1 mean that a digital measurement of 1.99 mm receives a score of 3 (moderate deviation), while an expert might generously round it to 2.0 mm and give a score of 4. This threshold effect hits marginal ridge thickness particularly hard because the ideal cut off (2.0 mm) is clinically critical for fracture resistance. Fourth, visual and tactile assessment of the marginal ridge is inherently challenging because the ridge is narrow and curved; small measurement errors are easy to miss. The difficulty of visually assessing marginal ridge thickness has been noted by Schepke and colleagues [7], who reported that digital evaluation of preparation margins is more reproducible than visual inspection, particularly for narrow anatomical structures. Despite these differences, digital assessment offers a more rigorous and objective evaluation of remaining tooth structure, which may ultimately predict fracture resistance more accurately than visual inspection.

It is important to emphasize that the digital scoring system was based on fixed millimetre thresholds derived from established preparation guidelines [26]. The precise mm ranges for each score (0-5) are shown in Table 1. These fixed ranges ensured objective and reproducible scoring without subjective interpretation, which explains why digital scores were frequently lower than those given by experts. Experts may apply a degree of clinical leniency that the software simply does not allow. The mean difference of 1.53 points, in favour of experts, is consistent with a systematic leniency of approximately 7 to 8 percent of the total score. That difference is clinically meaningful, but not so large as to invalidate the digital method. Effective assessment method must balance reliability, validity, educational impact, acceptability, and cost [25]. Our digital approach scores high on reliability and validity, though it involves a tradeoff in terms of institutional investment.

A notable aspect of this research is the use of Exocad DentalCAD as an accessible tool for quantitative cavity evaluation. Although Exocad is primarily used for CAD/CAM restoration design, we repurposed it to determine if its measurement capabilities could provide objective feedback for preclinical students' cavity preparation. To our knowledge, this is one of the first attempts to use general purpose CAD software rather than dedicated grading platforms like PrepCheck or E4D Compare for this purpose. Our protocol demonstrates that with careful standardization of alignment and reference points, Exocad can achieve reliability comparable to specialized platforms. This finding is of practical importance for dental schools in low resource settings, where specialized software may be cost prohibitive. By utilizing existing CAD software,

institutions can improve objectivity in preclinical education without substantial additional investment, though we acknowledge that the manual workflow is currently more time consuming than automated systems.

These results are broadly consistent with existing literature, aligning with studies that found digital systems tend to be stricter than faculty for both Class I and crown preparations [10,12,13].The close agreement between our reliability values and previous reports suggests that Exocad performs comparably to dedicated grading software. Our study extends these findings by providing per criterion weighted kappa values and a Bland Altman analysis specifically for Class I amalgam cavities, which are not extensively studied in digital assessment literature. Furthermore, our findings align with research using virtual reality simulators and 3D printed models, all of which suggest that digital tools offer a standardized and objective complement to traditional grading [31]. For example, Lee and colleagues observed significant variations in student manual dexterity in cavity preparation [32], through volumetric analysis of 3D printed teeth, a trend consistent with our own observed scoring ranges. While Zou and others confirmed that such automated systems produce more objective results than traditional expert panels[30].

The realism of preclinical training depends heavily on the models used, and our use of standardized typodont teeth allowed for a controlled and objective environment. While stone and basic plastic teeth have limitations, advanced phantom models with anatomically accurate teeth allow students to develop more authentic manual dexterity [21]. These benefits must be weighed against the challenges of cost and curriculum integration [24,25]. This study benefited from several strengths, including a large sample size of 115 preparations, blinded consensus grading, and a validated scoring system. However, some limitations remain, such as the use of a single tooth type and cavity class, which may limit generalization to more complex preparations like Class II cavities. Future work should address intra examiner reliability for digital measurements and explore the tactile differences between typodont and natural teeth.

This study benefited from several strengths, including a large sample size of 115 preparations, blinded consensus grading, and a validated scoring system. The study was conducted under standardised preclinical conditions with identical typodont teeth, burs, and scanning protocols However, some limitations remain, such as the use of a single tooth type and cavity class, which may limit generalization to more complex preparations like Class II cavities. The expert panel consisted of two examiners from one institution; results might differ with larger or international panels. Future work should address intra-examiner reliability for digital measurements and explore the tactile differences between typodont and natural teeth.

Educational implications

The educational implications of these findings support digital assessment as a powerful complementary tool rather than a total replacement for expert grading. Digital methods excel at providing quantitative feedback on specific dimensions, helping students understand their technical errors with precision. Conversely, expert assessment remains essential for evaluating qualitative aspects and making holistic judgments regarding clinical acceptability. A hybrid

approach; involving digital screening for objective measurements followed by expert review of borderline cases, may optimize both objectivity and educational soundness.

Conclusions

In conclusion, within the limitation of this study we can concluded that, within digital assessment using Exocad achieves moderate to good reliability compared to traditional expert consensus for Class I amalgam cavity preparations. Digital scores are systematically lower, particularly for marginal ridge thickness, because the software applies fixed thresholds without the clinical leniency often applied by human examiners. Based on these findings, we recommend a hybrid assessment model that combines the precision of digital tools with the clinical judgment and educational feedback of expert faculty. Future research should investigate whether adjusting digital scoring thresholds improves agreement with experts and whether digital feedback leads to accelerated skill acquisition in longitudinal studies. Additionally, multicenter research would help establish generalizable norms for digital assessment in preclinical operative dentistry.

Declarations

Ethics approval and consent to participate

This study was approved by the Scientific Research Ethics Committee (SREC) at the Faculty of Dentistry, University of Benghazi (Approval No. 0260). "attached file". The study involved a retrospective assessment of dental cavity preparations performed by second year students during their routine practical training; no human participants, human tissue, or identifiable data were directly involved.

All procedures were performed in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments. Informed consent was obtained verbally from all participants involved in this study. The study involved the assessment of routine educational cavity preparations, and all data were anonymized prior to analysis. The Ethics Committee approved the study design. All methods were performed in accordance with the relevant guidelines and regulations of the University of Benghazi.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Clinical trial number

Clinical trial number: not applicable.

Authors' contributions

NMB and HEM contributed to the study design. NMB, HEM, ZBE, and AMA contributed to the methodology. NMB and AMA performed the expert assessment, while HEM and ZBE performed the digital assessment. All authors (NMB, HEM, ZBE, and AMA) contributed to the data analysis and manuscript writing. All authors reviewed and approved the final manuscript.

Acknowledgements

The authors would like to thank the staff and students of the Faculty of Dentistry, University of Benghazi, for their support during data collection, and the preclinical simulation laboratory team for their technical assistance. We are also grateful to Dr Ahmad Shawqi for his guidance and support with the Exocad software.

References

1. Taylor CL, Grey N, Satterthwaite JD. Assessing the Clinical Skills of Dental Students: A Review of the Literature. JEL. 2013;2:p20. https://doi.org/10.5539/jel.v2n1p20
2. El-Kishawi M, Khalaf K, Al-Najjar D, Seraj Z, Al Kawas S. Rethinking Assessment Concepts in Dental Education. Dioguardi M, editor. International Journal of Dentistry. 2020;2020:1–9. https://doi.org/10.1155/2020/8672303
3. Renne WG, McGill ST, Mennito AS, Wolf BJ, Marlow NM, Shaftman S, et al. E4D compare software: an alternative to faculty grading in dental education. Journal of dental education. Wiley Online Library; 2013;77:168–75.
4. Sharaf AA, AbdelAziz AM, El Meligy OAS. Intra- and Inter-Examiner Variability in Evaluating Preclinical Pediatric Dentistry Operative Procedures. Journal of Dental Education. 2007;71:540–4. https://doi.org/10.1002/j.0022-0337.2007.71.4.tb04307.x
5. Satterthwaite JD, Grey NJA. Peer-group assessment of pre-clinical operative skills in restorative dentistry and comparison with experienced assessors. Eur J Dental Education. 2008;12:99–102. https://doi.org/10.1111/j.1600-0579.2008.00509.x
6. Callan RS, Cooper JR, Young NB, Mollica AG, Furness AR, Looney SW. Inter- and Intrarater Reliability Using Different Software Versions of E4D Compare in Dental Education. J Dent Educ. 2015;79:711–8.
7. Schepke U, Van Wulfften Palthe ME, Meisberger EW, Kerdijk W, Cune MS, Blok B. Digital assessment of a retentive full crown preparation—An evaluation of prepCheck in an undergraduate pre-clinical teaching environment. Eur J Dental Education. 2020;24:407–24. https://doi.org/10.1111/eje.12516
8. Tahani B, Rashno A, Haghighi H, Monirifard R, Khomami HN, Kafieh R. Automatic Evaluation of Crown Preparation Using Image Processing Techniques: A Substitute to Faculty Scoring in Dental Education. J Med Signals Sens. 2020;10:239–48. https://doi.org/10.4103/jmss.JMSS_5_20
9. Matthisson L, Zitzmann NU, Zaugg LK, Joda T. Potential of intraoral optical scanning to evaluate motor skills' improvement for tooth preparation: A prospective cohort study. Eur J Dental Education. 2022;26:669–75. https://doi.org/10.1111/eje.12745

10. Kim Y, Kim J, Jeong Y, Yun M, Lee H. Comparison of digital and conventional assessment methods for a single tooth preparation and educational satisfaction. Eur J Dental Education. 2023;27:262–70. https://doi.org/10.1111/eje.12799
11. Soundar J, Hameed F, Valluvan A, Rajambigai.M A. A COMPARATIVE EVALUATION OF THE ACCURACY OF TOOTH PREPARATION FOR VENEERS: CONVENTIONAL METHODS VERSUS DIGITALLY GUIDED TECHNIQUES - AN IN VITRO STUDY. JBSMFS. 2025;401–5. https://doi.org/10.58240/1829006X-2025.21.6-401
12. Sly MM, Barros JA, Streckfus CF, Arriaga DM, Patel SA. Grading Class I Preparations in Preclinical Dental Education: E4D Compare Software vs. the Traditional Standard. Journal of Dental Education. 2017;81:1457–62. https://doi.org/10.21815/JDE.017.107
13. Qutieshat A, Salem A, Kyranides MN. Dental Cavity Grading: Comparing Algorithm Reliability and Agreement with Expert Evaluation. Li Z, editor. International Journal of Dentistry. 2024;2024:3965641. https://doi.org/10.1155/2024/3965641
14. Azhari AA, Ahmed WM, Almaghrabi R, Almalki A, Merdad Y, Alrefai HJ, et al. Comparing the reliability of inter- and intra-grader using digital scanning vs. traditional visual method for evaluating preclinical class II composite preparation. The Saudi Dental Journal. 2024;36:1253–9. https://doi.org/10.1016/j.sdentj.2024.07.005
15. Schubert O, Erdelt K-J, Tittenhofer R, Hajtó J, Bergmann A, Güth J-F. Influence of intraoral scanning on the quality of preparations for all-ceramic single crowns. Clin Oral Invest. 2020;24:4511–8. https://doi.org/10.1007/s00784-020-03316-2
16. Zitzmann NU, Matthisson L, Ohla H, Joda T. Digital Undergraduate Education in Dentistry: A Systematic Review. IJERPH. 2020;17:3269. https://doi.org/10.3390/ijerph17093269
17. Malau-Aduli BS, Lee A, Alele F, Adu M, Drovandi A, Jones K. Preclinical to clinical transition experiences of dental students at an Australian Regional University. Eur J Dental Education. 2022;26:182–96. https://doi.org/10.1111/eje.12687
18. Perry S, Bridges SM, Burrow MF. A Review of the Use of Simulation in Dental Education. Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare. 2015;10:31–7. https://doi.org/10.1097/SIH.0000000000000059
19. Özveren N, Sevinç B, Baltacı E. Evaluation of Dental Students' Knowledge and Opinion About Learning with Virtual Reality. EADS. 2021;14–6. https://doi.org/10.52037/eads.2021.0026
20. Khattak O, Ganji KK, Iqbal A, Alonazi M, Algarni H, Alsharari T. Educational Videos as an Adjunct Learning Tool in Pre-Clinical Operative Dentistry—A Randomized Control Trial. Healthcare. 2022;10:178. https://doi.org/10.3390/healthcare10020178
21. Tezel H, Peskersoy C, Atalayın Özkaya Ç, Kemaloğlu H, Ercan Devrimci E, Karaçolak G, et al. The Effect of the Phantom Models Used in Preclinic Education of 2nd Year Dental Students on the Success of Amalgam Restorations. EÜ Dişhek Fak Derg. 2025;46:189–97. https://doi.org/10.5505/iads.2025.06926
22. Ayar MK. Relationship between students' performance on weekly task assessments and final practical exam at the preclinical restorative dentistry course. Eur Res J. 2019;5:607–12. https://doi.org/10.18621/eurj.400696
23. Nassar HM, Tekian A. Computer simulation and virtual reality in undergraduate operative and restorative dental education: A critical review. Journal of Dental Education. 2020;84:812–29. https://doi.org/10.1002/jdd.12138

24. Gandhi A. Advancements in Dental Education: The Role of Simulators in Training Future Dentists. Journal of Clinical Insights and Research in Dentistry. 2025;1:76–81. https://doi.org/10.5005/jcird-11034-0024
25. Van Der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. Medical Education. 2005;39:309–17. https://doi.org/10.1111/j.1365-2929.2005.02094.x
26. Sturdevant's Art and Science of Operative Dentistry [Internet]. Elsevier; 2019 [cited 2026 June 28]. https://doi.org/10.1016/C2015-0-05603-9
27. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33:159. https://doi.org/10.2307/2529310
28. Ardila CM, González-Arroyave D. Efficacy of CAD/CAM technology in dental procedures performed by students: A systematic scoping review of randomized clinical trials. Heliyon. 2023;9:e15322. https://doi.org/10.1016/j.heliyon.2023.e15322
29. Choi S, Choi J, Peters OA, Peters CI. Design of an interactive system for access cavity assessment: A novel feedback tool for preclinical endodontics. Eur J Dental Education. 2023;27:1031–9. https://doi.org/10.1111/eje.12895
30. Zou H, Jin S, Sun J, Dai Y. A Cavity Preparation Evaluation System in the Skill Assessment of Dental Students. Journal of Dental Education. 2016;80:930–7. https://doi.org/10.1002/j.0022-0337.2016.80.8.tb06173.x
31. Murbay S, Chang JWW, Yeung S, Neelakantan P. Evaluation of the introduction of a dental virtual simulator on the performance of undergraduate dental students in the pre-clinical operative dentistry course. Eur J Dental Education. 2020;24:5–16. https://doi.org/10.1111/eje.12453
32. Lee J, Lee J, Park J, Son S, Shin HS, Kim J. Development of a Quantitative Method to Assess Dental Students' Manual Dexterity in Tooth Cavity Preparation. Eur J Dental Education. 2025;eje.70057. https://doi.org/10.1111/eje.70057
33. B.V L. Cohen's Kappa Calculator: Weighted + Unweighted [Internet]. Lensym. [cited 2026 June 28]. https://lensym.com/tools/cohens-kappa-calculator. Accessed 28 June 2026