AlQalam
Alq J Med App Sci
https://journal.utripoli.edu.ly/index.php/Alqalam/index eISSN 2707-7179

Original article

Evaluation of Preclinical Class I Amalgam Cavity Preparation: A Comparison of Two Assessment Methods

Haneen Azzuz*, Fatma Soltan, Asma Abognah

Department of Conservative Dentistry and Endodontics, Faculty of Dentistry, University of Tripoli, Tripoli, Libya

ARTICLE INFO

Corresponding Email. *h.azzuz@uot.edu.ly*

Received*: 03-09-2024*
Accepted*: 06-11-2024*
Published*: 12-11-2024*

Keywords*. Assessment, Variability, Class I Amalgam, Reliability, Criteria*

ABSTRACT

Preclinical assessment forms an effective strategy in undergraduate training. It enhances the continuous development of skills, facilitates early intervention, and optimizes training resources. Constructive feedback provided to the student fosters continuous improvement in preparedness for professional practice and has implications for the development of a learning culture characterized by student activity. This indeed ensures that a smooth transition will be made toward clinical application of skills. The aim of the study was to evaluate examiner`s reliabilities while applying two different assessment methods: a Glance and Grade assessment method and a Checklist and Criteria assessment method. A total of 100 class I amalgam cavity preparations were assessed by three independent examiners to evaluate Class I amalgam cavity preparations on artificial first mandibular molar teeth prepared by second-year undergraduate dental students in a preclinical laboratory setting. Statistical analysis for inter-examiner and intra-examiner variability was tested using one-way ANOVA test and a paired t-test, respectively. Results of this study showed that intra-examiner variability was significantly reduced with Checklist and Criteria assessment method. The inter-examiner variability was present in a Glance and Grade assessment method while with the Checklist and Criteria assessment method, there was an extremely high level of agreement among the examiners. The study's findings indicated considerable variability among examiners using the Glance and Grade assessment method. However, the Checklist and Criteria assessment method demonstrated remarkable consistency among examiners, indicating that it may offer a more standardized and reliable approach to evaluation.

Cite this article. *Azzuz H, Soltan F, Abognah A. Evaluation of Preclinical Class I Amalgam Cavity Preparation: A Comparison of Two Assessment Methods. Alq J Med App Sci. 2024;7(4):1203-1208. https://doi.org/10.54361/ajmas.247442*

INTRODUCTION

Assessment is an essential component of the educational process across all disciplines. It involves evaluating the quantity, level, significance, value, or quality of the outcomes or products resulting from the learning experience [1]. Assessment involves the establishment of relevant standards and criteria, together with judgments about quality. It is integral to lifelong learning in the same way as any formal educational experience [2]. A dental student evaluation system in a preclinical dentistry is purposed to fulfill several objectives [3]. The assessment tasks seek to pinpoint strengths and weaknesses and to guide curriculum development. It promotes continuous improvement through constructive feedback, preparing students for professional practice and fostering a culture of learning and engagement. The conduction of preclinical assessments is an important tool to facilitate the transition of skills into the clinical phase [4].

Literature on dental education reviews different assessment methodologies; the choice among them, however, depends upon their intended use: summative, formative, or both. Assessment methods may be divided into those which are outcome-oriented, generally called summative assessments, and those which assess student performance during learning, usually referred to as formative assessments [5].

An effective assessment tool will have several important dimensions: reliability, validity, accountability, flexibility, comprehensiveness, feasibility, timeliness, and relevance [6]. Reliability refers to the consistency of scores across time, or, in different terms, that a reliable assessment tool would have no difference between the same tests administered by different persons at different times. Validity refers to what the actual entity it measures is, or what it is supposed to be measuring. Most of these requirements are quite difficult to fulfill in practical life, and practically all the commonly used assessment tools lack on many points. Assessment problems can also be exacerbated by the assessors themselves, with the variability among assessors not always reported, hence further confounding such study results.

Several studies have focused on the determination of examiner agreement in dentistry. The scores of inter-examiner agreement in different studies conducted on clinical and laboratory assessments within dental education have run the range from 0.012 to 0.94 [7-10]. A Glance and Grade assessment method was used in a study by Jenkins et al. (1996) to evaluate Class II cavity preparation. Within a thirteen-point grading system, it was found that the Class II cavities markings varied by up to seven points [11].

Research mainly indicates that intra-examiner variability is lesser than inter-examiner variability [9, 12, 13]. However, most of these studies find that even agreement does not occur. For instance, Satterthwaite and Grey, 2008, found when grading typodont preparations that the intra-examiner agreement for two experienced examiners was 0.53 [14]. This is in agreement with the study by Goepferd and Kerber 1980, who, based on similar results, reported intra-examiner agreement scores ranging between 0.62-0.68 as opposed to 0.3-0.47 between assessors [15].

Due to issues with examiner consistency, students may regard evaluation techniques as being somewhat arbitrary. Therefore, an objective and reliable system of assessment is indispensable in creating a productive learning environment and also for minimizing disputes between students and teachers on grading issues. The study investigates the comparison of the intra- and inter-examiner variability by two assessment methods for evaluation, namely Checklist and Criteria marking system versus Glance and Grade assessment method, in an effort to better understand the factors impacting evaluations in preclinical laboratory courses [13].

METHODS

The study involved assessing 100 class I amalgam cavity preparations on artificial first mandibular molar teeth, prepared by second-year undergraduate students in a preclinical laboratory setting. The checklist criteria used in this study were based on the principles of amalgam cavity preparation in Sturdevant's Art and Science of Operative Dentistry.

Assessment Methods

Glance and Grade Assessment Method

Scoring: This assessment method of cavity preparations relied on visual inspection, whereby every preparation was given a score out of 20. In contrast, this system had the advantage of quickness in giving an estimate, albeit subjective.

Checklist and Criteria Assessment Method:

The criteria were designed by three faculty who came up with a checklist that underscored five very important items for evaluation: a) Outline Form, 4 points; b) Flat Pulpal Floor, 4 points; c) Depth of the Cavity, 4 points; d) Roundation of the angles, 4 points; e) Thickness of the Marginal Ridges, 4 points.

In total: Each tooth could be awarded a maximum score of 20 points to enable more structured and comprehensive assessments of cavity preparations. The examined teeth were labeled from 1 to 100 to ensure consistency in assessment (figure 1). Measurements were made with the aid of a periodontal probe with 3 mm engraved markings.

Figure 1. Labelled examined teeth

Examiners

Three examiners conducted this study, utilizing two different assessment methods. The examiners were faculty members with over 10 years of experience working with preclinical conservative dentistry students. The examiners first evaluated the preparations by the Glance and Grade assessment method followed by the Checklist and Criteria assessment method.

Data Management

Data management involved cleaning and preparing scores from both the Glance and Grade and the Checklist criteria assessment methods. Data were input in SPSS version 25 and cleaned for accuracy; missing values and outliers were checked and cleaned to ensure reliability in the analysis.

Statistical Analysis

There were mainly two types of statistical analysis used: *1. Intra-examiner Reliability***:** The test of the significant difference between scores by the same examiners using the two assessment methods was done by the use of a paired t-test. The p value was considered to be less than 0.05. *2. Inter-examiner Reliability***:** Comparison of the average scores between different examiners was done using one-way ANOVA to check if there was any statistically significant difference in grading by the different examiners. The p-value for the differences was considered less than 0.05.

RESULTS

Inter-Examiner Variability

Evaluation of Glance and Grade Assessment Method

Statistical analysis of the Glance and Grade assessment method revealed significant difference in the mean scores assigned by the examiners. While Examiner 1 scored a mean of 10.32 with a SD of 3.278, Examiners 2 and 3 gave higher mean scores of 12.20 and 12.67, respectively, with SDs of 3.803 and 3.172, respectively. The results of the ANOVA indicated that the F-value was statistically significant at 13.152, with the p-value equal to 0.001, indicating that the difference in scoring among the examiners was statistically significant (see Table 1).

Table 1. Descriptive Statistics and ANOVA Results for the Glance and Grade Marking System

Examiner	(Mean ± SD)	F	P-value
Examiner 1	**10.32 ± 3.278**	**13.152**	**0.001**
Examiner 2	**12.20 ± 3.803**		
Examiner 3	**12.67 ± 3.172**		

The post-hoc tests provided a clearer overview of the discrepancies in the scores across examiners. This revealed that the scores for Examiner 1 were significantly lower than Examiner 2, at a mean difference value of 1.880 and a p-value of 0.001. The scores given by Examiner 1 were also significantly lower than those of Examiner 3, with a mean difference

of 2.350 and a p-value of 0.001. On the other hand, the two examiners in Examiner 2 and Examiner 3 did not present any significant statistical difference, as manifested in a mean difference of 0.470 and a p-value of 0.626. (Table 2).

Table 2. Post Hoc Test Results for the Glance and Grade Marking System

Comparison	Mean Difference	P-value
Examiner 1 vs. Examiner 2	**1.880***	**0.001**
Examiner 1 vs. Examiner 3	**2.350***	**0.001**
Examiner 2 vs. Examiner 3	**0.470**	**0.626**

**Significant at the 0.05 level.*

Evaluation of the Checklist and Criteria Assessment Method

Descriptive statistics for the Checklist Criteria indicated that the mean scores assigned by the examiners were very close in value. The mean for Examiner 1 was 13.22, with a SD of 2.915; for Examiner 2, the mean was 13.16, with a SD of 3.271; and that for Examiner 3 was 13.32, with a SD of 2.971. These are supported by the results of the ANOVA, which yielded an F value of 0.07 and a p-value of 0.932, indicating that there is no significant difference between examiners' scores for the Checklist Criteria assessment method. Such a result suggests an extremely high agreement level of the examiners in their evaluation of cavity preparation using the Checklist Criteria assessment method. (Table 3.)

Table 3. Descriptive Statistics and ANOVA Results for the Checklist and Criteria

Examiner	(Mean ± SD)	F	P-value
Examiner 1	**13.22 ± 2.915**	**0.07**	**0.932**
Examiner 2	**13.16 ± 3.271**		
Examiner 3	**13.32 ± 2.971**		

Intra-Examiner Variability

Table 4 shows the t-test, indicating that, in comparison, the mean scores between the Glance and Grade assessment method and the Checklist and Criteria assessment method for each of the examiners were significantly different The mean score of Examiner 1 was 10.32 (SD = 3.277) in the Glance and Grade assessment method evaluation, which is remarkably lower compared with the mean score derived from the Checklist and Criteria assessment method evaluation, which was 13.22 with a SD of 2.914. Thus, a t-test analysis produced a p-value of 0.001, showing a statistically significant difference between these two methods. Similarly, Examiner 2 reported an average of 12.20 (SD = 3.803) for Glance and Grade assessment method evaluation, while the average for Checklist and Criteria assessment method evaluation was 13.16 (SD = 3.271). By applying a t-test, one gets a p-value of 0.001, demonstrating that the difference between the methods is statistically significant. In the Glance and Grade assessment method, the average score, as provided by Examiner 3, was 12.67 (SD = 3.172), while in the Checklist and criteria assessment method evaluation, the mean score was 13.32 (SD = 2.970). Using the t test analysis, the p-value was 0.008, which implies that these evaluations are significantly different.

Table 4. Comparison of Mean Scores between Glance and Checklist Evaluations by Examiners

Examiner	Glance (Mean ± SD)	Checklist (Mean ± SD)	t	P-value
Examiner 1	**10.32 ± 3.277**	**13.22 ± 2.914**	**-13.71**	**0.001**
Examiner 2	**12.20 ± 3.803**	**13.16 ± 3.271**	**-3.76**	**0.001**
Examiner 3	**12.67 ± 3.172**	**13.32 ± 2.970**	**-2.71**	**0.008**

DISCUSSION

Assessment for learning is an educational approach designed to engage both educators and students in enhancing the learning process and cultivating a favorable perspective on future learning [16]. During preclinical practice, the students undergo exercises that help them learn technical manual skills that are important in achieving high competence in conservative dentistry. In order to encourage effective learning and to minimize grade discrepancies between learners and instructors, an objective and reliable assessment method is integrally necessary [17].

This study found a significant difference between the Glance and Grade assessment method and the Checklist Criteria assessment method. This supports the findings of articles written by Sherwood, I. A., and Douglas, G. V. who advised that the evaluation of pre-clinical operative work should be done using an objective checklist criterion [18]. Furthermore, Scott et al. revealed that the use of a criteria checklist in the scoring system enhances the objectivity of the evaluation

[19].
Numerous studies have shown that using an objective checklist scoring system significantly reduces intra-examiner variation and greatly improves intra-examiner reliability [20, 21]. Haj-Ali et al. recommended the establishment of an evaluation system in order to reduce the inconsistencies among dental educators. This system should provide objective criteria, a suitable scoring scale, and a system of training for the assessors to enhance their ratings. Moreover, they point out that these calibration and training sessions should be based on a standard method as it has been demonstrated to help improve the agreement between the examiners and also assist in maintaining the agreement for a period of 10 weeks [22].
The outcomes of this study showed that the Glance and Grade assessment method was found to be less reliable with poor inter-examiner and intra-examiner reliability when compared to the Checklist Criteria assessment method. This is in agreement with the findings of Houpt and Kress, who emphasized the need for written criteria in helping examiners maintain consistent standards of judgment [23].
This study revealed that the application of Checklist Criteria assessment method improved the intra-examiner reliability and inter-examiner variability among all examiners.

CONCLUSION

The results of this study revealed significant inter-examiner variability in the Glance and Grade Marking method; the marks given by Examiner 1 are considerably lower than the marks of Examiners 2 and 3. Indeed, this points to the serious question of the reliability of grading by different examiners. The checklist and criteria marking method through direct comparison are very consistent among the examiners, implying that this method provides a more standardized and consistent approach to evaluation. Additionally, the intra-examiner reliability test revealed that all examiners using Glance and Grade and Checklist ratings were significantly different from each other, with the Checklist Criteria rating always being higher. The Checklist Criteria assessment method therefore seems to be an effective method for the evaluation of cavity preparations as it tends to give more consistent and higher scores among examiners. In summary, these findings emphasize the necessity of standardized assessment procedures.

Recommendation

For greater consistency and reliability in assessments, it is essential for examiners to become familiar with the criteria through practice. Inter-examiner variation could be reduced if examiners gain experience with the assessment method and complete calibration prior to evaluating students' work.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

REFERENCES

1. Taylor C, Grey N, Satterthwaite JD. Assessing the Clinical Skills of Dental Students: A Review of the Literature. J Educ Learn. 2013 Jan;2(1):20. doi: 10.5539/jel.v2n1p20.
2. Atia A, El-Attug M. Profiling of Accreditation in Libyan Higher Education Institutions. Open Journal for. 2021 Jun 6:17.
3. Shugars DA, May KN, Vann WF. Comprehensive evaluation in a preclinical restorative dentistry technique course. J Dent Educ. 1981 Dec;45(12):801-3.
4. El-Kishawi M, Khalaf K, Al-Najjar D, Seraj Z, Al Kawas S. Rethinking Assessment Concepts in Dental Education. Int J Dent. 2020 Oct 14;2020:8672303.
5. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. Med Teach. 2011;33(6):478-485.
6. Turnbull J, Gray J, MacFadyen J. Improving in-training evaluation programs. J Gen Intern Med. 1998 May;13(5):317-23.
7. Dhuru VB, Rypel TS, Johnston WM. Criterion-oriented grading system for preclinical operative dentistry laboratory course. J Dent Educ. 1978 Sep;42(9):541-546.
8. Fuller JL. The effects of training and criterion models on interjudge reliability. J Dent Educ. 1972 Apr;36(4):19-22.
9. Lilley JD, ten Bruggen Cate HJ, Holloway PJ, Holt JK, Start KB. Reliability of practical tests in operative dentistry. Br Dent J. 1968 Sep 3;125(5):194-7.
10. Quinn F, Keogh P, McDonald A, Hussey D. A study comparing the effectiveness of conventional training and virtual reality simulation in the skills acquisition of junior dental students. Eur J Dent Educ. 2003 Nov;7(4):164-9.
11. Jenkins SM, Dummer PM, Gilmour AS, Edmunds DH, Hicks R, Ash P. Evaluating undergraduate preclinical operative skill; use of a glance and grade marking system. J Dent. 1998;26(8):679-684.
12. Deranleau NJ, Feiker JH, Beck M. Effect of percentage cut-off scores and scale point variation on preclinical project evaluation. J Dent Educ. 1983 Oct;47(10):650-5.

13. Sharaf AA, AbdelAziz AM, El Meligy OA. Intra- and inter-examiner variability in evaluating preclinical pediatric dentistry operative procedures. J Dent Educ. 2007 Apr;71(4):540-4.
14. Satterthwaite JD, Grey NJ. Peer-group assessment of pre-clinical operative skills in restorative dentistry and comparison with experienced assessors. Eur J Dent Educ. 2008 May;12(2):99-102.
15. Goepferd SJ, Kerber PE. A comparison of two methods for evaluating primary class II cavity preparations. J Dent Educ. 1980 Sep;44(9):537-42.
16. Elmaroush M, Ben Hamida S, Elgarboae Z, Mohsen N. Variability in assessment of preclinical prosthodontics work of undergraduate students. Khalij-Libya J Dent Med Res. 2022;6(2):129–33.
17. Huth KC, Baumann M, Kollmuss M, Hickel R, Fischer MR, Paschos E. Assessment of practical tasks in the Phantom course of Conservative Dentistry by pre-defined criteria: a comparison between self-assessment by students and assessment by instructors. Eur J Dent Educ. 2017 Feb;21(1):37-45.
18. Sherwood IA, Douglas GV. A study of examiner variability in assessment of preclinical class II amalgam preparation. J Educ Ethics Dent. 2014;4:12–7.
19. Scott BJ, Evans DJ, Drummond JR, Mossey PA, Stirrups DR. An investigation into the use of a structured clinical operative test for the assessment of a clinical skill. Eur J Dent Educ. 2001 Feb;5(1):31-7.
20. Gansky SA, Pritchard H, Kahl E, Mendoza D, Bird W, Miller AJ, Graham D. Reliability and validity of a manual dexterity test to predict preclinical grades. J Dent Educ. 2004 Sep;68(9):985-94.
21. Park RD, Susarla SM, Howell TH, Karimbux NY. Differences in clinical grading associated with instructor status. Eur J Dent Educ. 2009 Feb;13(1):31-8. doi: 10.1111/j.1600-0579.2008.00534.x.
22. Haj-Ali R, Feil P. Rater reliability: short- and long-term effects of calibration training. J Dent Educ. 2006 Apr;70(4):428-33.
23. Houpt MI, Kress G. Accuracy of measurement of clinical performance in dentistry. J Dent Educ. 1973 Jul;37(7):34-46.

تقييم تحضير تجويف الأملغم من الفئة الأولى في المرحلة ما قبل السريرية: مقارنة بين طريقتين للتقييم

حنين عزوز * , فاطمة سلطان , اسماء ابوجناح

قسم العلاج التحفظي و علاج الجذور , كلية طب و جراحة الفم و الأسنان , جامعة طرابلس

المستخلص

يشكل التقييم في المرحلة ما قبل السريرية استراتيجية فعالة في تدريب طلاب المرحلة الجامعية. فهو يعزز التطور المستمر للمهارات ويسهل التدخل المبكر ويعزز من كفاءة استخدام موارد التدريب. كما أن النقد البناء الذي يتلقاه الطالب يساهم في تحسين استعداده المستمر لممارسة المهنة ويؤثر على تطوير ثقافة التعلم للطلاب. وهذا يضمن انتقالاً سلسًا نحو التطبيق السريري للمهارات. هدفت هذه الدراسة إلى تقييم دقة الممتحنين عند تطبيق طريقتين مختلفتين للتقييم: طريقة "النظرة والتقييم" وطريقة "قائمة المراجعة والمعايير". تم تقييم 100 تجويف من الفئة الأولى للأملغم من قبل ثلاثة ممتحنين مستقلين لتقييم تحضير تجويف للأملغم على أسنان صناعية للضرس السفلي الأول, قام بتحضيرها طلاب طب الأسنان في السنة الثانية في المعمل. و قد تم إجراء التحليل الإحصائي لاختبار التباين بين الممتحنين و تقييم كل ممتحن لنفسه باستخدام اختبار (ANOVA)أحادي الاتجاه واختبار (t) المزدوج على التوالي. أظهرت نتائج هذه الدراسة أن التباين داخل الممتحن نفسه كان منخفضًا بشكل ملحوظ عند استخدام طريقة "قائمة المراجعة والمعايير". و لقد وُجد تباين بين الممتحنين في طريقة "النظرة والتقييم"، بينما كان هناك مستوى عالٍ جدًا من الاتفاق بين الممتحنين عند استخدام طريقة "قائمة المراجعة والمعايير". أشارت نتائج الدراسة إلى وجود تباين كبير بين الممتحنين عند استخدام طريقة "النظرة والتقييم". ومع ذلك أظهرت طريقة "قائمة المراجعة والمعايير" اتفاقًا ملحوظًا بين الممتحنين، مما يشير إلى أنها قد توفر نهجًا أكثر موثوقية وموحدًا للتقييم.

الكلمات المفتاحية :التقييم، التباين، الأملغم من الفئة الأولى، الدقة، المعايير.