European Journal of Dental Education ISSN 1396-5883

Utilising an innovative digital software to grade pre-clinical crown preparation exercise

E. T. Kateeb[1], M. S. Kamal[2], A. M. Kadamani[2], R. O. Abu Hantash[2] and M. M. Abu Arqoub[2]

[1] Department of Periodontology and Preventive Dentistry and Oral Health Research and Promotion Unit, Al-Quds University, Jerusalem, Palestine,
[2] Department of Fixed and Removable Prosthodontics, Al-Quds University, Jerusalem, Palestine

Keywords
dental education; assessment; educational technology; prosthodontics; crown preparation.

Correspondence
Elham Kateeb
Faculty of Dental Medicine
Al-Quds University Main Campus, Abu Dis,
P.O Box 89,
Jerusalem, Palestine
Tel: 00972599510404
Fax: 0097022791065
e-mail: Elhame20@gmail.com

Accepted: 18 March 2016

doi: 10.1111/eje.12204

Correction added on 5 July 2016, after first online publication: Co-author's spelling corrected.

Abstract

Background: Accurate assessment of dental students' pre-clinical work is the most critical component of the dental education process. Thus, this study came to investigate the effectiveness of using technology in students' pre-clinical work evaluation; by comparing grades generated from a digital assessment software of a prepared tooth and a traditional visual inspection carried out by four calibrated faculty members.

Methods: Ninety-six teeth were prepared for a ceramo-metal crown by fourth year dental students. The four examiners and the digital grading software evaluated independently each preparation once. A random sample of 20 preparations were graded twice to assess intra-rater reliability. Inter-class correlation (ICC) was used to measure agreement among the four examiners, and between the examiners and the digital grading software. Paired student *t*-test was used to assess the accuracy of grades generated from visual inspection when compared to the digital grading system.

Results: Intra-rater reliability for examiners 1 and 2 were 0.73 and 0.78 and for the digital grading system was 0.99. The inter-rater reliability among the four examiners was very good, ICC of 0.76. However, the agreement between scores produced by the examiners and the digital system were mostly in the low to moderate range. The paired *t*-test demonstrated statistically significant differences between each examiner and the digital grading by 6—25 grades.

Conclusions: This study demonstrates that the digital grading system used in this study can reliably scan and compare students' tooth preparations to a known gold standard. Results of this study suggests that using digital grading will preclude the variability and the subjectivity that usually result from the traditional visual inspection grading.

Introduction

Accurate assessment of dental students' pre-clinical work is the most critical component of the dental education process (1). Assessment has an important role in achieving planned educational objectives and should not be a goal in itself (2). And for assessment to be effective, it should provide consistent and accurate feedback to students to help them achieve high level of competency before progressing to patient care (1).

The traditional method that is commonly used to assess students' pre-clinical dental work such as tooth preparations usually depends on the judgment of experienced restorative dentistry specialists mainly utilising visual inspection. Unfortunately, this method fails to provide students with consistent feedback due to many sources that contribute to disagreement including grading scale, grader calibration, training and subjective influences (3). Different studies found significant disagreement between graders when evaluating students' pre-clinic assignments (4–6). A more recent work that evaluated consistency in pre-clinical grading found significant disagreement between examiners in almost all types of tooth preparations (7).

Even in studies that graders were calibrated to standardise the grading process, results showed a significant element of uncertainty and inconsistency amongst evaluators (8). Continuing training and calibration sessions may minimise these differences; however, this needs time and resources that are usually not available for busy faculty members (9).

On the other side, this inconsistency in grading may interfere with the learning process when students perceive it as a form of favouritism, discrimination and unfairness (1). This usually leads to students focusing more on the grade than the actual learning experience.

Recently, dental education leaders demonstrated great interest in improving grading systems. As an example, the U.S. Commission on Dental Accreditation (CODA) has suggested predetermined, standardised, reliable and valid grading forms to be used in the US dental schools (10). Those standards require dental schools to develop robust assessment tools along with protocols to give effective feedback to students. Investigators concluded that to achieve this goal, dental schools need to remove the human element and adopt more objective methods (6, 11).

This can be achieved by adding an additional dimension to the learning process by using technology in feedback such as a grading software that can provide assessment and communicate feedback to students and faculty members (12). This technology can help evaluators achieve a more objective assessment and help students improve their self-assessment (13). Consistent and accurate feedback from a digital grading software can help students differentiate between weaknesses in hand skills and deficiencies in conceptual knowledge (13).

Thus, this study investigated the effectiveness of using technology in students' pre-clinical assignment evaluation by comparing grades generated from a digital grading software (Nissan Fair Grader 100®, Nissin Dental Products Inc., Kyoto, Japan) and four calibrated faculty members. This software scans a student's prepared tooth and compares it to a known (software driven) standardised preparation. An actual numerical evaluation is generated by this system.

Studies that assessed the use of digital grading systems in students' evaluation agreed on the great potential of digital technology to produce more objective assessment. Those studies mainly used (E4D Compare) (E4D Technologies LLC, Richardson, TX, USA) digital software and evaluated different students' pre-clinical skills, such as crown preparations, dental anatomy and intracoronal restorations (1, 14, 15).

The hypothesis of this study was that the (Nissin Fair Grader 100®) software is more consistent and therefore less variable when evaluating students' tooth preparations compared with the traditional visual inspection.

Methods

Ninety-six teeth were prepared by fourth-year dental students at Al-Quds University, Faculty of Dental Medicine as part of their fixed prosthodontics course final exam. The preparations were performed as a 'practical examination' after adequate didactic, laboratory and clinical instructions on the proper parameters necessary to accomplish an ideal preparation. The assignment was to prepare tooth #1.6 (maxillary right first molar) for a ceramo-metal crown using a Nissin Simple Root Tooth Model (permanent tooth) A5A-200® (Nissin Dental Products INC., Kyoto, Japan) fixed in the jaw of a simulation unit.

The students were allowed 1 h for the preparations. Preparations were then graded (double blinded) by four faculty members specialising in fixed prosthodontics (examiners 1, 2, 3 and 4). Examiners 2, 3 and 4 had been involved in teaching fixed prosthodontics for almost 10 years, whilst faculty member 1 had only been teaching this course for the past 4 years.

The preparations were graded by the examiners on a raw scale from 0 to 100. Examiners were calibrated to grade against the ideal 'gold standard' preparation. Calibration was performed on what constitutes an ideal preparation and how to score deviations from ideal. Additionally, faculty members assessed a sample of 10 preparations to ensure that all of them agree independently on the different evaluation criteria that they set for this practice such as taper, amount of reduction and type of finish line. Figure 1 demonstrates criteria used in the visual grading.

For the first time at Al-Quds University, Faculty of Dental Medicine, a digital grading system was introduced to ensure accuracy and fairness in students' grading. Nissin Fair Grader 100® was used in this study as the digital grading system, and its consistency was assessed by grading a random sample of 20 tooth preparations twice.

Nissin Fair Grader 100® scans the preparations in three dimensions and superimposes the prepared tooth on the master gold standard preparation to detect the differences. The scanned images consist of 300 000 points of cloud data that produce high accuracy replica of the tooth preparation. The Fair Grader 100® produces a PDF report that is saved to the hard disc and shows deviations from a gold standard preparation in horizontal and vertical cross sections. Figure 2 shows Fair Grader 100® evaluation criteria.

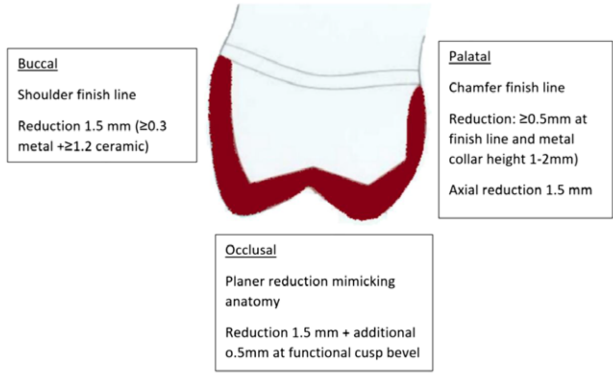
In the first step, the students' preparations were removed from the simulation unit, labelled by a given ID number, blinded and graded independently by the four faculty members (examiners 1, 2, 3 and 4). In the second step, the tooth preparations were scanned, and stored in the digital grading system (Nissin Fair Grader 100®) by a trained teaching assistant. The software performed the grading task by comparing the student's preparation to a gold standard preparation. In this study, the gold standard was created by Nissin Fair Grader 100® ideal preparation criteria that was approved by the Prosthodontic Department faculty members. The software gives the final grade of the student's preparation on a (0—100) scale by measuring any discrepancy in reduction (over-reduction or under-reduction) than the gold standard set by the software. In this application, Nissin Fair Grader 100® software tolerance area was set to 0.25 as the acceptable range that students' preparations can vary from the ideal. Figure 3 shows the final evaluation report produced by Nissin Fair Grader 100®.

These two methods of evaluating a student's ability to prepare a tooth for a ceramo-metal single crown were compared. The first method involved a faculty member examiner comparing the student's tooth preparation to the gold standard preparation based on visual inspection of the agreed-upon criteria and providing a grade that ranges from 0 to 100.

The second method utilised the software assessment of the tooth preparation using the method explained above. More

CRITERIA USED IN VISUAL GRADING

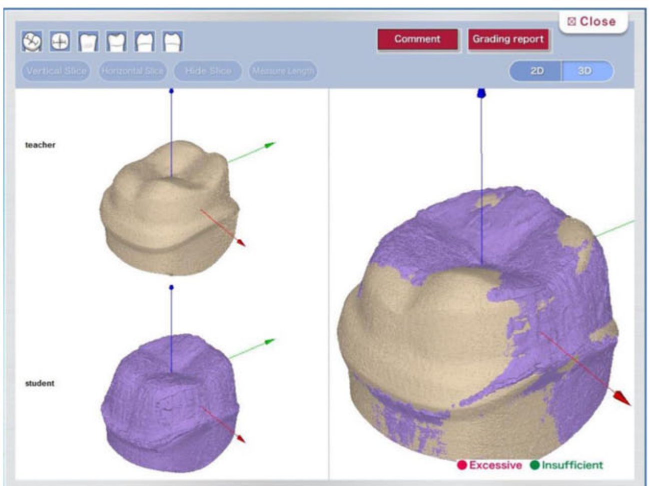
Criteria	Operational Definition	Weight
Finish line	Deep chamfer bucally, light chamfer palatally, wingless preparation	20%
Amount of axial reduction	1.2 mm bucally, 0.5 – 0.8 mm palatally	20%
Parallelism	Free of under cuts, parallel axial walls, taper 4- 14 degrees	20%
Planner occlusal reduction	Nonfunctional cusp areas 1.5 mm reduction, functional cusp bevel areas 2.0 mm	20%
Finishing	Over all shape , line angels, smoothening, foundation	20%



Tooth number 16 prepared for a ceramo-metal crown: (Wingless preparation)

Fig. 1. Criteria used in visual grading. Tooth number 16 prepared for a ceramo-metal crown: (Wingless preparation).

The Grading software comparing the participant's preparation scan vs. The ideal Master preparation scan	The Fair Grader 100 ®software grading settings



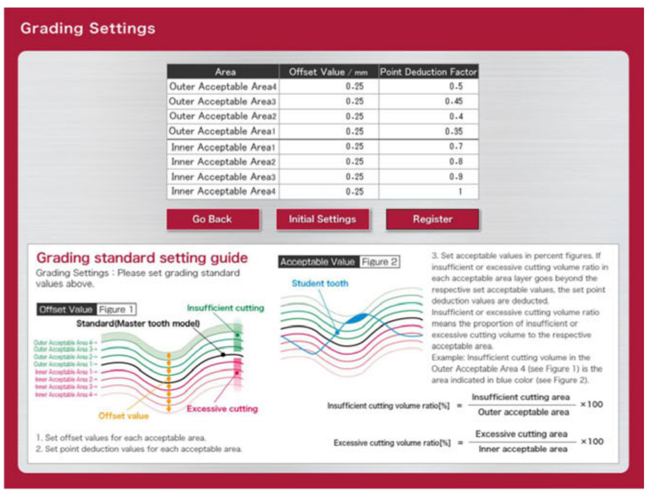


Fig. 2. Fair Grader 100™ Digital Software Grading Criteria.

details on how the software works can be found at the Fair Grader 100®official website (16).

It was hypothesised that the software would be more precise in grading the students and that traditional visual inspection grading method would have a greater variability. It is also hypothesised that if we consider the digital grading as the gold standard, examiners would show different degrees of accuracy in their grading using the visual inspection. The Research Ethical Board at Al-Quds University approved all aspects of this research.

Statistical analysis

The four examiners and the digital grading system (Nissin Fair Grader 100®) evaluated independently each student's preparation once. The evaluation provided by the digital system and the four examiners was on a continuous measure, ranging from 0 to 100. Grades from the four examiners using the visual inspection were averaged (mean rating); this average is usually used as the student's final grade.

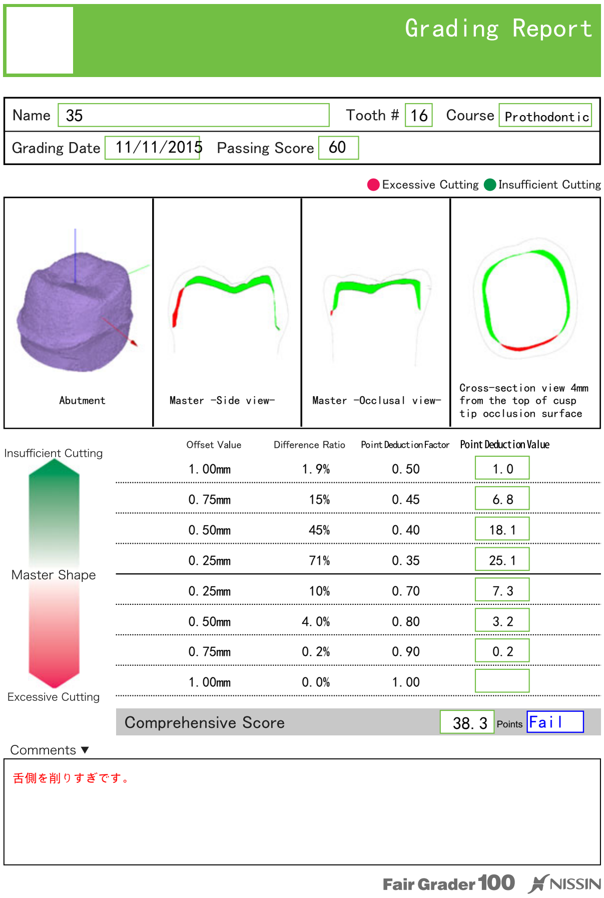


Fig. 3. Grading Report Produced By the Fair Grader 100™ Digital Software (Nissin Dental Products Inc.)

Descriptive statistics such as the mean, standard deviations and the range for the digital grading system, the individual rating of examiners 1, 2, 3 and 4 and the mean rating were generated.

Twenty tooth preparations were selected randomly and were re-evaluated by examiners 1 and 2 after a month from the first evaluation to assess the intrarater reliability of each examiner. This was carried out using SPSS interclass correlation (ICC) single measure one-way random definition.

Inter-rater reliability amongst the four examiners was assessed using the interclass correlation (ICC) (17). Inter-class correlation average measurements were generated using SPSS absolute agreement and consistency definitions assuming a two-way mixed effects model where examiners' effects are random and measures' effects are fixed.

Reliability tests measure precision and repeatability by measuring the degree to which two methods produce the same results when they measure the same sample (18); however, accuracy of a grading method can be measured by assessing the agreement of this method against a gold standard (18). If we consider the digital grading with its high precision as the gold standard, we can assess the accuracy of each of the four examiners using different statistical methods.

One way is to check whether the mean grade for each examiner is equal to the mean grade for the digital system. This was considered by calculating the differences between the examiner's score and the digital grading score and testing whether they differ from zero using *paired t*-test (17). Looking at the standard deviation (SD) for each difference gave us an idea about the agreement between the two methods. However, this reflects one aspect of the agreement because the means can be equal, whilst the (random) differences between measurements can be huge. For that, we used Bland–Altman plots and limits of agreement to assess any systematic difference between the digital grading and the individual examiner rating (i.e. fixed bias) and to identify possible outliers (19). Bland–Altman was constructed by plotting |Examiner x − digital rating| against (Examiner x + digital rating)/2, and limits of agreement were generated by calculating the mean difference ± 1.96 SD of the difference (17). The limits of agreement represent the range of values in which agreement between the two methods will lie for approximately 95% of the sample (18).

Spearman's rank correlation was used to check bivariate linear association between each pair of examiners. We also used Spearman's rank correlation to assess the association between the mean rating and the digital grading system. For all calculations, the software SPSS version 20 was used (20).

Results

The final sample included the grades for 91 students' preparations. Five preparations were excluded because they were performed on the wrong tooth number.

Four experienced and calibrated examiners were responsible for visually grading the preparations. The grading was performed once for the 91 preparations by the four examiners and twice for a random sample of 20 teeth by two examiners. Table 1 shows descriptive statistics of students' grades by the four examiners and the digital grading system.

The average time for visually grading a total of 91 samples was 3 h and 40 min (2.5 min/sample/examiner). In contrast, it took a little over 3 h for one teaching assistant to scan the 91 preparations and get the results using Nissin Fair Grader 100® [100 s (1 min and 40-s) to scan one tooth preparation].

The intrarater reliability for visual scoring amongst examiners 1 and 2 was very good. The single measure of the ICC for both examiners was 0.73, 95% CI [0.44—0.88] for Examiner 1 and 0.77 for Examiner 2, 95% [0.51—0.90]. The strong Spearman correlation coefficients, especially for examiner 2 (Examiner 1: 0.69, $P < 0.001$; Examiner 2: 0.93, $P < 0.0001$)

TABLE 1. Descriptive statistics of students' grades by the four examiners and the digital grading system

	n	Minimum	Maximum	Median	Mean	SD
Examiner 1	91	0	90	60	57.4	21.6
Examiner 2	91	10	85	53	48.9	17.1
Examiner 3	91	30	89	70	68.1	11.5
Examiner 4	91	30	90	60	57.7	18.1
Digital grading	91	0	71	47	42.2	20.1

SD, standard deviation.

indicate that, between the two grading trials, samples were ranked similarly to each other. This implies that tooth preparations that scored well on the first grading trial also tended to score well on the second grading trial.

Regarding the inter-rater reliability amongst the four examiners, authors first were interested in assessing the degree that examiners' ratings were consistent with one another such that higher ratings by one examiner corresponded with higher ratings from another examiner. This was performed by the ICC mixed model with SPSS consistency definition. The average measure of ICC amongst the four examiners was excellent, 0.82, 95% [0.75—0.87]. On the other hand, the average measure of the 'Absolute Agreement' type of ICC (the degree that examiners agreed on the absolute values of their ratings) amongst the four examiners showed less inter-rater reliability of 0.76, 95% [0.61—0.85].

Bivariate correlation results showed that Examiner 1 and Examiner 2 grades were highly correlated, $r = 0.73$, $P < 0.0001$ and examiners 3 and 4 had the poorest correlation, $r = 0.47$, $P < 0.0001$.

To evaluate the agreement between the scores assigned by the visual grading technique and those produced by the digital grading technique, we considered the ICC and the Spearman rank correlation coefficient between individual scores for each faculty evaluator, as well as the mean of the four faculty members' scores. Table 2 demonstrates those values.

The majority of the ICC values were in the low-to-moderate range. The highest agreement between the digital system and the four examiners was highest for examiners 1 and 2 and lowest for examiners 3 and 4. When we used the overall mean of faculty grades and compared to the digital grading system, the agreement dropped significantly to 0.58 with correlation coefficient of 0.63. The mean rating and individual ratings by the four examiners were consistently higher than the grades produced by the digital grading system.

The reproducibility of the grades using the Fair Grader 100® was excellent; Spearman rank correlation was $P = 0.998$, $P < 0.0001$ and ICC was .998, $P < 0.0001$. This high precision of the Fair Grader 100® qualified this device to be considered as the gold standard in further analysis. The authors used the Fair Grader 100® to assess the accuracy of each individual examiner's grades against grades produced by the digital system. The paired *t*-test demonstrated statistically significant differences between all examiners and the digital grading system by 6—25 grades. Examiner 2 on average had the smallest difference (digital grade – examiner visual grade); however, the standard deviation for this examiner was one of the largest, indicating high variation in his grades. Table 3 shows the paired *t*-test results.

Bland–Altman plots (Fig. 4) showed no systematic bias for the four examiners. This was demonstrated by having most of the differences clustered around the mean (the fixed bias) from both sides.

A close examination of the Bland–Altman plots shows the following:

- In plot #1(Examiner #1 – Digital grading), differences were clustered around their mean very closely compared to other plots and outliers greater or lower than the agreement lines were very few. This suggests that Examiner #1 had the best agreement with the digital grading system.
- The differences in plot #3 (Examiner #3 – Digital grading, *y*-axis) are dependent on the grade value (the average of the two methods on the *x*-axis). This was demonstrated by how differences have a pattern when they are scattered around the mean.

TABLE 2. Agreement between the digital grading system and each of the four examiners and their average ratings

Comparison	Intraclass correlation[a]	95% CI		Spearman's Correlation	
Digital grading and Examiner 1	0.65	0.15	0.83	$P = 0.63$	$P < 0.0001$
Digital grading and Examiner 2	0.54	0.30	0.69	$P = 0.40$	$P < 0.0001$
Digital grading and Examiner 3	0.30	−0.19	0.61	$P = 0.47$	$P < 0.0001$
Digital grading and Examiner 4	0.54	0.07	0.75	$P = 0.49$	$P < 0.0001$
Digital grading and average ratings	0.57	−0.041	0.798	$P = 0.63$	$P < 0.0001$

CI, confidence interval.

[a]Two-way mixed effects model where people effects are random and measures effects are fixed.

Discussion

Department of Fixed and Removable Prosthodontics at Al-Quds University is responsible for establishing the basic concepts and hand skills of different dental prosthetic procedures such as crown and bridge preparation. The department aims to prepare students for their future career in dentistry using the best available evidence and technology in teaching and in assessing students' dental work.

TABLE 3. Paired t-test to assess accuracy of different examiners' ratings

		Paired differences				
					95% CI of the Difference	
		Mean	SD	SEM	Lower	Upper
Pair 1	Examiner 1 – Software assessment	15.187	18.517	1.941	11.330	19.043
Pair 2	Examiner 2 – Software assessment	6.659	20.596	2.159	2.370	10.949
Pair 3	Examiner 3 – Software assessment	25.901	17.920	1.878	22.169	29.633
Pair 4	Examiner 4 – Software assessment	15.484	19.273	2.020	11.470	19.497

SD, standard deviation; SEM, standard error mean, CI, confidence interval.

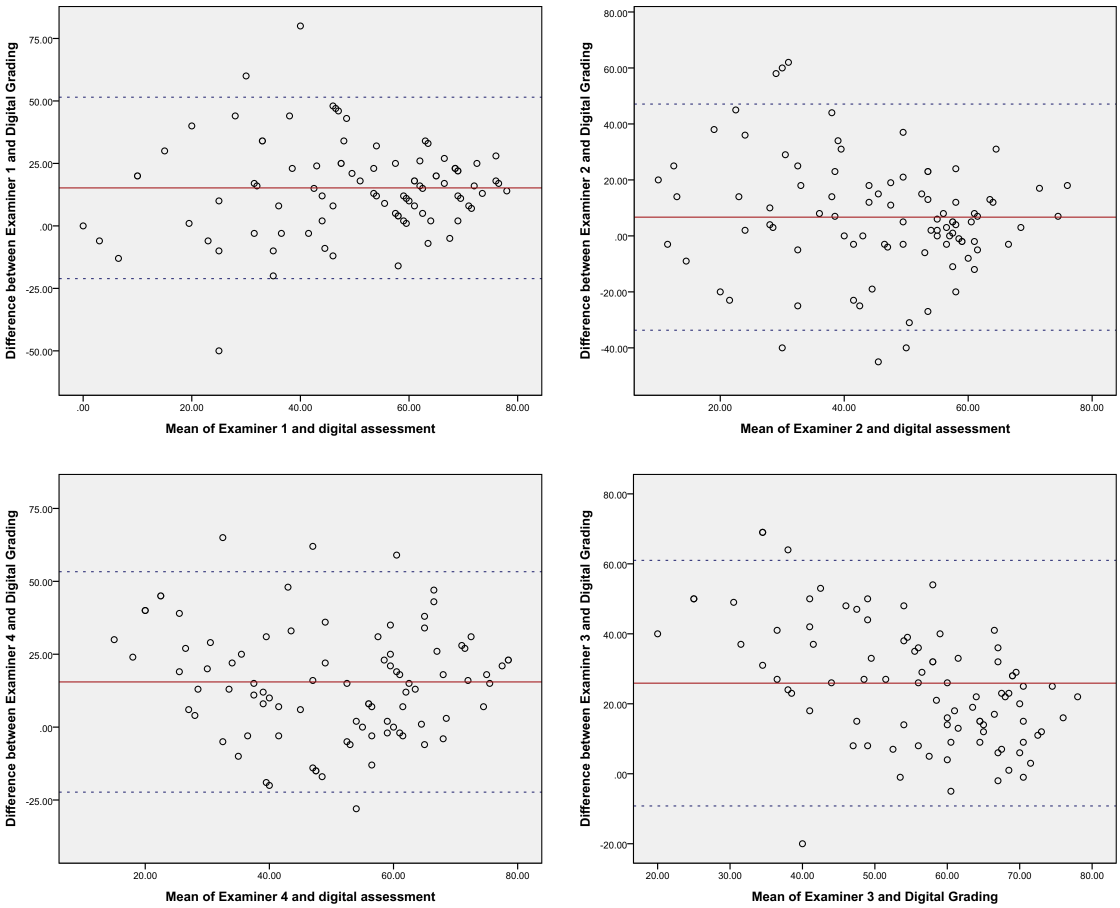


Fig. 4. Bland-Altman plots and limits of agreements.

Students usually trust technology more than human judgment when it comes to the evaluation of their work (21). They perceive the inconsistency in assessment feedback from different faculty as a bias, subjectivity or discrimination, which makes them tend to question or challenge the grade or the grading criteria most of the time (22). This distrust usually leads to shift the focus to what is the grade rather than what it actually represents. Grading pre-clinic exercises such as tooth preparation exercise is usually used to show the student the deviation from the ideal and to encourage him or her to identify deficiencies and work to improve (22).

Thus, in the current study, the department wanted to evaluate the use of a new digital grading system in minimising the variability that usually results from the traditional visual inspection. The department was hoping that using the new system to evaluate students' tooth preparations would increase the accuracy and the precision of the evaluation process and to take the subjectivity and the inconsistency out of the assessment process. A digital grading system with its high reliability and accuracy may shift back the focus of the students to the learning process and away from arguing the validity of their grade.

In the current study, students scored low by the digital grading system because this was their first exposure to the Fair Grader 100®. It is well known that before a student can perform a certain procedure, it is crucial that the student knows exactly what is to be achieved (7). Students in this study never used the digital system before for practice and although the tooth preparation criteria of the course didactic and laboratory instruction agrees with the digital system criteria, students still need to use the Fair Grader 100® in their daily practice to score higher in the final exam.

This system can provide valuable, on-time, feedback to students when they practise independently outside of established laboratory times. Fair Grader 100® provides students with a hard copy of their assessment report showing the gold standard abutment image, an image of their preparation and an overlapping comparison image. These images are colour-coded, indicating areas of deviation from ideal and where excessive or under cutting occurred. This detailed feedback, in addition to the high reliability and consistency of the digital grading system, demonstrated in this study should allow students to learn and develop these skills more efficiently. However, more research should investigate the importance of this new system in improving students' learning process. An interesting future research question can compare the grades of students in this study with next-year students who will be getting feedback from the digital grading system regularly over the course of the academic year for different pre-clinic exercises.

Another reason of the low grades students got in this exercise could be due to selecting the digital grading system default stringent tolerance level for this exercise. However, an advantage of the digital grading system is that the grade distribution can be adjusted by modifying the tolerance level without influencing the relative ranking of students' scores.

Reliability is the measure of how well examiners agree with others or themselves (7). Our results demonstrated very good intrarater reliability. Although neither of Examiner 1 nor Examiner 2 achieved the minimum acceptable intrarater agreement of 0.8., they were very close and they scored better than graders in other studies that reported fair to moderate agreement when grading pre-clinical intracoronal restorations and dental morphology exercise (4,14).

The 'absolute agreement' of the inter-rater reliability in this study was less than the intrarater agreement, which has been expected and reported in other studies (14, 23). Although the four examiners did not agree very well on the absolute grades, they agreed better on students' tooth preparations ranks. This means the student's work that ranked well by one examiner ranked well by other examiners.

The reproducibility of the digital grading system was almost perfect in this study (ICC, 0.99), which agrees with other studies that got excellent results of intrarater reliability of a different type of digital grading system, ICC of 0.93 (14) and ICC of 0.97 (22).

The agreement between the digital grading system and individual examiners was very modest. One explanation could be that comparing the two methods is invalid because digital grading system and visual grading in this application are measuring different things. This was even emphasised further by the results of the paired *t*-test which highlighted an important limitation of the Fair Grader 100®.

The Nissin Fair Grader 100® software main concept in evaluating tooth preparation is to compare the student's preparation in terms of over-reduction and/or under-reduction from a known gold standard. Obviously, other factors are important in terms of good tooth preparation such as smoothness of the surfaces and finish line configuration. Those other criteria that Nissin software did not address were important part in the faculty members' judgment which made the traditional visual inspection grades vary significantly from the digital grades.

Currently, the software can calculate and display taper, total occlusal convergence, reduction lingual wall and axial wall height and undercuts. Those pieces of information are very valuable in providing accurate feedback to students. However, this digital grading system needs to find a way to automatically evaluate the marginal configuration of the students' preparations as compared to the ideal.

Despite the previous limitation, the digital grading system seems to save the time of the busy faculty members. In this application, the time required for the four examiners to produce a grade that can be fair for the student was 3.5 times greater than what it took one teaching assistant to grade the same tooth preparations using the high reliable digital grading system.

Findings from this study suggest that there is great potential in introducing digital grading system technology in students' pre-clinic dental work evaluation. However, any investment in technology should be founded on a good understanding of real and expected outcomes. That is why more research is needed to better understand what works and what does not and to use this feedback to improve the outcomes, and the efficiency of learning. More detailed score that addresses different assessment criteria of tooth preparation would be a good improvement of the digital grading system's current general score. This will help students to self-assess their performance and improve their learning experience.

Conclusion

This study demonstrates that a digital grading system like Fair Grader 100® can reliably scan and compare student tooth preparations to a known gold standard. Using this method will preclude the variability and the subjectivity that usually result from the traditional visual inspection grading and makes it feasible to accurately and consistently assess students' work. In addition, this new method will provide immediate feedback to students in examinations and daily practice, and reduce the workload of faculty members, which is a huge plus.

Conflict of interests

The authors declare that there is no affiliation or involvement in an organisation or entity with a financial or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1 Renne WG, McGill ST, Mennito AS, et al. E4D compare software: an alternative to faculty grading in dental education. J Dent Educ 2013: 77: 168–175.
2 Krupat E, Dienstag JL. Assessment is an educational tool. Acad Med 2009: 84: 548–550.
3 Feil PH, Gatti JJ. Validation of a motor skills performance theory with applications for dental education. J Dent Educ 1993: 57: 628–633.
4 Lilley JD, Bruggen Cate HJ, Holloway PJ, Holt JK, Start KB. Reliability of practical tests in operative dentistry. Br Dent J 1968: 125: 194–197.
5 Fuller JL. The effects of training and criterion models on inter-judge reliability. J Dent Educ 1972: 36: 19–22.
6 Salvendy G, Hinton WM, Ferguson GW, Cunningham PR. Pilot study on criteria in cavity preparation. J Dent Educ 1973: 37: 27–31.
7 Sharaf AA, AbdelAziz AM, El Meligy OAS. Intra- and inter-examiner variability in evaluating preclinical pediatric dentistry operative procedures. J Dent Educ 2007: 71: 540–544.
8 Hauser AM, Bowen DM. Primer on preclinical instruction and evaluation. J Dent Educ 2009: 73: 390–398.
9 Arevalo CR, Bayne SC, Beeley JA, et al. Framework for e-learning assessment in dental education: a global model for the future. J Dent Educ 2013: 77: 564–575.
10 Commission on Dental Accreditation. Accreditation standards for dental education programs. Chicago, IL: American Dental Association, 2006.
11 Schiff AJ, Salvendy G, Root CM, Ferguson GW, Cunningham PR. Objective evaluation of quality in cavity preparation. J Dent Educ 1975: 39: 92–96.

12 Friedman CP. The research we should be doing. Acad Med 1994: 69: 455–457.

13 Morrow JA, Pulido MT, Smith PB, McDaniel TF, Willcox AB. Effective use of e-grading in the dental simulation clinic. J Dent Educ 2014: 78: 829–837.

14 Kwon SR, Restrepo-Kennedy N, Dawson DV, et al. Dental anatomy grading: comparison between conventional visual and a novel digital assessment technique. J Dent Educ 2014: 78: 1655–1662.

15 McPherson KR, Mennito AS, Vuthiganon J, et al. Utilizing self-assessment software to evaluate student wax-ups in dental morphology. J Dent Educ 2015: 79: 697–704.

16 Nissin Fair Grader official website. Available at: http://www.nissin-dental.cn/product/detail.html?id=241. Retrieved November 28 2015

17 Khoza-Shangase K, Kassner L. Automated screening audiometry in the digital age: exploring uhear™ and its use in a resource-stricken developing country. Int J Technol Assess Health Care 2013: 29: 42–47.

18 Hanneman SK. Design, analysis, and interpretation of method-comparison studies. AACN Adv Crit Care 2008: 19: 223–234.

19 Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil 1998: 12: 187–199.

20 IBM Corp. Released 2011. IBM SPSS statistics for windows, version 20.0. Armonk, NY: IBM Corp.

21 Junco R, Mastrodicasa J. Connecting to the net generation: what higher education professionals need to know about today's students. Washington, DC: NASPA, 2007.

22 Knight GW. Toward faculty calibration. J Dent Educ 1997: 61: 941–946.

23 Satterthwaite JD, Grey NJA. Peer-group assessment of preclinical operative skills in restorative dentistry and comparison with experienced assessors. Eur J Dent Educ 2008: 12: 99–102.