ACP AMERICAN COLLEGE OF PROSTHODONTISTS
Your smile. Our specialty.®

Effects of Evaluator's Fatigue and Level of Expertise on the Global and Analytical Evaluation of Preclinical Tooth Preparation

Mohammad D. Al Amri, BDS, MS, FRCDC, Haneef R. Sherfudhin, BDS, MDS, & Syed R. Habib, BDS, FCPS

Department of Prosthetic Dental Sciences, College of Dentistry, King Saud University, Riyadh, Saudi Arabia

Keywords
Evaluation; grading; tooth preparation; expertise; fatigue; rubric.

Correspondence
Mohammad D. Al Amri, Department of Prosthetic Dental Sciences, College of Dentistry, King Saud University, P. O. Box 60169, Riyadh 11545, Saudi Arabia.
E-mail: dramri@yahoo.com.

Funding provided by College of Dentistry Research Center and Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia (research project # FR 0298).

The authors declare that they have no conflicts of interest.

Accepted September 15, 2016

doi: 10.1111/jopr.12558

Abstract

Purpose: To investigate the effects of evaluator fatigue and level of expertise on the grading of preclinical tooth crown preparations, by global and analytical methods of evaluation.

Materials and Methods: The study had a double-blind design. Two faculty members, each with more than 10 years of clinical and teaching experience, and two demonstrators with no teaching experience evaluated tooth preparations on maxillary central incisors and mandibular first molars. As a test of the effect of fatigue, preparations were globally (subjective grading) and analytically (criteria-based grading) graded on day 1 (after evaluators had been on duty continuously for 8 hours) and day 2 (in the morning after evaluators had sufficient sleep). Evaluators worked under the same circumstances and did not communicate with each other. The assigned textbooks were used to develop the criteria for grading (rubric) and the predefined exclusion criteria. Grades were recorded and statistically analyzed using statistical software. The paired-sample *t*-test and Mann-Whitney U test were used for multiple comparisons. Level of significance was set at $p < 0.05$.

Results: An inconsistency in preclinical tooth preparation evaluation was found to exist by both global and analytical methods. Junior faculty tended to award higher grades than senior faculty did. Furthermore, higher grades were scored by the analytical method. More clinical and academic experience did not guarantee intra- and interexaminer reliability. Younger faculty appeared to tolerate fatigue better than older faculty. Likewise, global evaluation appeared to be more influenced by fatigue than was the analytical method.

Conclusion: There were variations in grading, with no consistently preferred grading method. Evaluator performance after continuous 8-hour duty had no significant effect on preclinical tooth preparation evaluation. Level of expertise did not affect preclinical evaluation regardless of the grading method used.

A plethora of grading tools, including older mechanical and newer computer-based systems, has been developed over the years for the objective assessment of dental students' clinical work.[1-3] Tooth preparation constitutes a basic technique for dental treatment and therefore plays an important role in preclinical and clinical dental education. Reliability in preclinical or clinical evaluation can be a concern to faculty who must make these assessments, and any lack of evaluation consistency can lead to stress and confusion for dental students.[1] Variations in the grading of tooth preparations have been reported by academic clinicians, showing a lack of consistency.[4] Myers[5] concluded that a source of frustration for both dental students and clinical instructors is the subjectivity associated with the clinical evaluation of student performance. Others also reported significant disagreement between examiners, as well as wide intra-examiner variation when the same rater evaluated the same operative procedure on a second occasion.[6,7] In more recent years, investigations have concentrated on the development of grading systems based on specific checklists and criteria (analytical) as an alternative to the 'glance and grade' (global) method, to improve rater reliability.[8,9]

Investigators have reported attempts to develop reliable laboratory and clinical evaluation systems.[6,7,9-11] However, few studies, regardless of level of success, have used an analytical procedure to identify those components of the evaluation system that, if refined further, could improve reliability.[10] The ongoing challenges for dental educators are implementing and identifying effective methods for assessing dental student

performance. One solution to grading dilemmas can be found in the use of rubrics, scaled tools with levels of achievement and clearly defined criteria placed in a grid.[8] According to O'Donnell et al,[8] "Rubrics establish clear rules for evaluation and define the criteria for performance. Rubrics speak to both teaching and learning expectations and outcomes and can provide faculty members with a tool that can be useful in evaluating dental student performance."[8] Rubrics can also provide students with clear expectations of performance, an opportunity to self-assess, and timely, detailed feedback.[9,10] Goepferd and Kerber[12] and Vann et al[13] reported that the development of an analytical approach using detailed checklists (rubrics) improved examiner reliability; however, Vann et al[13] reported no difference between "glance and grade" and checklist methods of grading.

Several factors can be anticipated to influence grading.[14-18] Studies have investigated the influence of clinical site,[16] student's and preceptor's gender,[16] and preceptor's academic rank and level of experience on grading.[18] Preceptors with higher academic ranks tended to award better grades, whereas female preceptors gave significantly lower grades, regardless of the student's gender, and created a higher failure rate.[16] This suggests an interaction in the grading process between the gender of the examinees and the gender of their attending evaluators. Fatigue has been associated with professional disability and disengagement.[19-32] Its adverse effect on clinical (psychomotor) and cognitive skills of dentists and physicians has been studied in abundant indexed publications.[25-32] However, research on grading performance has focused on interexaminer and intra-examiner reliability, with no attention to fatigue and its effect on preclinical and clinical evaluations.[6-11,33]

Students tend to perceive that evaluation methods are somewhat arbitrary, due to a lack of examiner consistency. This concept can produce a negative effect on undergraduate confidence and performance level, and undermine the learning process. We hypothesized that evaluator fatigue and level of expertise would affect the grading of a tooth preparation. Therefore, this study was conducted to investigate the effects of evaluator fatigue and level of expertise on the grading of preclinical tooth preparations by both global and analytical methods of evaluation.

Materials and methods

Study population and design

All included students agreed to be part of this investigation and signed the consent form. The study was approved by the College of Dentistry Research Centre (CDRC # FR 0298). It was conducted in a preclinical prosthodontics course for third-year students at the College of Dentistry, King Saud University, Riyadh, Saudi Arabia. Seventy-four students enrolled in the course participated in the study. The students were asked to prepare two ivory teeth (maxillary central incisor and mandibular first molar) on dentoforms (PRO2001-UL-UP-FEM-32; Nissin Dental Products Inc., Kyoto, Japan) for metal ceramic (MC) crown preparations in 140 minutes. All the students had the same level of experience and training, with no repeaters among them. Environment, equipment, and instruments were kept standard.

Four examiners with two levels of expertise (two faculty prosthodontists, each with more than 10 years of clinical and teaching experience and two demonstrators who had graduated from the same program 1 year previously) evaluated tooth preparations using the global and analytical methods of evaluation. Both evaluations were anonymously and simultaneously performed by the four evaluators under the same circumstances without communication among them. The study had a blind design with anonymously representative numbers (from 1 to 74) for students and letters for evaluators (A, B, C, and D).

Criteria for evaluation (rubrics)

To standardize time allocated for grading, evaluators were given a maximum of 2 minutes per tooth to award a grade. Global (subjective) evaluation was carried out according to the evaluators' personal experience and definition according to their educational/scientific background. For the analytical (objective or criteria-based) method of grading, the evaluators and students were familiarized with the evaluation criteria in advance through direct instructions. Furthermore, detailed instructions were given to the students 3 days before the exercise and again immediately before the examination. The rubric used was developed according to the published criteria and defined principles of tooth preparation from the assigned textbooks (Table 1).[34,35] Moreover, the rubric reliability was validated through a pre-exercise pilot study of 10 standard tooth preparations under similar times, conditions, and circumstances. Evaluators repeated grading twice with a 5-day interval in between. The intra- and interrater agreements were satisfactorily high (kappa scores were 0.90 and 0.86, respectively). All students were asked to fabricate a preoperative putty silicone (Virtual putty; Ivoclar Vivadent AG, Schaan, Liechtenstein) index (stent) to help evaluate the amount of tooth reduction.

Exclusion criteria

Exclusion criteria included the following:

1. Altered dentoforms, including any evidence of tampering or attempts to remove the screws after the dye was applied.
2. Major tissue damage (soft or hard) resulting in lacerations, major damage to the soft tissue, incorrect tooth prepared or improper extension of the preparation, adjacent tooth damaged requiring restoration.
3. Excessive over- or under-reduction that does not reflect the principles of acceptable tooth preparation with the preoperative silicone putty index.
4. Continuing to work after the published cut-off time (140 minutes).
5. Working on a manikin model in a manner that does not simulate actual patient conditions.
6. Failure to complete the examination within the allotted time. (No grace period, make-up time, or second effort was allowed for any part of this examination.)
7. Receiving assistance from another person during the course of the examination.
8. Preparing a tooth other than the one approved by the examiners. This is considered major hard tissue damage.

Table 1 Criteria (rubrics) used to evaluate tooth preparations for metal-ceramic restoration

	Grading criteria			
Feature	1.5 points	1.0 point	0.5 point	0 point
Occlusal/incisal reduction	Proper reduction - Supporting areas or porcelain = 1.5-2.0 mm - Nonsupporting areas or metal = 1.0-1.5 mm - Incisal reduction = 2 mm	Moderately under-reduced - Supporting areas or porcelain <1.5 mm - Nonsupporting areas or metal <1.0 mm - Incisal reduction = 1.5 mm	Moderately over-reduced - Supporting areas or porcelain ≥2.5 mm - Nonsupporting areas or metal ≥2 mm - Incisal reduction = 2.5 mm	Severely over-reduced or under-reduced - Supporting areas or porcelain >3.0 mm or <1.0 mm - Nonsupporting areas or metal >2.5 mm or <0.5 mm - Incisal reduction >3 mm or <1 mm
Axial reduction	Proper reduction – 0.7 to 1 mm for metal – 1.2 to 1.5 mm for	Moderately under-reduced – <0.7 mm for metal, <1.2 mm for porcelain or – Lack of rounded line angles or point angles	Moderately over-reduced – 1.5 to 2 mm for veneered metal – 1 to 1.5 mm for metal	Severely over-reduced or under-reduced – >2 mm for veneered metal or >1.5 mm for metal – <1 mm for veneered metal, or <0.5 mm for metal
Two-plane reduction	veneered metal, and rounded line angles and point angles	Preparation has two planes providing adequate material bulk for strength/esthetics		Preparation has single-plane reduction providing inadequate material bulk for strength/esthetics
Taper	Retentive walls: 6°-10° for anterior teeth or up to 15° for posterior teeth	Taper present but nearly parallel (< 6°) Over-tapered reduction: anterior, 10°-15°; posterior, 15°-20°	Undercuts visually present Over-tapered reduction: anterior, 15°-20°; posterior, 20°-25°	Severe undercuts present Severe overtapering of any axial surface: anterior teeth >20°; posterior teeth >25°
Margin placement	Margins extended to specified target (even with free gingival margin or 0.5 mm supragingivally)	Moderately overextended (not more than 0.5 mm subgingivally) or moderately under-extended (not more than 1 mm supragingivally)	Significantly over-extended (not more than 1.0 mm subgingivally) or significantly under-extended (not more than 1.5 mm supragingivally)	Severely overextended (more than 1.0 mm subgingivally) or severely underextended (more than 1.5 mm supragingivally)
Finish, margins, and walls (heavy chamfer)		- Margins and walls are smooth - Margins are continuous and well-defined	- Moderate roughness of margins and walls - Margins are moderately noncontinuous with moderate lack of definition	- Significant roughness of margins and walls - Margins are noncontinuous - Lack of definition of finish line
Preservation of adjacent tissue		Adjacent teeth and gingiva are unaffected by preparation	Adjacent teeth and gingiva are minimally damaged	Adjacent teeth and gingiva are severely damaged
Time management		Student ends the examination 20 minutes early (120 min)	Student ends the examination 10 minutes early (130 min)	Student ends the examination on time (140 min)

9. Any incomplete grading forms with at least one missing grade were eliminated.

Testing the effect of fatigue (duty-related)

For investigation of the effect of the evaluator's fatigue on grading, two evaluations were conducted under two different circumstances and at two different times (day 1 and day 2). On day 1, the evaluation session took place immediately after the examination exercise at the end of the working day (from 5:00 to 8:00 p.m.) after the evaluators had completed 8 hours of daily continuous academic and clinical duties. All four evaluators described being exhausted after having a long day before participating in the grading. The second evaluation (day 2) was scored during the weekend at early morning (from 8:00 to 11:00 a.m.) after evaluators had at least 8 hours of sleep. All four evaluators reported being fresh and energized after having sufficient rest and sleep.

Table 2 Descriptive statistics (means ± SD) of global and analytical evaluations on days 1 and 2

	Grading method	Level of expertise	Mean	SD	N
Day 1	Global	Faculty	6.90	0.83	296
		Demonstrators	7.30	0.83	296
		Total	7.10	0.85	592
	Analytical	Faculty	7.30	1.35	296
		Demonstrators	7.60	1.07	296
		Total	7.50	1.22	592
	Total	Faculty	7.10	1.14	592
		Demonstrators	7.40	0.97	592
		Total	7.30	1.07	1184
Day 2	Global	Faculty	7.40	4.72	296
		Demonstrators	7.20	0.67	296
		Total	7.30	3.36	592
	Analytical	Faculty	7.20	1.35	296
		Demonstrators	7.60	1.00	296
		Total	7.40	1.21	592
	Total	Faculty	7.30	3.47	592
		Demonstrators	7.40	0.87	592
		Total	7.30	2.53	1184

SD = standard deviation.

Table 3 Means (± SD) and significance levels of evaluators using the global and analytical evaluations

	Global	Analytical	
Evaluators	Mean ± SD	Mean ± SD	Significance (*p*)
Faculty 1 (A)	7.01 ± 0.63	7.05 ± 1.37	0.69
Faculty 2 (B)	6.95 ± 1.00	7.70 ± 1.25	<0.001
Demo 1 (C)	6.97 ± 0.89	7.56 ± 1.26	<0.001
Demo 2 (D)	7.66 ± 0.61	7.74 ± 0.84	0.11

Statistical analysis

Grades were recorded and analyzed by SPSS for Windows, v20 (SPSS Inc., Chicago, IL), with the paired-sample *t*-test and Mann-Whitney U test for multiple comparisons. The level of significance was set at $p < 0.05$.

Results

Descriptive statistics

In total, 74 male dental students and four male full-time faculty members (two faculty prosthodontists and two demonstrators) participated in the present investigation, with no drop-outs, exclusions, or missing grades. The average ages of students and evaluators were 24 ± 0.84 (from 23 to 26) and 37 ± 12.50 (from 28 to 55) years, respectively. On days 1 and 2, 2368 total grades were submitted by the four evaluators (Table 2). There was no interaction between the method of grading and the level of expertise ($p = 0.68$). Comparison of the grades of the two evaluation methods showed that analytical grading consistently yielded higher mean scores than did global grading (Table 3). This difference was significant only for Faculty 2 and Demonstrator 1 ($p < 0.001$). Furthermore, higher mean grades were awarded by junior faculty compared with senior faculty. When grades for days 1 and 2 were compared, it was found that higher grades were given after evaluators had sufficient rest and sleep (day 2).

Results related to the effect of expertise on grading

A statistically significant difference ($p < 0.001$) was found between Demonstrator 2 and the rest of the examiners for the global method of grading (Table 4). For the analytical method, a statistically significant difference ($p < 0.01$) was found between Faculty 1 and the rest of the examiners. For both methods, a statistically significant difference was found between Faculty 1 and Demonstrator 2 ($p < 0.001$); however, there was no statistically significant difference between the global and analytical methods of grading for Faculty 2 and Demonstrator 1 ($p = 0.99$ and 0.78, respectively) (Table 4). The results also showed a significant difference ($p < 0.05$) between the grading of anterior and posterior teeth only when the analytical method was used.

Results related to the effect of fatigue on grading

The mean global and analytical grades, correlations, and statistical significance are presented in Table 5. The intra-examiner difference in grading was not significant between days 1 and 2 for global ($p = 0.16$) and analytical ($p = 0.93$) methods. Faculty 2, using global grading, demonstrated the highest difference between days 1 and 2. The highest (0.76) and lowest (0.24) correlations between the grades on days 1 and 2 were achieved by the same evaluator (Faculty 2) using the analytical and global methods, respectively. The overall correlation between the grades on days 1 and 2 was high with the analytical method (0.71) and weak for the global method (0.22) of grading. Plotted comparisons of grades on days 1 and 2 with both the global and analytical methods are presented in Figures 1 and 2, respectively.

Discussion

Tooth crown preparation was selected for investigation because it constitutes a basic technique for dental treatment and therefore plays an essential role in preclinical dental education.[2,36,37] Moreover, learning the widely differing forms of tooth preparation places high demands on both students and faculty in terms of 3D conceptualization, precision, reproducibility, and evaluation.[37] The shortage of senior faculty in dental schools places an extra burden on the existing staff regarding consistent student evaluation to avoid conflicts with the students. One way of overcoming this problem is by the use of digital devices in dental education. Another short-term response to this shortage can be the hiring of dental specialty residents, demonstrators, and part-time contributors as student instructors. Therefore, the purpose of this study was to investigate the influence of fatigue and level of expertise on the evaluation of tooth crown preparation done by dental students in a preclinical fixed prosthodontics course.

Table 4 Interexaminer comparisons for both grading methods

	Global grading			Analytical grading		
Comparison	Mean difference	Std. error	Sig.	Mean difference	Std. error	Sig.
Faculty 1 vs Faculty 2	0.06	0.10	0.94	−0.65*	0.15	<0.001
Faculty 1 vs Demo 1	0.04	0.09	0.96	−0.51*	0.15	0.01
Faculty 1 vs Demo 2	−0.65*	0.07	<0.001	−0.70*	0.13	<0.001
Faculty 2 vs Demo 1	−0.01	0.11	0.99	0.14	0.15	0.78
Faculty 2 vs Demo 2	−0.70*	0.09	<0.001	−0.04	0.12	0.99
Demo 1 vs Demo 2	−0.69*	0.09	<0.001	−0.18	0.13	0.46

*The mean difference was significant at $p < 0.05$.

Table 5 Mean grades and SDs using global and analytical evaluations on days 1 and 2

	Evaluator	Mean grades on day 1 (SD)	Mean grades on day 2 (SD)	Correlation	*p* Value
Global evaluation (N = 296)	Faculty 1	7.01 (0.62)	6.97 (0.68)	0.68	0.48
	Faculty 2	6.95 (1.00)	7.87 (6.63)	0.24	0.09
	Demo 1	6.96 (0.89)	7.09 (0.69)	0.61	0.04
	Demo 2	7.65 (0.60)	7.40 (0.61)	0.61	<0.001
	Mean	7.14 (0.85)	7.33 (3.36)	0.22	0.16
Analytical evaluation (N = 296)	Faculty 1	7.04 (1.36)	6.73 (1.35)	0.73	<0.001
	Faculty 2	7.69 (1.25)	7.70 (1.18)	0.76	0.92
	Demo 1	7.56 (1.26)	7.64 (1.21)	0.62	0.34
	Demo 2	7.74 (0.84)	7.70 (0.72)	0.69	0.51
	Mean	7.51 (1.17)	7.51 (1.18)	0.71	0.93

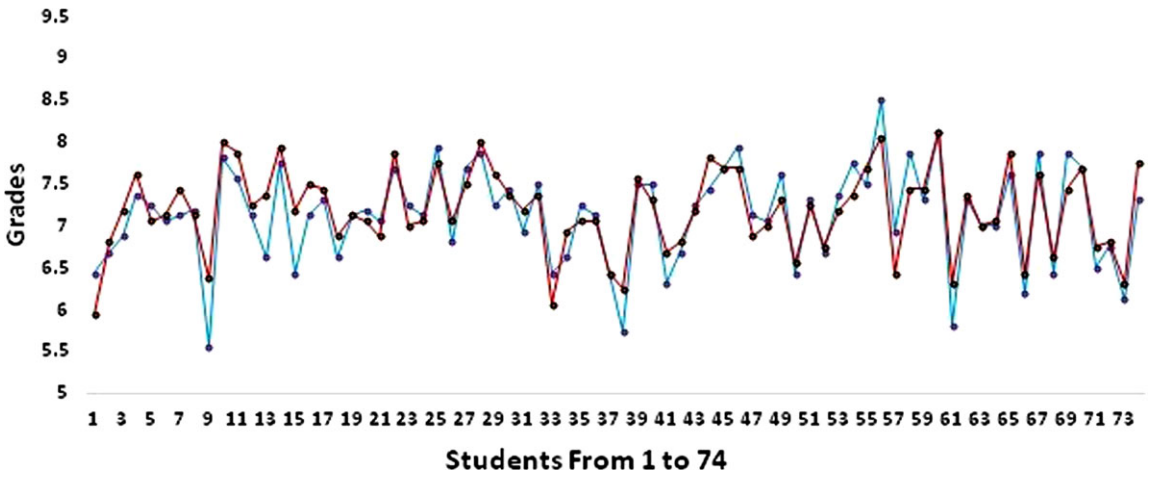

Figure 1 Comparison of global grades of the students on day 1 (blue) and day 2 (red).

The benefits of using dental specialty residents and demonstrators as student instructors have been demonstrated.[38,39] These instructors can serve to compensate for faculty shortage, thereby improving the student-instructor ratio and the quality of education provided. It has been suggested that as instructor experience increases, intra- and interexaminer variability decreases.[40] Nimmo et al[38] compared standardized teaching evaluations for full-time faculty and freshly graduated instructors, and found no statistically significant difference in students' ratings of faculty members and freshly graduated instructors for preclinical prosthodontics and occlusion courses.[39] Furthermore, other researchers[14,15] have reported good agreement in grading between senior and junior faculty members or senior students. Likewise, the present results showed that level of expertise had no significant effect on grading.

The present study showed that inconsistency among examiners (senior vs. junior faculty) in evaluating a preclinical tooth preparation for a full veneer MC restoration by no specified

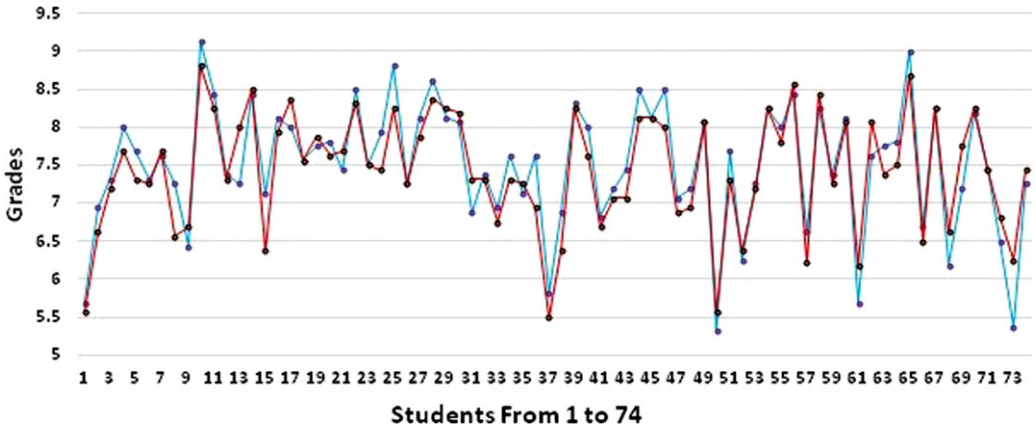

Figure 2 Comparison of analytical grades of the students on day 1 (blue) and day 2 (red).

method (global vs. analytical) resulted in superior interexaminer reliability. This observation is in agreement with earlier reports.[3,6,7,10,37] However, several factors could have influenced the current outcomes. One critical factor is the validity and reliability of the rubric used. In the development of effective rubrics, an important factor is the use of clear criteria to rate students' work, and direct observation of the performance being evaluated.[37] Therefore, the rubric used was presented to evaluators and students and pre-validated with adequate intra- and interrater agreements. Although it was not clear if global grading was superior to analytical grading in this study, it was interesting that all examiners (faculty members and junior demonstrators) scored higher grades with the analytical grading method. This could be due to students receiving maximum or minimum scores for different individual criteria within the analytical grading, leading to overall higher scores; whereas, in the global method, the same student may get a lower score because the grading is based on the overall clinical acceptability of the preparation. Another factor to be considered here is the inclusion of time management as one of the criteria within analytical grading, since it may not be considered during global grading.

It has also been reported that cognitive and psychomotor skills are negatively affected by fatigue during tasks that require a combination of psychomotor and cognitive skills.[28] However, continuous 8-hour duty had no significant effect on preclinical tooth grading in the present study. It was hypothesized that differences would be significant if duty was for a longer period of time and that younger individuals tend to tolerate fatigue more than older ones. This assumption could in part justify the lowest correlation between the 2-day grades scored by the oldest evaluator (55 years old). Fatigue is an individual response and must be countered in an individualized manner. Although frequent short breaks and relaxation exercises can reduce this type of stress, all grading should be performed in one sitting.[23,24,32] Further studies are needed to test these hypotheses.

As with any investigation, there were limitations to this study. First, although evaluations were repeated 74 times by all instructors in all sessions, the small number of evaluators is a concern in the current investigation. Therefore, future studies with larger number of assessors are recommended. Second, the rubric used in this investigation was not revised and evaluated by external audit body. Also, there were no attempts for training or calibration of evaluators other than during the pilot study. It was hypothesized that training and calibration against a gold standard would have improved interrater agreements; however, validation tests indicated reliable agreements. Furthermore, the large number of evaluations per session ($n = 296$) undertaken by each instructor could have affected evaluators' performance and judgments. Although inclusion of only male participants and evaluators could be considered a limitation of the present study, it is pertinent to mention that the intentional inclusion of unisex participants could have helped eliminate any potential gender bias, which has been reported in previous studies.[16,17] Level of expertise in the current study was merely based on credentials and years of service, irrespective to other potential co-factors like age and postgraduate program. The fact that this study was conducted in association with one clinical procedure within a single department at one academic institution indicates that the present results may not generalize to all such institutions or other departments. Finally, a common limitation in investigating fatigue effect is how fatigue was defined. Although the definition used in this study reflects the daily activities of dental educators who usually undertake such evaluations, it was subjective and exclusive of other somatic effects; however, as a feeling of tiredness, fatigue should be distinguished from weakness. Therefore, e-grading software programs have been advocated to overcome these human-associated limitations. Nevertheless, such systems are expensive, and their benefits have not been fully supported in the literature.Future studies are warranted to improve conventional methods of evaluation and to investigate the use of simulation technology in the assessment of students' psychomotor learning potential.

Conclusions

Within the limits of this study, the following can be concluded:

1. Variability in preclinical tooth preparation evaluation exists in both global and analytical methods. Junior faculty

tended to award higher grades as compared with senior faculty. Furthermore, insignificant higher grades were scored by the analytical method.
2. Level of expertise did not affect preclinical evaluation regardless of the grading method used. More clinical and academic experience did not guarantee intra- and interrater reliability.
3. Evaluator performance after continuous 8-hour duty had no significant effect on the evaluation of preclinical tooth preparation.
4. A high correlation between grades on days 1 and 2 was associated with the analytical evaluation method and demonstrators. The global evaluation and senior faculty seemed to be more influenced by fatigue than were their counterparts.
5. Further randomized controlled clinical and preclinical trials are recommended to investigate factors affecting evaluation in dental education. Attention and efforts should be focused on standardization and reliability.

Acknowledgments

The authors would like to thank the College of Dentistry Research Center and Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia for funding this research project (research project # FR 0298).

References

1. Schiff AJ, Salvendy G, Root CM, et al: Objective evaluation of quality in cavity preparations. J Dent Educ 1975;39:92-96
2. Quinn F, Keogh P, McDonald A, et al: 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;7:164-169
3. Esser C, Kerschbaum T, Winkelmann V, et al: A comparison of the visual and technical assessment of preparations made by dental students. Eur J Dent Educ 2006;10:157-161
4. Seymour KG, Samarawickrama DY, Zou L, et al: Assessing the quality of shoulder preparations for metal ceramic crowns. Eur J Prosthodont Restor Dent 1999;7:125-129
5. Myers B: Beliefs of dental faculty and students about effective clinical teaching behaviors. J Dent Educ 1977;41:68-76
6. Fuller JL: The effects of training and criterion models on interjudge reliability. J Dent Educ 1972;36:19-22
7. Lilley JD, ten Bruggen Cate HJ, Holloway PJ, et al: Reliability of practical tests in operative dentistry. Br Dent J 1968;125:194-197
8. O'Donnell JA, Oakley M, Haney S, et al: Rubrics 101: A primer for rubric development in dental education. J Dent Educ 2011;75:1163-1175
9. Mould MR, Bray KK, Gadbury-Amyot CC: Student self-assessment in dental hygiene education: A cornerstone of critical thinking and problem-solving. J Dent Educ 2011;75:1061-1072
10. Sharaf AA, AbdelAziz AM, El Meligy OA: Intra- and inter-examiner variability in evaluating preclinical pediatric dentistry operative procedures. J Dent Educ 2007;71:540-544
11. Rodger S, Coleman A, Caine AM, et al: Examining the inter-rater and test-retest reliability of the student practice evaluation form-revised (spef-r) for occupational therapy students. Aust Occup Ther J 2014;61:353-363
12. Goepferd SJ, Kerber PE: A comparison of two methods for evaluating primary class ii cavity preparations. J Dent Educ 1980;44:537-542
13. Vann WF, Machen JB, Hounshell PB: Effects of criteria and checklists on reliability in preclinical evaluation. J Dent Educ 1983;47:671-675
14. Nick DR, Clark M, Miler J, et al: The ability of dental students and faculty to estimate the total occlusal convergence of prepared teeth. J Prosthet Dent 2009;101:7-12
15. 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;12:99-102
16. Golubic R, Golubic K: What do grades in clinical subjects depend on? Case study of the zagreb university school of medicine. Croat Med J 2004;45:67-71
17. Rand VE, Hudes ES, Browner WS, et al: Effect of evaluator and resident gender on the american board of internal medicine evaluation scores. J Gen Intern Med 1998;13:670-674
18. Park RD, Susarla SM, Howell TH, et al: Differences in clinical grading associated with instructor status. Eur J Dent Educ 2009;13:31-38
19. Humphris GM, Kaney S: Examiner fatigue in communication skills objective structured clinical examinations. Med Educ 2001;35:444-449
20. Gorter RC, Storm MK, te Brake JH, et al: Outcome of career expectancies and early professional burnout among newly qualified dentists. Int Dent J 2007;57:279-285
21. Pohlmann K, Jonas I, Ruf S, et al: Stress, burnout and health in the clinical period of dental education. Eur J Dent Educ 2005;9:78-84
22. Smith MR, Marcora SM, Coutts AJ: Mental fatigue impairs intermittent running performance. Med Sci Sports Exerc 2015;47:1682-1690
23. Wasoski RL: Stress, professional burnout and dentistry. J Okla Dent Assoc 1995;86:28-30
24. Harris NO, Crabb LJ: Ergonomics. Reducing mental and physical fatigue in the dental operatory.Dent Clin North Am 1978;22:331-345
25. Gerdes J, Kahol K, Smith M, et al: Jack Barney award: The effect of fatigue on cognitive and psychomotor skills of trauma residents and attending surgeons. Am J Surg 2008;196:813-819; discussion 819-820
26. Vered Y, Zaken Y, Ovadia-Gonen H, et al: Professional burnout: Its relevance and implications for the general dental community. Quintessence Int 2014;45:87-90
27. Gorter RC, Freeman R: Burnout and engagement in relation with job demands and resources among dental staff in northern ireland. Commun Dent Oral Epidemiol 2011;39:87-95
28. Kahol K, Leyba MJ, Deka M, et al: Effect of fatigue on psychomotor and cognitive skills. Am J Surg 2008;195:195-204
29. Yamany T, Woldu SL, Korets R, et al: Effect of postcall fatigue on surgical skills measured by a robotic simulator. J Endourol 2015;29:479-484
30. Smith MR, Coutts AJ, Merlini M, et al: Mental fatigue impairs soccer-specific physical and technical performance. Med Sci Sports Exerc 2016;48:267-276
31. Hopstaken JF, van der Linden D, Bakker AB, et al: A multifaceted investigation of the link between mental fatigue and task disengagement. Psychophysiology 2015;52:305-315
32. Fukui K, Arai E: [mental fatigue of dentists during dental treatment with special regard to the results of the brain "excite meter" test]. Josai Shika Daigaku Kiyo 1986;15:173-182

33. Dhuru VB, Rypel TS, Johnston WM: Criterion-oriented grading system for preclinical operative dentistry laboratory course. J Dent Educ 1978;42:528-531
34. Shillingburg HT, Hobo S, Whitsett LD, et al: Preparations for full veneer crowns. In Shillingburg HT, Hobo S, Whitsett LD, et al (eds): Fundamentals of Fixed Prosthodontics (ed 3). Chicago, Quintessence, 1997, pp 142-151
35. Rosenstiel SF, Land MF, Fujimoto J: The metal-ceramic crown preparation. In Rosenstiel SF, Land MF, Fujimoto J (eds): Contemporary Fixed Prosthodontics, Vol 1 (ed 4). St. Louis, Elsevier Mosby, 2006, pp 272-285
36. Nishida M, Sohmura T, Takahashi J: Training in tooth preparation utilizing a support system. J Oral Rehabil 2004;31:149-154
37. Karl M, Graef F, Wichmann M, et al: Evaluation of tooth preparations - a comparative study between faculty members and pre-clinical students. Eur J Dent Educ 2011;15: 250-254
38. Nimmo A, Mitchell GS, Penfield RD, et al: Evaluation of dental students as instructors in preclinical prosthodontics and occlusion courses. J Prosthodont 2007;16:400-405
39. Nimmo A, Mitchell GS, Echeto L, et al: Effect of dental students as instructors on preclinical performance in prosthodontics. J Dent Educ 2008;72:1488-1492
40. Bedi R, Lo E, King NM, et al: The effect of pictorial criteria upon the reliability of assessments of cavity preparations. J Dent 1987;15:222-224