

Intra-and Inter-examiner Variability in Evaluating Impression Procedures at the Undergraduate Level

Manal R. Alammari, Yaser M. Alkhiary and Alsayed A. Nawar

Department of Oral and Maxillofacial Rehabilitation, Division of Removable Prosthodontics, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

KEYWORDS Intra-examiner. Inter-examiner. Evaluation. Variability. Impression Making

ABSTRACT Many authors have reported attempts to develop reliable clinical evaluation systems. On reviewing the literature, there is lack of data concerning evaluation and grading of clinical procedures for construction of removable prostheses. This study was conducted to compare intra-and inter-examiners variability in two evaluation methods: glance and grade (global), and check list criteria (analytical) for evaluating final impression making for completely edentulous patients. Three faculty staff members with more than ten years of clinical and teaching experience evaluated final impression made by thirty dental students in removable prosthodontics clinical sessions. Final maxillary and mandibular impressions were graded blindly by each evaluators (A, B, and C) three times, the values were statistically analyzed using Wilcoxon signed rank test and Friedman test setting value of significance at 5 percent. The study revealed that, among the examiners, the intra-examiner variability was statistically significant in most situations for global method. Moreover, there was a statistically significant difference among the three evaluators for half of the evaluation using the same method of grading. Using the analytical method caused an improvement in variability. Analytical method for evaluating final impression making is highly recommended as it is an effective method which satisfies practicality, equality and time management.

INTRODUCTION

Reliability in clinical evaluation presents serious problems to faculty who must render such judgments, and any lack of evaluation consistency can also be a source of confusion and stress for dental students (Sharaf et al. 2007).

In 1997 a researcher concluded that subjectivity associated with clinical evaluation of students performance remains a source of disappointment for both dental students and clinical instructors (Meyers 1997). If consistent and reliable evaluation is absent in the clinical setting, students become confused regarding the standard of performance expected (Haj and Feil 2006). Moreover, progress toward competency can be delayed. In addition, the ability and motivation to self-evaluation, a skill necessary for lifelong learning may disappear when the student is confronted with contradictory feedback from faculty. Calibration in clinical evaluation

can also be a source of frustration for faculty, especially who strive to teach effectively and grade reliability (Jacks et al. 2008). Studies assessing inter-rater agreement among medical and dental faculty have yielded poor results. Lilley and colleagues as well as Fuller found not only significant disagreement between examiners, but also broad intra-examiner variation when the same rater evaluated the same operative procedure on a second time (Lilley et al. 1968; Fuller 1972).

Sharaf and his colleagues recently have concentrated on the development of making systems centered on specific criteria and check lists as an alternative to the glance and grade method, in order to improve rater performance, but the results have been equivocal (Sharaf et al. 2007). Several researchers found that development of an analytical approach using detailed checklists enhanced examiner reliability (Goepferd and Kerber 1980; Dhuru et al. 1978). However, other investigators reported no difference between glance and grade and checklists of assessment (Vann et al. 1983).

Problems with examiner consistency may lead students to recognize that evaluation methods are somewhat uninformed. This concept can determine the learning process and produce a negative effect on undergraduate confidence and performance. A method of assessment of both objectivity and reliability is essential (Sharaf et

Address for correspondence:

Dr. M.R. Alammari,
Department of Oral and Maxillofacial Rehabilitation,
King Abdulaziz University,
Dental Hospital,
Kingdom of Saudi Arabia
PO Box 132817
Jeddah 21382
Saudi Arabia
Telephone: +966 (2)6403443 Ext: 23273
Fax: +966(2)6403316
E-mail: malammari@kau.edu.sa

al. 2007). Therefore to endorse an effective system of learning and to reduce friction between students and faculty over the issue of grading, an objective, reliable and practical methods needs to be used.

Objectives

Literature review revealed a lack of data concerning evaluation and grading of preclinical and clinical procedures for construction of removable prostheses. This preliminary study was done to compare intra-and inter-examiners variability by two evaluation methods: glance and grade (global), checklist and criteria (analytical) for final impressions of completely edentulous patients. Also, to determine an effective method which satisfy practicality, reliability and time management.

MATERIALS AND METHODS

After statistical consultation, it was found that a sample size of thirty final impressions is sufficient to conduct the study. Thirty final impressions for completely edentulous patients made by thirty dental students in clinical removable prosthodontics sessions evaluated. Three specialists evaluated the impressions separately, and each procedure was given a score on a 1 to 10 scale. The three evaluators were faculty members with both master's degrees and Ph.Ds. in removable prosthodontics and have been practicing and teaching removable prosthodontics for more than ten years.

After an ethical approval granted by Faculty of Dentistry ethical Committee, thirty completely edentulous patients were selected from out-patients clinic, removable prosthodontics division, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia. They were selected under the following criteria: No previous complete denture experience, normal and healthy oral mucosa and normally formed ridge. After the patient's consent had obtained, the following procedures were performed by sixth year dental students"

Preliminary impressions of both edentulous jaws were made using irreversible hydrocolloid (normal set) in perforated stock trays (Jeltrate, Dentsply Ltd, UK)

On study casts, custom trays (Kemdent, UK) with spacer were fabricated. Custom trays were

evaluated intra-orally and over-extended borders were reduced.

Border molding was done using low fusing compound (KEMCO Green Tracing Sticks, Kemdent Works, UK) and excess material was removed.

Definitive impressions for both upper and lower arches were made using polyvinyl silicoxane impression material (Aquasil Ultra, Dentsply/Caulk, USA)

The impressions were collected after each session and were given a number code.

Upper and lower impressions for each case were graded blindly by each of the three investigators (A, B, and C) three times.

For the first evaluation, each author graded impressions with the ten-point scale using glance and grade variability.

After completion of the first two evaluations, certain criteria were established (Fig.1) for evaluating and grading of maxillary and mandibular impressions.

Using the criteria and checklist, the third evaluation was performed blindly and graded using the same ten-point scale.

Statistical Analysis

The values were tabulated, and statistical analysis was performed using SPSS package version 16.0 to test the intra-and inter-examiner variability among the three examiners. Statistical analysis was done using Wilcoxon signed rank test and Friedman test setting value of significance at 5 percent. Z value was the calculated statistic that was compared with the tabulated Z_x , and the P value was used to indicate statistical significance.

RESULTS

Table 1 revealed that for almost half of the measurement there is a significant difference among the evaluators except for mandibular impression for evaluator A ($p= 0.897$), Maxillary impression for evaluator B ($P=0.984$) and mandibular impression for evaluator C ($p= 0.246$).

Table 2 revealed a statistically significant differences between maxillary and mandibular impression for the two attempts of grading using the global method for all the evaluators except for evaluator B in the second Glance and

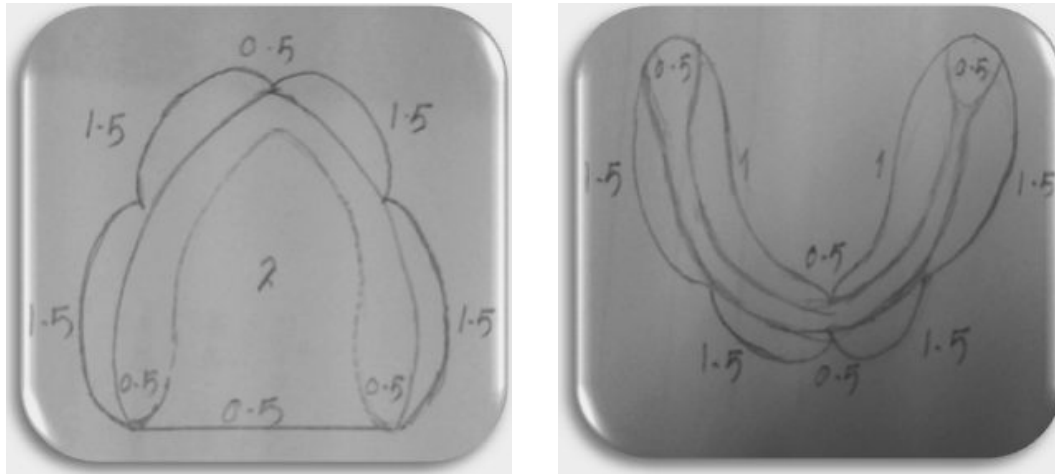


Fig. 1. Illustrative presentation of analytical method with grade of each part and landmark

Check list criteria:

- | | | |
|------------------------------------|------------|---|
| ✓ Presence of surface details | ✓ Rounded | ✓ Even thickness of impression material |
| ✓ Impression is supported by tray. | ✓ No voids | ✓ No pressure area(s). |

Table 1: Intra-examiner variability using glance and grade method of evaluation and grading of scale from 1 to 10

Examiner	Arch	Global 1 st vs. 2 nd reading	
		Z	P
A	Maxillary	2.13	0.033*
	Mandibular	0.13	0.897
B	Maxillary	0.02	0.984
	Mandibular	1.89	0.042*
C	Maxillary	2.29	0.022*
	Mandibular	1.16	0.246

Z = Wilcoxon Signed Ranks Test

* P < 0.05 (Significant).

Grade attempt (P=0.227) as well as when the average of the two reading were calculated for evaluator B (P=0.082).

Intra-examiner variability for maxillary versus mandibular impression evaluation using the analytical method is shown in Table 3. There was a significant difference between the evaluations grades for the maxillary from the mandibular, higher grades were given to the maxillary arch impression from all the evaluators. While inter examiner variability was tested using the Friedman test (Stricker 2008) followed

Table 2: Intra-examiner variability (Maxillary versus mandibular) for Glance and grades method of evaluation and grading in scale of 1 to 10

Method	Examiner	Side	Percentiles			Maxillary vs Mandibular	
			25 th	50 th	75 th	Z	P
Global (1 st Reading)	A	Maxillary	6.9	7.5	8.0	2.39	0.017*
		Mandibular	6.4	7.0	7.5		
	B	Maxillary	7.9	8.0	8.5	2.07	0.039*
		Mandibular	7.0	8.0	8.5		
	C	Maxillary	7.4	8.0	8.0	2.20	0.028*
		Mandibular	6.5	7.5	8.0		
Global (2 nd Reading)	A	Maxillary	7.0	7.5	8.5	2.82	0.005*
		Mandibular	6.0	7.0	8.0		
	B	Maxillary	7.0	8.5	9.0	1.21	0.227
		Mandibular	7.0	8.0	8.5		
	C	Maxillary	7.5	8.3	8.5	2.85	0.004*
		Mandibular	7.0	7.5	8.0		
Average of Two Readings of Global Method	A	Maxillary	7.0	7.5	8.3	2.86	0.004*
		Mandibular	6.2	7.3	7.6		
	B	Maxillary	7.5	8.3	8.8	1.74	0.082
		Mandibular	7.2	7.9	8.3		
	C	Maxillary	7.8	8.0	8.3	2.94	0.003*
		Mandibular	6.8	7.5	8.0		

Table 3: Intra-examiner variability (Maxillary versus mandibular) for analytical method

Method	Examiner	Side	Percentiles			Maxillary vs Mandibular	
			25 th	50 th	75 th	Z	P
Analytical	A	Maxillary	7.0	7.5	8.5	2.95	0.003*
		Mandibular	4.5	6.0	8.0		
	B	Maxillary	6.0	7.0	7.6	3.46	0.001*
		Mandibular	4.0	5.8	7.0		
	C	Maxillary	6.4	7.0	7.5	3.33	0.001*
		Mandibular	5.0	6.0	7.0		

Z = Wilcoxon Signed Ranks Test

*P < 0.05 (Significant)

by (Conover 1999). Then, post-hoc test for pair wise was used for comparisons between the means of the ranks. All tests were two-sided and the 0.05 level was used to indicate statistical significance.

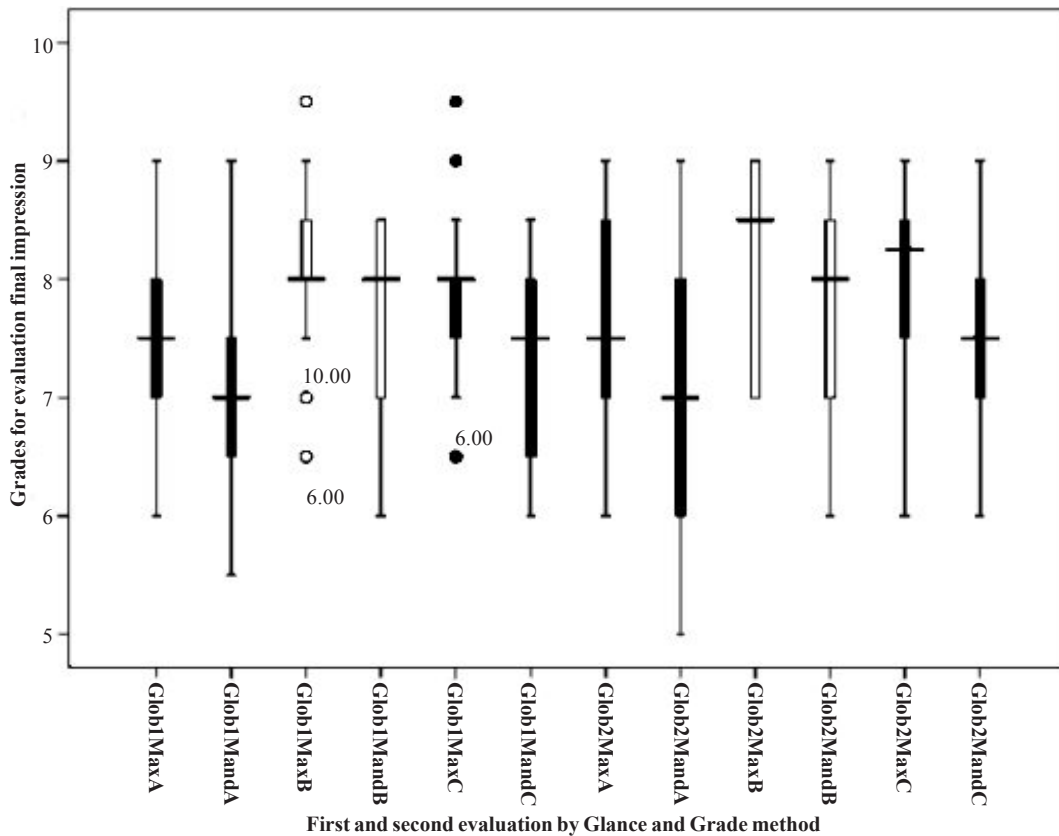
Inter-examiner variability has been shown in Table 4. For the global method, examiners were inconsistent in their measurements for both maxillary and mandibular arches (Friedman $\chi^2 = 12.58$ and 19.02 respectively, $P < 0.01$).

Table 4: Inter-examiner variability comparing glance and grade method to analytical method

	Maxillary		Mandibular	
	χ^2	P	χ^2	P
Global	12.58	0.002*	19.02	0.000*
Analytical	5.02	0.160	4.02	0.134

 χ^2 = Friedman Test * P < 0.05 (Significant)

Using the analytical method, examiners were consistent on the maxillary and on the mandibu-

**Fig. 2. Comparison between the evaluation by glance and grade (Global) method of for each evaluators in the first and second attempt**

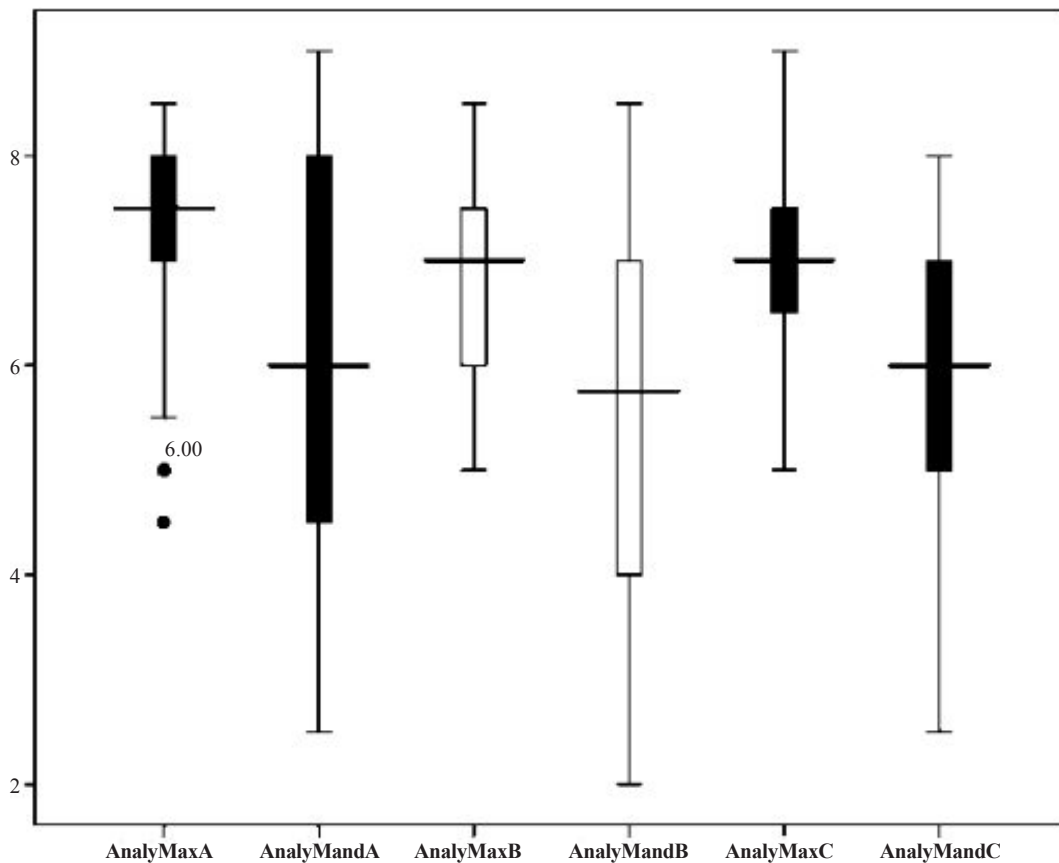


Fig. 3. Difference between the evaluation of maxillary and mandibular impression using the analytical method for grading

lar impression's evaluation and grading from 1 to 10.

Figure 2 illustrates that there was always higher reading on the maxillary than the mandibular impression and these differences between the two arches were more prominent with the analytical method Figure 3. The later method showed lower median and higher inter-quartile range than the global one.

DISCUSSION

The results from this study support the notion of inconsistency among evaluators and/or examiners in evaluating the clinical performance in impression making for undergraduate dental students. There was a significant inter-examiner variability by using the most common method of evaluation which is evaluation

by glance and grade (Global), this was support the finding of several researchers ((Lilley et al. 1968; Fuller 1972; Salvendy et al. 1973; and Jenkins et al. 1998).

In an effort to reduce variability among examiners, Geopferd and Kerber (1980) used analytical system for evaluation using specific criteria. They reported that the technique was superior to the glance and grade method in reducing the variability among examiners. The researchers' results agreed with their finding as seen in Table 4. It was found that global reading higher than analytical readings for all examiners and on both maxillary and mandibular impressions with a lot of variation and outliers as shown in Figure 1. Moreover, our results did not agree with the work of Vann et al. (1983) who reported that no method resulted in more reliability between examiners.

In many teaching dental schools and due to practice circumstances, the glance and grade method is still applied. It is important to develop a practice reproducible, easily applicable and fair method to accurately measure the impression making procedure as this step has a major impact on the following stages of denture making. It is anticipated that by using analytical method, it might help the student to learn what are the landmarks that should be recorded and what they have to look for before submitting their impression and call for supervisor's approval. The authors thought that this diagram is easier to use and to remember than a written checklist. However, the analytical method needs staff training and as suggested by Jenkins et al. (1998). It is found in our results that the maxillary impressions always have a higher grades in both methods of evaluation in contrast to the mandibular one, it may be due to the ease of recording the anatomical land mark in the maxillary arch than those in the mandibular.

CONCLUSION

Among the three examiners, the level of intra-examiner variability was statistically significant for most the evaluation of final impression making clinical procedure for Glance and grades method of evaluation and grading in scale of 1 to 10. Moreover, there was a statistically significant difference among the three evaluators for half of the evaluation using the same method of grading. Using the analytical method caused improvement in variability. These finding indicate that the problem in inconsistency of evaluating undergraduate dental students for final impression making for completely edentulous patients in removable prosthodontics exists.

RECOMMENDATIONS

Finding another method like the analytical explained in this paper is highly needed as it showed that it is an effective methods that satisfy the practicality, and equality. Moreover, The examiners consistency is crucial in teaching and learning process of the students. Therefore, new

evaluation methods and techniques of standardised assessment need to be implemented to promote efficient system of learning and evaluation. Finally, using the criteria established may be useful to encourage a greater understanding of the anatomical landmarks and impression recording skills.

ACKNOWLEDGEMENTS

We wish to acknowledge King Abdulaziz University for sponsoring the study.

REFERENCES

- Conover WJ 1999. *Practical Non-parametric Statistics*. New York: John Wiley & Sons, Inc.
- Dhuru VB, Rypel TS, Johnston WM 1978. Criterion-oriented grading system for preclinical operative dentistry laboratory course. *J Dent Educ*, 42(9): 528-531.
- Fuller JL 1972. The effects of training and criterion models on inter-judge reliability. *J Dent Educ*, 36(4): 19-22.
- Goepferd SJ, Kerber PE 1980. A comparison of two methods for evaluating primary Class II cavity preparations. *J Dent Educ*, 44(9): 537-542.
- Haj-Ali R, Feil P 2006. Rater reliability: Short-and long-term effects of calibration training. *J Dent Educ*, 70(4): 428-433.
- Jacks ME, Blue Ch, Murphy D 2008. Short-and long-term effects of training on dental hygiene faculty members, capacity to write SOAP notes. *J Dent Educ*, 72(6): 719-724.
- Jenkins SM, Drummer PM, Gilmore AS, Edmunds DH, Hicks R, Ash P 1998. Evaluating undergraduate preclinical operative skill: Use of glance and grade marking system. *J Dent*, 26(8): 679-684.
- Lilley JD, Bruggen CHJ, Holloway PJ, Holt JK, Start KB 1968. Reliability of practical test in operative dentistry. *Br Dent J*, 125(5): 194-197.
- Meyers B 1997. Beliefs of dental faculty and students about effective teaching behaviors. *J Dent Educ*, 41(2): 68-76.
- Salvendy G, Hinton WM, Ferguson GW, Cunningham PR 1973. Pilot study on criteria in cavity preparation. *J Dent Educ*, 37(10): 27-31.
- Sharaf AA, Abdel Aziz AM, El-Meligy OA 2007. Intra-and inter-examiner variability in evaluating preclinical pediatric dentistry operative procedures. *J Dent Educ*, 71(4): 540-544.
- Stricker D 2008. Bright Stat com: Free statistics online. *Computer Methods and Programs in Biomedicine*, 92: 135-143.
- Vann WF, Machen JB, Hounshell PB 1983. Effect of criteria and checklists on reliability in preclinical evaluation. *J Dent Educ*, 47(10): 671-675.