

Comparison between Unstructured oral examination "UOE" and Objective Structured oral examination "OSOE" in department of fixed prosthodontics/dental faculty university of Tripoli

DR Hussein S. GHIRRI<sup>1</sup> DMD, MSc. C.A.S.P.

DR Amina Rajab Elsalhin BDS, MDentSc, PhD

Dr Milad M Elshah BDS, MSc, DDS

Fixed prosthodontics department/Faculty of Dentistry/Tripoli University

Key word, oral exam, fixed prosthodontics, structured oral exam,

**Abstract.**

Aim: the aim of this study was to compare reliability of two different oral exam examinations commonly used alternatively in Dental school of Tripoli university (Objective Structured Oral Exam, and Unstructured Practical Oral exam.).Methodology: The method that has been adapted in this study were clinical and self-structured questioner, and statistically a descriptive and inferential statistical analyses was used, the relative variation, Pearson's correlation test, and "ICC" i.e. Interclass Correlation Coefficient respectively...i.e. quantitative, descriptive correlation study, Result: the inferential statistical analyses yielded a "coefficient of variation" value for  $R^{1U}$  and  $R^{2U}$  and for  $R^{1S}$  and  $R^{2S}$  as (28.455, 34.930) and (10.870, 16.028) respectively. Cronbach's alpha reliability was found (0.455) for  $R^{1U}$  and  $R^{2U}$  raters and (0.951) and was (0.463)  $R^{1U}$  and  $R^{2U}$ , and for  $R^{1S}$  and  $R^{2S}$  was significant (P0.001) with value of (0.951 which rated excellent). Bivariate correlation was significant and with value of (0.906) test for Structured oral exam, and was not significant with value of (0.054) for unstructured oral.

Conclusion: We concluded that the Structured oral exams is more reliable than unstructured oral exam.

**Introduction**

"Assessment is an integral part of the educational process at any level and in any discipline. It is a process during which, consideration is given to the amount, level, worth, value or quality of outcomes or products of the learning process".<sup>[1]</sup> Assessments can take a

multitude of formats, and can be classified in many ways; broadly speaking educational assessments are usually classified as summative or formative. Summative assessments are designed to evaluate knowledge and provide formal recognition.<sup>[2]</sup> They are usually used at the

end of a course, or unit, and often used to determine student progression. Formative assessments are used as more of a can be reflected upon in order to make any required improvements. Formative assessments are usually not used for formal recognition, but to aid the learning process. [3] The challenge for effective assessment is to manage the grading process. To do this “faculty must abandon three common false hopes that belie the context and the complexity of the grading process: 1) The false hope of total objectivity in grading; 2) The false hope of total agreement about grading; and 3) The false hope of a one-dimensional student motivation for learning”. [4] The most common examination tools applicable in dental schools, either all, or partially, depends wither the exam is formative, or summative. These are short assay, multiple choices, oral, and practical. Usually oral exams are attainable either by direct questioning dialog between the examiner and the student, either singular, and the assessment based on student knowledge, or in groups of two or more, and the grading assigned according to comparison between candidate performance. Reliability of oral exam in part depend on the observed agreements between two or more sets of grades assigned by

diagnostic tool to provide feedback about the students' progression, which independent examiners, and on the part on the fact that there are Possibility of abuse of personal contact, and There may not be enough adequately trained examiners. However, validity of any test is related to the content and construct of a test, while reliability is related to the score. Reliability is an absolutely essential quality of tests, which means consistency in scores regardless of when and how many times a particular test is taken. The more similar the scores would have been, the more reliable the test is said to be. [5] There are two components of reliability: the performance of candidates from occasion to occasion, and the reliability of the scoring. In the same way, to make an oral test reliable, testers should try out to achieve consistent performances from candidates and to achieve scoring reliability. Difference among raters or examiners in grading is expected due to its multifactorial nature, which makes it something crucial, and should be checked every now and then. The question what is the tolerance difference margin that can be accepted between the means of two or more examiners testing a student otherwise the reliability of the test in doubt. There is no such thing named

tolerance margin or certain factor to adjust the final grad that should be

assigned to the student

### The problem.

The senior staff members of fixed prosthetic department of dental school/ Tripoli university noticed that; there are very wide ranges between the scores assigned by oral examiners set station (very evident variation in the agreement

and correlation inter-raters) in commonly used oral examination test (Unstructured Oral Test) in comparison to that of (Structured oral examinations (OSOE), which have better validity and reliability.<sup>[6]</sup>

### The study's question:

Dose our commonly and more frequently used oral exam (unstructured) in

prosthodontics Department is reliable in comparison to (Structured) test exam?

### The Aim:

The aim of this study was to compare two different oral exam tests commonly used alternatively in Dental school of Tripoli university (Objective

Structured Oral Exam, and Unstructured Oral Exam.) reliability as method, by standardizing (fixation) the examiners.

### Methods and materials

This was a quantitative, descriptive correlation study, to evaluate the reliability of two different oral exams (Objective Structured Oral Exam "OSOE", and Unstructured Oral Exam "UOE") consisted of two blindly selected examiners from the list of examiners were assigned by the prosthodontics department to conduct the spring semester 8 summative exam held at 23/06/2017. The same examiners were selected deliberately to conduct the competition exam to win a job in dental school/university of Tripoli as an assistant

research, at 25/07/2017, the second part (Objective Structured Oral Exam) of the study, to hold any biases from using different raters as this exam was designed to be implemented as. The examiners were instructed to keep their scores of both exams separately, and were handed to the primary authors. A "systematic random sampling was implemented in selecting students' scores used in this study. The procedure was executed by: picking up the first student in main result list as the maiden subject and thereafter pick every

third student till the last subject in the list. The unstructured list result exam consisted of 30 students score from a list of 90 students attended the Unstructured oral exam, and a list of 29 students score were Rater" and number "1 & 2" to differentiate between the two raters, and we assigned for (Unstructured exam) the Alphabetic letters "U"&"S" to denote for (Structured exam) that indicates the method of oral

### 1. Unstructured oral test:

The method of the exam that the students went throw in unstructured exam was as usually the department executes it. Each student was summoned singularly to oral exam station, which formed of two examiners. Each examiner commonly has to ask the student one question he composed on the spot from what the

### 2. Structured oral test:

The method of exam structured exam that the students went throw as Structured oral exam was differing from that of unstructured exam, as the questions of this exam and accompanied answers were constructed, prepared and organized as it follows:

A six questions as well as their accompanied answers were elaborately structured to cover certain topics has the same weight, and each of the two questions paired randomly were printed on a separate papers slip, as well as its answers on paper

selected from a list of 89 students attended structured exam. To distinguish between the two oral exam methods and raters. The two raters involved in the study were denoted aliphatic with litter "R" indicates " exam involved. Therefor the exam unstructured event labeled as ( $R^{1U}$  &  $R^{2U}$ ) and structured oral exam labeled as ( $R^{1S}$  &  $R^{2S}$ ).

student was taught, and the student has to answer it verbally, then each examiner has to rate the student independently, after the student dismissed, the final score will be made after negotiation and averaging of both two scores. The students were rated on scale of 20 marks according to the regulation 10 for each question.

sheet bears student identification space to be filled by the examiners during the examination was on. All students enrolled in the competition were hold in closed room, and summoned not in pairs as usually occurs, but in singular manner to oral exam station, that composed of the raters conducted the unstructured oral test exam. Student was asked to withdrew blindly one of the three wrapped question slips from well shacked pack container (each student has the right to withdraw only two slips), open it, spread it, and read

his question number, then he can proceed reading his question and answering it whenever he is ready. The examiners at mean time were withdrawn the compatible sheet answer and wrote the student name on the answering sheet and independently tack on each right answers, the student verbally replayed on the coordinating answer. When the student quiet answering,

### Data analysis.

After all data was collected, entered to computer by means of Windows Excel Office 2017, cleaned, organized, and summarized. As we have two different data sets measure values, what needed in situations like these is, to measure the relative variation (percentage), rather than measuring the absolute variation. Such a measure is found in the coefficient of variation, which expresses the standard deviation as a percentage of the mean to verify dispersion value. Henceforth, all data was transferred to IBM SPSS version 24 software, where then was transferred to "Z" score values, to standardize the two different exam scale were used in assessment of the students (20, and 30). In order to evaluated Reliability, we need to compare two properties the first is the correlation or the constancy, and the second is test the inter-rater's agreement in each linearly paired of assigned scores in

the student asked to withdraw his second question slip, and so forth with other candidates. After each student completed his exam, and dismissed, each examiner collected all the scores each student obtained out of 30 marks, and wrote it on the student's answering sheet. Then at the end of the exam the final score of each student was averaged.

each test separately, thus the following test has been implemented:

1. Pearson's correlation test was used to establish the correlation inter-raters regarding that the Statistical significance was set at  $P < 0.05$ . an r value of a higher than 0.7 was regarded as strong correlation, and between 0.5 and 0.7 as moderate correlation and below 0.5 is week correlation
2. "ICC" i.e. Interclass Correlation Coefficient, conducted to evaluate the inter-rater's agreement, at significance set "P" value of 0.05, and we considered according to Fleis 1986 the agreement score of higher than 0.08 as excellent and less than 0.08 as good and less than 0.05 poor.

## Results and discussion.

SPSS software version 24 generated the summarized results appeared on the table (1) from the following predetermined test; Intra-class Correlation Coefficient 'ICC', and Bivariate Correlation, and by excel 2017 we calculated the coefficient of variation

1. By carefully reviewing the section " coefficient of variation " values in table (1), you can easily figure out obvious high

disparity in the coefficient of variation score (relative variance) between unstructured oral exam and Structured oral exam, as much as (28.455, 34.930) and (10.870, 16.028) respectively, this high magnitude of differences indicates a wide dispersion of scores around the mean in the Unstructured oral exam.

<b>Table (1) Summary of Interclass correlation coefficient and bivariate correlation results for two raters' implemented two different oral exam tests.</b>						
			Unstructured oral exam		structured oral exam	
			Z*R1u	ZR2u	ZR1s	ZR2s
<b>The Coefficient of Variation</b>			28.455	34.930	10.870	16.028
<b>Cronbach's Alpha</b>	value		.455 low		.951 high	
	rating		unacceptable		highly acceptable	
<b>Intra-class Correlation Coefficient 'ICC'</b>	Average Measures		.463c poor		.953c excellent	
	Significance		.054		.000	
<b>Bivariate Correlation</b>	ZR1	Pearson Correlation	1	.294	1	.906* * high
		Sig. (2-tailed)		.115		.000
		N	30		29	
	ZR2	Pearson Correlation	.294	1	.906* *	1
		Sig. (2-tailed)	.115		.000	

	N	30	29
<b>Two-way mixed effects model where people effects are random and measures effects are fixed.</b>			
<b>a. The estimator is the same, whether the interaction effect is present or not.</b>			
<b>b. Type A Intra-class correlation coefficients using an absolute agreement definition.</b>			
<b>c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.</b>			
<b>Z* all values are in Z scores.</b>			
<b>** Correlation is significant at the 0.01 level (2-tailed).</b>			

2. by reviewing the table (2) which showing the SPSS output result of Inter Correlation Coefficient test "ICC" between rater  $R^{1U}$  and rater  $R^{2U}$  scores. The Reliability Cronbach's alpha test, scored (0.455), i.e. only 45% of variances were consistent, and since it was less than 0.7 the data test considered has very low
3. reliability. Moving to ICC table, with expectation to see small agreement score under the "average measure" column and eventually was (0.463) and not significant, which classified as poor level of agreement.

Table (2) showing Reliability Statistics result between  $R^{1U}$  &  $R^{2U}$  raters into unstructured oral exam.

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.455	.455	2

4.

Table (3) showing the result of Intraclass Correlation Coefficient between  $ZR^{1U}$  &  $ZR^{2U}$  raters into unstructured oral exam.

	Intraclass Correlation <sup>b</sup>	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df 1	df 2	Sig

Single Measures	.301 <sup>a</sup>	-.071-	.596	1.83 3	29	29	.05 4
Average Measures	.463 <sup>c</sup>	-.152-	.747	1.83 3	29	29	.05 4
Two-way mixed effects model where people effects are random and measures effects are fixed.							
a. The estimator is the same, whether the interaction effect is present or not.							
b. Type A Intraclass correlation coefficients using an absolute agreement definition.							
c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.							
D, "Z" alphabetic indicates, that data was transformed to "Z" value.							

Table (3) showing Bivariate Correlation (Pearson Correlation) result test values magnitude, between  $R^{1u}$  and  $R^{2u}$ , which revealed a value of (0.294) at (P value of  $0.115 < 0.05$ ) which classified as poor, and not significant.

<b>Table (3) showing Bivariate Correlation result between two raters sores into Unstructured oral exam</b>			
		R1u	R2u
R1u	Pearson Correlation	1	.294
	Sig. (2-tailed)		.115
	N	30	30
R2u	Pearson Correlation	.294	1
	Sig. (2-tailed)	.115	
	N	30	30

3. by reviewing the table (2) which showing the SPSS output result of Inter Correlation Coefficient test "ICC" between rater R<sup>1S</sup> and rater R<sup>2S</sup> scores. The Reliability Cronbach's alpha test, scored ((0.951)), i.e. 95% of variances were consistent, and since it was more than 0.7 the data test considered has very high reliability. Moving to ICC table, with expectation to see high agreement score under the "average measure" column and eventually was (0.953) and significant, which classified as excellent level of agreement.

Table (4) Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.951	.951	2

4.

Table (5) Intraclass Correlation Coefficient							
	Intraclass Correlation <sup>b</sup>	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df 1	df 2	Sig
Single Measures	.909 <sup>a</sup>	.816	.956	20.377	28	28	.000
Average Measures	.953 <sup>c</sup>	.899	.978	20.377	28	28	.000
Two-way mixed effects model where people effects are random and measures effects are fixed.							
a. The estimator is the same, whether the interaction effect is present or not.							

b. Type A Intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Table (6) showing Bivariate Correlation (Pearson Correlation) result test values magnitude, between  $R^{1S}$  and  $R^{2S}$ , which revealed a value of (0.910) at (P value of  $0.001 > 0.05$ ) which is significant and classified as very strong correlation.

<b>Table (6)Correlations</b>			
		R1s	R2s
R1s	Pearson Correlation	1	.910**
	Sig. (2-tailed)		.000
	N	29	29
R2s	Pearson Correlation	.910**	1
	Sig. (2-tailed)	.000	
	N	29	29
**. Correlation is significant at the 0.01 level (2-tailed).			

### Conclusion:

The Structured oral exams was found more reliable than unstructured oral exam. Our commonly used unstructured oral exam should have avoided, or if not possible, some of experienced staff members only should conduct it i.e. the answer to the

study question is: our commonly and more frequently used oral examination test in prosthodontic Department is not reliable and worth revising or replaced with the alternative Structured oral exams.

### References.

1. Topping, K. (1998). Peer assessment between students in Colleges and Universities. *Review of Educational esearch*, 68, 249-276.

2. Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education* ,21 ,(3) 239-250. <http://dx.doi.org/10.1080/0260293960210304>
  3. Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self - assessment . *Assessment and Evaluation in Higher Education* , 25 ,(1) <http://dx.doi.org/10.1080/02602930050025006>
  4. Cross, K. P. "Classroom Research: Helping Professors Learn More about Teaching and Learning." In P. Seldin (ed.), *How Administrators Can Improve Teaching: Moving from Talk to Action in Higher Education*. San Francisco: Jossey- Bass, 1990.
- Hughes, Arthur. (2000). *Testing for Language Teachers*. Beijing: Foreign Language Teaching and Researching Press.
6. Pearce G, Lee G. Viva voce (oral examination) as an assessment method insights from marketing students. *J Mark Educ* 2009; 31:120-30.