



COMPARATIVE STUDY BETWEEN FACULTY AND STUDENT ASSESSMENTS OF IMPRESSIONS MADE BY UNDERGRADUATES

Amal A. Swelem * and Mohamed H. Abdelnabi **

ABSTRACT

Objective: Competency-based curricula aim to equip students with sufficient skills to enable them to deliver dental care safely. Assessment represents an essential component of a successful competency-based educational experience. This study investigated faculty and self-assessment of impressions, the cornerstones for any prosthodontic treatment, made by dental students in their first clinical prosthodontic course.

Methods: 237 “first-attempt” preliminary impressions (115 maxillary; 122 mandibular) were assessed by 2 calibrated professional faculty and self-assessed by the students themselves. Assessment was based on specific clearly-defined criteria and a 3-level rating rubric: 0(non-acceptable); 1(acceptable); 2(ideal). Faculty assessed the impressions independently then together to agree on its acceptability. Impressions were counted acceptable when at-least 1 was scored for all criteria and students were considered competent in this clinical skill. Comparison between faculty scores and between faculty and self-assessment scores were statistically analyzed ($P < 0.05$).

Results: Based on faculty assessment, out of the 237 impressions; 2.5% (4 maxillary; 2 mandibular) were ideal, 62.45% (80 maxillary; 68 mandibular) were acceptable and 35% (31 maxillary; 52 mandibular) were unacceptable. Tray selection and border thickness were the most identified defective criteria. Inter-assessor agreement exceeded 95% in most criteria. There were statistically significant differences between faculty and student self-assessment scores with the latter giving higher scores.

Conclusions: Students, in their first clinical course, usually require more than one attempt to make acceptable impressions. Competency was higher for maxillary impressions. Clearly-defined criteria and rubric-based assessment resulted in satisfactory agreement levels between assessors. Undergraduate students, at this educational level, tend to over-rate their work.

* Associate Professor, Oral and Maxillofacial Prosthodontic Department, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia; Removable Prosthodontic Department, Faculty of Oral and Dental Medicine, Cairo University, Egypt.

** Associate Professor, Oral and Maxillofacial Prosthodontic Department, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia; Removable Prosthodontic Department, Faculty of Dentistry, Minia University, Egypt.

INTRODUCTION

Over the past years there has been an on-going shift towards competency-based curricula in health professions education including under- and post-graduate dental clinical programs.¹ One of the objectives of competency-based undergraduate curriculum is to equip students with necessary skills to qualify them to deliver dental care to their patients safely.²

Assessment represents a critical and essential component of a successful educational experience. It ensures that the students are acquiring the required skills,³ as well as, it determines the students' capacity to implement and integrate the various learning domains that collectively define competent practice.⁴ However, assessment is not a unitary concept as regards methodology. Kramer et al³ listed and described different strategies for assessing dental students' attainment of competence including written assessments (MCQs, short answer and structured essays), faculty assessment by observation (global ratings, structured observation, standardized oral exam), multisource assessment (Self or peer assessment, patient survey, standardized patients), stimulation (virtual reality and models), multi-competency comprehensive assessment (OSCE and Triple Jump Exercise ;TJE) and work samples (portfolios and record reviews).

Faculty assessment by structured observation is based on specific criteria and rating scales.³ The students' performance is assessed by one or more raters based on a checklist with agreed upon entries that address the critical aspects of the performance. Students are rated on a scale consisting of a range of points where the lowest point indicates that the competency has not been successfully achieved while the highest indicates that most or all aspects of the competency have been achieved successfully. This type of assessment is used extensively for assessing performance of clinical procedures and psychomotor skills.³

Self-assessment is a significant component

of formative evaluation that encourages students to assess their own performance as part of an ongoing process of reflection on their learning progression and self-regulation.⁵ It is often used as interchangeable descriptors of the students' ability to identify their inherent shortcomings and strengths.^{3,6} In dentistry, which is considered one of the self-regulated professions, self-assessment and self-directed learning concepts are integral to undergraduate education, ongoing professional development and the maintenance of competence.⁶ It has been recognised that this approach is most effective, and its benefits maximized, when used in comparison to or combination with assessment by experienced, competent evaluators.³

The present study was dedicated to faculty and self-assessment of preliminary impressions for completely edentulous arches made by undergraduate students in their first clinical prosthodontic course. Impression recording was the chosen clinical skill since impressions represent the cornerstones for any prosthodontic treatment, the accuracy of which influences the quality, hence longevity of the final prostheses. The following were investigated: (i) Students' competency achievements in making first-attempt acceptable impressions (ii) levels of agreement between faculty assessments and (iii) levels of agreement between faculty and students' assessments.

MATERIALS AND METHODS

In a series of meetings and discussions, two professional clinical faculty members (authors) agreed upon a number of criteria that could be used to assess a completely edentulous preliminary alginate impression. Both faculty had PhDs in Removable Prosthodontics and have been teaching and practicing Removable Prosthodontics for more than ten years. The final assessment was based on 6 specific clearly defined criteria and a 3-level rating rubric for each criterion as presented in **Table 1**. The rubric assessed students' abilities to make acceptable maxillary and mandibular

TABLE (1) The 6 main criteria for assessing an alginate preliminary impression for completely edentulous arches with the 3-level rating rubric for each criterion

Criteria	0 (Non-acceptable)	1 (Acceptable)	2 (Ideal)
Tray selection (TS)	<ul style="list-style-type: none"> • A partially edentulous tray or completely edentulous non-perforated tray has been selected • Absence of 2 mm clearance space between the tray and the border tissues • Tray borders are very short or very long 	<ul style="list-style-type: none"> • A completely edentulous perforated tray has been selected • Tray is covering all or most of the denture bearing areas • Presence of at least 2mm clearance between the tray and the border tissues • Tray borders are shorter than the vestibule by 2mm in nearly all areas 	<ul style="list-style-type: none"> • A completely edentulous perforated tray has been selected • Tray is covering all denture bearing areas • Presence of at least 2mm clearance between the tray and the border tissues • Tray borders are shorter than the vestibule by 2mm all around
Tray modification (TM)	<ul style="list-style-type: none"> • Tray is not properly modified by wax in any of the deficient areas 	<ul style="list-style-type: none"> • Tray properly modified by wax in most of the deficient areas 	<ul style="list-style-type: none"> • Tray properly modified by wax in all deficient areas
Alginate mix (AM)	<ul style="list-style-type: none"> • Alginate is not properly mixed and of improper consistency (flowy or granular) • Air bubbles in the mix 	<ul style="list-style-type: none"> • Alginate is properly mixed and of proper consistency • Minute air bubbles in the mix 	<ul style="list-style-type: none"> • Alginate is properly mixed and of proper consistency • No air bubbles in the mix
Tray insertion and seating (TI)	<ul style="list-style-type: none"> • Loaded tray is not centralized during insertion • Loaded tray is inserted far more anteriorly/far more posteriorly • Loaded tray is not completely seated 	<ul style="list-style-type: none"> • Loaded tray is centralized during insertion • Loaded tray is inserted slightly more anteriorly /slightly more posteriorly • Loaded tray is completely seated 	<ul style="list-style-type: none"> • Loaded tray is centralized during insertion • Loaded tray is inserted properly anteroposteriorly • Loaded tray is completely
Impression surface and denture bearing area (IS)	<ul style="list-style-type: none"> • Alginate is not sufficiently set before removal • Impression is detached or separated from the tray • Impression surface did not record most of the details of the denture bearing areas • Presence of many small voids/ large void in critical areas of the denture bearing surface • Presence of more than 3 pressure areas void in critical areas of the denture bearing surface 	<ul style="list-style-type: none"> • Alginate is sufficiently set before removal • Impression is not detached or separated from the tray • Impression surface recorded most of the details of the denture bearing areas • Presence of 2-3 small voids in non-critical areas of the denture bearing surface • Presence of 1-2 small pressure areas in non-critical areas of the denture bearing surface 	<ul style="list-style-type: none"> • Alginate is sufficiently set before removal • Impression is not detached or separated from the tray • Impression surface recorded all relevant details of the denture bearing areas • Absence of voids in the denture bearing areas • Absence of pressure areas in the denture bearing areas
Impression borders' extension and thickness (IB)	<ul style="list-style-type: none"> • Impression borders are under-or over-extended all around or in most of the areas • Impression borders are very thick or very thin in all areas or most of the areas • Presence of many small voids/ large void or many pressure areas in the impression borders 	<ul style="list-style-type: none"> • Impression borders are well extended and of proper thickness in most areas • Impression borders in most of the areas are rounded and smooth • Presence of 2-3 small voids or pressure areas in the impression borders 	<ul style="list-style-type: none"> • Impression borders are well extended and of proper thickness all around • Impression borders all around are rounded and smooth. • Absence of voids and pressure areas in the impression borders

preliminary impressions for a completely edentulous case. The rubric was designed to encourage objective assessment amongst faculty and students. Based on this structured objective assessment, impressions made by undergraduate students who were in their first clinical Prosthodontic course were assessed by the two faculty members and self-assessed by the students themselves. At this stage, formal instructions of all clinical aspects of impression making had been delivered to the students through a 2 hour lecture. This was followed by video demonstrations then 3-hour group-based (approximately 8 students/group) clinical demonstrations on the recording of maxillary and mandibular impressions for completely edentulous patients. A detailed illustrative step-by-step PDF document with explanatory photos and text had also been distributed. Students were initially instructed to choose an appropriately sized stock tray and record maxillary and mandibular alginate impressions of each other, one acting as the patient and one as the clinician. This was carried out in 2 clinical sessions (3 hours each) under close faculty supervision, where the students received instant formative feedback or some assistance when necessary. They were then instructed to independently make one maxillary and one mandibular alginate impression for their completely edentulous cases, and these impressions were assessed. All cases were chosen with fairly well-developed ridges to try to standardize case difficulty for the students in their first clinical encounter. To standardize the procedure itself, all students used the same regular-set alginate material and were instructed to manually mix it using the powder: liquid measures provided by the manufacturer.

A total of 237 first-attempt alginate preliminary impressions (115 maxillary and 122 mandibular) were assessed. Once an impression was produced, it was self-assessed by the student then by the 2 assessors independently. The assessors then conferred to decide whether the impression was acceptable or not. Impressions were considered acceptable when at-least 1 was scored for all

criteria (by the two Faculty) and students were considered competent in this clinical skill. Students were instructed to repeat unacceptable impressions however the assessment of the remade impressions was not included in this study.

The number of ideal, acceptable and non-acceptable first-attempt impressions were recorded. All assessment values were recorded, tabulated and statistically analysed using SPSS for Windows version 16. Agreement between the assessors (Faculty-Faculty and Faculty-Student) was studied using percentage agreement and kappa analysis to test the inter-examiner variability as well as faculty-student variability. Wilcoxon signed rank test was used to indicate statistical significance ($P < 0.05$).

RESULTS

Based on overall faculty assessment and agreement, out of the 237 impressions; 2.5% (4 maxillary and 2 mandibular) were ideal, 62.45% (80 maxillary and 68 mandibular) showed some defects but were still acceptable and 35% (31 maxillary and 52 mandibular) were rated unacceptable (**Fig. 1**). Competency was greater with maxillary than mandibular impressions.

Agreement between the two faculty members as well as between the faculty members and students is shown in **Fig. 2**. There was some variation in the amount of agreement between the faculty assessors for the different criteria. However, agreement exceeded 95% in most criteria. 100% inter-assessor agreement was recorded for the alginate mix (AM) criterion while the lowest % agreement (92%) was detected for the impression border (IB) criterion. Kappa statistics revealed very strong (>0.8) to almost perfect (>0.9) agreement for all criteria with statistically insignificant differences between the 2 faculty members (**Table 2**). On the otherhand, % agreement between students' and faculty ratings ranged from 52% to 75% with students giving higher scores. The same trend was observed when the students' ratings were compared to both 1st and 2nd faculty ratings. The lowest % agreements

were recorded for the tray selection (TS) and tray modification (TM) criteria, while the highest % agreements were recorded for the alginate mix (AM) and tray insertion (TI) criteria. Kappa statistics

revealed slight (<0.2) to moderate (<0.6) agreement for all criteria with highly statistically significant differences between the students and the 2 faculty members in all six criteria. (Tables 3 and 4).

TABLE (2) Comparison between faculty assessments

Criterion Code	Frequency of rating by 1 st Faculty Assessor			Frequency of rating by 2 nd Faculty Assessor			% Agreement	Kappa	P-value (WSR)
	0	1	2	0	1	2			
TS	44 (18.2%)	143 (59.1%)	42 (17.4%)	42 (17.4%)	147 (60.7%)	48 (19.8%)	96 %	0.923	1.000
TM	11 (4.5%)	176 (72.7%)	50 (20.7%)	10 (4.1%)	173 (71.5%)	54 (22.3%)	97 %	0.928	0.059
AM	9 (3.7%)	152 (62.8%)	76 (31.4%)	9 (3.7%)	152 (62.8%)	76 (31.4%)	100 %	1.000	1.000
TI	9 (3.7%)	161 (66.5%)	67 (27.7%)	10 (4.1%)	162 (66.9%)	65 (26.9%)	98 %	0.954	0.180
IS	13 (5.4%)	156 (64.5%)	68 (28.1%)	12 (5%)	153 (63.2%)	72 (29.8%)	96 %	0.922	0.096
IB	27 (11.2%)	160 (66.1%)	50 (20.7%)	27 (11.2%)	162 (66.9%)	48 (19.8%)	92 %	0.843	0.637

TS: Tray selection, TM: Tray modification, AM: Alginate mix, TI: Tray insertion and seating, IS: Impression surface and denture bearing area, IB: Impression borders' extension and thickness. WSR: Wilcoxon Signed Ranks Test

TABLE (3) Comparison between 1st Faculty member and students

Criterion Code	Frequency of ratings by Students			Frequency of ratings by 1 st Faculty			% Agreement	Kappa	P-value (WST)
	0	1	2	0	1	2			
TS	13 (5.4%)	124 (51.2%)	100 (41.3%)	44 (18.2%)	143 (59.1%)	42 (17.4%)	52%	0.185	<0.01*
TM	4 (1.7%)	155 (47.5%)	118 (48.8%)	11 (4.5%)	176 (72.7%)	50 (20.7%)	55%	0.154	<0.01*
AM	6 (2.5%)	109 (45%)	122 (50.4%)	9 (3.7%)	152 (62.8%)	76 (31.4%)	74%	0.522	<0.01*
TI	8 (3.3%)	111 (45.9%)	118 (48.8%)	9 (3.7%)	161 (66.5%)	67 (27.7%)	75%	0.531	<0.01*
IS	9 (3.7%)	94 (38.8%)	134 (55.4%)	13 (5.4%)	156 (64.5%)	68 (28.1%)	64%	0.376	<0.01*
IB	12 (5%)	126 (52.1%)	99 (40.9%)	27 (11.2%)	160 (66.1%)	50 (20.7%)	68%	0.422	<0.01*

TS: Tray selection, TM: Tray modification, AM: Alginate mix, TI: Tray insertion and seating, IS: Impression surface and denture bearing area, IB: Impression borders' extension and thickness.

WSR: Wilcoxon Signed Ranks Test

***: Highly significant at P<0.01**

TABLE (4) Comparison between 2nd Faculty member and students

Criterion Code	Frequency of ratings by Students			Frequency of ratings by 2 nd Faculty			% Agreement	Kappa	P-value (WSR)
	0	1	2	0	1	2			
TS	13 (5.4%)	124 (51.2%)	100 (41.3%)	42 (17.4%)	147 (60.7%)	48 (19.8%)	52%	0.164	<0.01*
TM	4 (1.7%)	155 (47.5%)	118 (48.8%)	10 (4.1%)	173 (71.5%)	54 (22.3%)	56%	0.167	<0.01*
AM	6 (2.5%)	109 (45%)	122 (50.4%)	9 (3.7%)	152 (62.8%)	76 (31.4%)	74%	0.522	<0.01*
TI	8 (3.3%)	111 (45.9%)	118 (48.8%)	10 (4.1%)	162 (66.9%)	65 (26.9%)	74%	0.525	<0.01*
IS	9 (3.7%)	94 (38.8%)	134 (55.4%)	12 (5%)	153 (63.2%)	72 (29.8%)	66%	0.408	<0.01*
IB	12 (5%)	126 (52.1%)	99 (40.9%)	27 (11.2%)	162 (66.9%)	48 (19.8%)	69%	0.436	<0.01*

TS: Tray selection, TM: Tray modification, AM: Alginate mix, TI: Tray insertion and seating, IS: Impression surface and denture bearing area, IB: Impression borders' extension and thickness.

WSR: Wilcoxon Signed Ranks Test

*: Highly significant at P<0.01

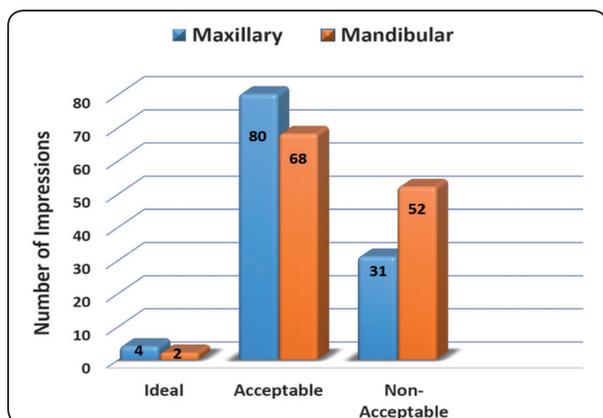


Fig. (1) Acceptability of “First attempt” preliminary impressions made by undergraduate dental students, in their first clinical prosthodontic course, based on overall faculty assessment.

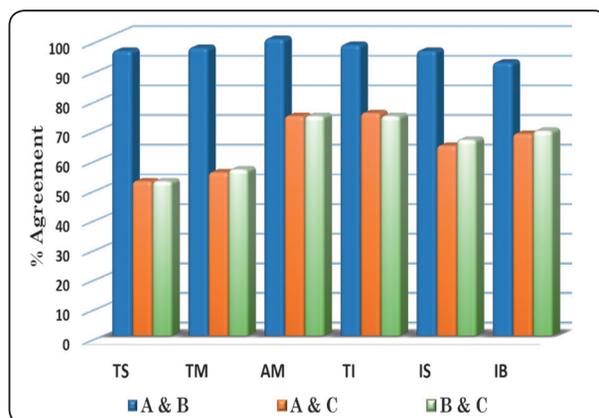


Fig. (2) % Agreement between the three assessor pairs: First and second Faculty (A and B); First Faculty and students (A and C) and second Faculty and students (B and C). TS: Tray selection, TM: Tray modification, AM: Alginate mix, TI: Tray insertion and seating, IS: Impression surface and denture bearing area, IB: Impression borders' extension and thickness.

DISCUSSION

The current study revealed that still 35 % (more than one third) of the students made unacceptable impressions despite the lecturing, formative instructions, explanatory documents, clinical demonstrations and supervised self-practice. This indicates that achieving competence in a “clinical”

skill, especially fundamental and core skills as the one at hand, is challenging especially at this early stage of a clinical course where some students may require more practice and training through several attempts to make acceptable impressions. By subsequently receiving further clinical instruction and supervision, they would be able to develop more clinical skill as they progressed in the course.

It is noteworthy that competency was more achievable with maxillary than mandibular impressions. This may be attributed to the fact that border molding procedures in the mandibular arch depend more on the patient and his tongue movements which are uncontrollable by the operator. Hence, in real life, clinical steps that require some "patient performance" are actually more challenging. In this case, variability in patients' compliance with the instructions and cooperation, as well as, their performance capability play an influential role in the student's competency achievement. This is totally different from pre-clinical steps, where patients are not involved, or clinical steps that are totally operator-dependent. This throws the light on a very entitled question of whether competency in such steps should be assessed differently considering that it is "student and patient dependent".

Standardized assessment is very critical in the educational process as lack of evaluation consistency can be a source of confusion and stress for students affecting thereby their confidence and performance.⁷ The current study revealed satisfactory, in fact, very high levels of agreement between the two faculty members in general. This could be attributed to several factors. First, assessors experience was found to have an influence on agreement scores.⁸ It has been demonstrated that experience significantly improves both intra- and inter-examiner agreement as more experienced examiners showed more consistency in their scores than less-experienced or junior staff.⁹ It could hence be recommended that assessments be carried out only by experienced faculty that have the same level of experience. Unfortunately this does not accurately reflect the situation in most dental schools. Secondly, levels of agreement are dependent upon the clarity of the assessment criteria and their relevance to the skills being assessed.⁵ It is acknowledged that any assessment of performance of a clinical skill cannot be totally objective.¹⁰

However the lack of significant differences between the faculty members may also be attributed to their beforehand close collaboration in deciding on clearly-defined relevant criteria, and the amount of preparation they put in discussing together how judgements could be made on the selected criteria for acceptable performance. This standardization is vital for a more reliable assessment.⁵ Moreover, studies revealed that using such an analytical approach for evaluation with specific criteria and a rubric-based rating significantly reduced variability among examiners.^{11,12} Recently, researches have concentrated on the development of such grading systems as an alternative to the conventional glance and grade method in an attempt to improve rater performance. Results however have been controversial. Some researchers reported that it did not result in any significant improvement in the inter-examiner variability^{7,13} while others found that it substantially reduced variability among examiners.^{10,12} Our results agree with the latter studies. According to Kramer et al³, structured observation assessments with checklists and rating scales are ideal for evaluating clinically relevant competencies, however they are still susceptible to some subjectivity on the part of raters. This statement could therefore explain the difference in inter-examiner agreement among five out of six criteria, which although still high, yet differed and ranged from 92% to 98%. This implies some subjectivity in judgement and could often be explained by different expectations on behalf of the assessors.¹⁰ The lowest % agreement (92%) was recorded for the impression borders' extension and thickness. This finding is in agreement with that of Scott et al 2001,¹⁰ who reported that judging the appropriate extension of an impression is particularly challenging. The authors believe that had more specific details such as "to the full depth of the vestibule" and "2-3mm thick" been added to the extension and thickness rubric respectively, more objectivity of assessing this criterion would have been established. This will

be considered in the future forms. It has also been mentioned that some criteria are normally more easily assessed than others.¹⁰ That was true for the alginate mix criterion in which the inter-examiner agreement reached 100%.

Self-assessment has long been identified as a vital component in developing the self-directed and analytical minds of health professionals. It is believed that this subsequently leads to lifelong learning and ensures satisfactory professional ongoing self-regulation, which are pre-requisites for being an effective clinician.⁶ However, educational research indicates that students, especially in their early undergraduate years, are not accurate assessors of their own performance.^{5,14,15} According to Kramer et al 2009,³ dental students are best capable of self-assessing their performance if provided with clearly defined criteria and a set of standards of acceptable levels of knowledge and performance. This was applied in the current study where the students were asked to self-assess themselves using the same detailed, clearly defined, rubric-based criteria forms used by the faculty members to encourage objective assessment amongst faculty and students. However, highly significant differences still existed between the student and faculty ratings for all six criteria. In fact, % agreement ranged from 52 to 75 % revealing slight (<0.2) to moderate (<0.6) agreement among the different criteria. San Diego et al⁵ explained in an earlier study that at this stage, concepts and skills that students may have acquired in assessing their own-work may not have been enough to enable them to judge a 'correct performance' criterion. According to their belief, self-assessment skills require expertise and practice, and that more and more training for undergraduate students might narrow the difference between theirs and their instructors' assessment ratings.⁵ This highlights once more the influence of experience in competency assessments, especially those involving clinical skills. This explains why it has been recommended to reserve self-assessment for pre-clinical competencies that do not involve

demonstrations of highly precise skills that are best evaluated by experienced clinicians.³

In-depth analysis of the results also revealed two worth mentioning observations. First observation was that the students over-rated their work. This contradicts the results reached by Ericson et al 1997,¹⁶ who reported that the students tended to under-score their performance more frequently than over-score it. However, our results matched and confirmed findings reached by other investigators.^{5,14,15} These studies demonstrated that non-experts, in this case, the students, are likely to be less critical in marking, especially if they are not familiar with the learning objectives and outcomes.^{17,18} It has been suggested that usually beginners tend to overestimate their competence, whereas more mature learners tend to underestimate it.¹⁵ It was obvious that the frequency of a "0 score" given by the students to themselves is remarkably low compared to faculty ratings. This could be attributed to either of two reasons. Either the students are truly incapable of detecting their shortcomings in this early stage or they are just "playing smart" drawn from their fear of losing marks. This seems as an interesting point that entails further investigation. Second observation was that the highest percentage of student-faculty agreement was recorded for the alginate mix and tray insertion criteria. The same trend was observed for inter-examiner % agreement. This again emphasizes the proposition that some criteria are practically easier to assess than others even for students.

One of the limitations of the current study was that it assessed only the quality of the end product (impression) and did not assess all criteria that contribute to the whole clinical procedure including student's attitude and communication skills with the patient throughout the procedure, patient-operator position, and the practice of appropriate infection control measures. Student demographic characteristics including age, gender and social background were also not considered. It is possible

that they may have had an influence on the accuracy of their self-ratings. These two points therefore could be the scope of future researches.

Assessment had to be done immediately after impression making since alginate impressions are to be poured within 30 minutes at-most. Hence both the students and the faculty had to make rapid judgements on whether the criteria were met. This may have possibly accounted for some of the differences in agreement between the assessors. For the same reason, intra-examiner variability was not applicable for this particular step as it was not possible to re-assess or re-evaluate the same impression hours or days later. An alternative approach was to take a photo of the impression which could then be viewed at a later stage. However this would also depend on the quality of the photographs not to mention that it is only a 2-dimensional representation of the actual product.

CONCLUSIONS

Within the limitations of the current study, the following could be concluded:

- Students, in their first clinical Prosthodontic course, usually require more than one attempt to make acceptable alginate impressions for completely edentulous cases. Competency was higher with maxillary than mandibular impressions.
- Clearly-defined criteria and rubric-based assessment resulted in satisfactory agreement levels between faculty assessors verifying assessment consistency.
- Undergraduate students, at this early educational level, tend to overrate their work.

REFERENCES

1. Chambers DW, Glassman P. A primer on competency-based evaluation. *J Dent Ed* 1997; 61: 651–666.
2. Scott BJ, Evans DJ, Drummond JR, Mossey PA, Stirrups DR. An investigation into the use of a structured clinical operative test for the assessment of a clinical skill. *Eur J Dent Educ*. 2001; 5:31-7.
3. Kramer GA, Albino JEN, Andrieu SC et al. Dental student assessment toolbox. *J Dent Educ* 2009; 73: 12-35
4. Albino JEN, Young SK, Neumann LM et al. Assessing dental students' competence: Best practice recommendations in the performance assessment literature and investigation of current practices in predoctoral dental education. *J Dent Educ* 2008; 72: 1405-35.
5. San Diego JP, Newton T, Quinn BFA, Cox MJ, Woolford MJ. Levels of agreement between student and staff assessments of clinical skills in performing cavity preparation in artificial teeth. *Eur J Den Educ*. 2014;18: 58-64.
6. Redwood C, Winning T, Townsend G. The missing link: self-assessment and continuing professional development. *Aust Dent J*. 2010; 55: 15-9.
7. Sharaf AA, Abdelaziz AM, El-Meligy OAS. Intra- and Inter-examiner variability in evaluating preclinical pediatric dentistry operative procedures. *J Den Educ*. 2007; 71: 540-4.
8. Mast TA, Bethart H. Evaluation of clinical dental procedures by senior students. *J Dent Educ* 1978; 42: 196-7
9. Jenkins SM, Dummer PM, Gilmour AS, Edmunds DH, Hicks R, Ash P. Evaluating undergraduate preclinical operative skill; use of a glance and grade marking system. *J Dent*. 1998; 26: 679-84.
10. Scott BJJ, Evans DJP, Drummond JR, Mossey PA, Stirrups DR. An investigation into the use of a structured clinical operative test for the assessment of a clinical skill. *Eur J Dent Educ*. 2001; 5: 31-7.
11. Dhuru VB, Rypel, TS, Johnston WM. Criterion-oriented grading system for preclinical operative dentistry laboratory course. *J Dent Educ* 1978; 42: 92-6.
12. Goepferd SJ, Kerber PE. A comparison of two methods for evaluating primary class II cavity preparations. *J Dent Educ* 1980; 44: 537-42.
13. Vann WF, Machen JB, Hounshell PB. Effects of criteria and checklist on reliability in preclinical evaluation. *J Den Educ*. 1983; 47: 671-5.
14. Stacey MA, Morgan MV, Wright C. The effect of clinical targets on productivity and perceptions of clinical competency. *J Dent Ed* 1998; 62: 409–14.

15. Mattheos N, Nattestad A, Falk-Nilsson E, Attstrom R. The interactive examination: assessing students' self-assessment ability. *Med Educ* 2004; 38: 378–89.
16. Ericson D, Christersson C, Manogue M, Rohlin M. Clinical guidelines and self-assessment in dental education. *Eur J Dent Educ*. 1997; 1:123-8.
17. Boud D, Falchikov N. Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *High Educ* 1989; 18: 529–49.
18. Wetherell J, Mullins G, Hirsch R. Self-assessment in a problem based learning curriculum in dentistry. *Eur J Dent Educ* 1999 Aug; 3: 97–105.