RedFame

Journal of Education and Training Studies
Vol. 10, No. 4; October 2022
ISSN 2324-805X E-ISSN 2324-8068
Published by Redfame Publishing
URL: http://jets.redfame.com

Developing Self-Assessment Skills in Objective Structured Clinical Examinations

Edward Hilton Iv[1], Muath A. Aldosari[2,3], Rahen P. Kakadia[3], Luis Lopez[3], Aram Kim[3], Sang E. Park[3]

[1]US Air Force, Washington, DC, USA

[2]College of Dentistry, King Saud University, Riyadh, Kingdom of Saudi Arabia

[3]Harvard School of Dental Medicine, Boston, MA, USA

Correspondence: Sang E. Park, Harvard School of Dental Medicine, 188 Longwood Ave, Boston, MA 02115, USA.

Received: April 19, 2022 Accepted: June 27, 2022 Online Published: August 1, 2022

doi:10.11114/jets.v10i4.5537 URL: https://doi.org/10.11114/jets.v10i4.5537

Abstract

Introduction: The purpose of the study was to describe the integration of self-assessments into the Objective Structured Clinical Examination (OSCE) as a part of reflective practice in clinical education. An emphasis was placed on the process of student learning to complement the existing comprehensive patient care assessment model.

Methods: Student self-assessment data was captured three times throughout patient care experiences during the clinical education period and measured against corresponding faculty assessments of students at these times in the predoctoral dental program.

Results: Analysis revealed that there were some disciplines in which faculty and student assessments were moderately reliable; however, as students progressed through the clinical program, their overall self-assessments did not have a higher correlation with faculty assessments.

Conclusion: The OSCE can be designed as a useful tool in measuring non-traditional competencies and provide an opportunity for students to self-assess their learning. However, further emphasis on self-assessment skills for students needs to be incorporated as they progress through clinical programs.

Keywords: OSCE, self-assessment skills, predoctoral education, curriculum design, student performance

1. Introduction

With the many advances in health sciences, educational structure and delivery to health professional students have remained relatively slow-paced as in-class lectures and multiple-choice assessments still predominate student learning. (Prober & Heath, 2012; Kramer et al., 2009; Park & Howell, 2015). It is important to engage the students in the learning process to promote critical thinking skills as active learning stimulates higher-order thinking, problem solving and critical thinking skills while providing feedback to both students and teachers (National Research Council, 2000; McKeachie, Pintrich, Lin, & Smith, 1986; Krupat, Richards, Sullivan, Fleenor, & Schwattzstein, 2016; Tain, Schwartzstein, Friedland, & Park, 2017). Development of self-assessment skills is an important component of active learning and encourages students to explore values that are not easily measurable. Students' performance has shown improvements in self-assessments with repeated experiences with no difference between self-assessed and faculty-guided remediation (White, Ross, & Gruppen, 2009; Kim, Chutinan, & Park, 2015).

The current Commission on Dental Accreditation (CODA) standards include student ability to self-assess, including the development of professional competencies and the demonstration of professional values and capacities associated with self-directed lifelong learning. (Commission on Dental Accreditation, 2012). The ability of students to evaluate their own work through purposeful self-assessment is a valuable skill for students to acquire during dental school in order to practice as a competent health care provider (Regehr & Eva, 2006; Eva & Regehr, 2005). Assessments should reflect these characteristics, and evaluation instruments need to be incorporated with specific criteria that can contribute to effective student learning in the dental program. However, a better understanding of accuracy in self-assessment is necessary to offer its effectiveness in measurement and whether the caliber of students have any correlation to tendencies in over or under-rating themselves (Falchikov & Boud, 1989; Brown, Andrade, & Chen, 2015). It is believed that teaching novice learners how to accurately self-assess can contribute to improved performance (Langendyk, 2006).

The objective structured clinical examination (OSCE) is a form of assessment that can be designed to evaluate higher-level cognitive skills such as critical thinking, problem solving, and communication ability contributing to life-long learning skills (Illeris, 2003; Bandura, 1986; Park, Anderson, & Karimbux, 2016; Park, Price, & Karimbux, 2018). It can serve as a formative or summative assessment tool as students transition from preclinical to clinical dentistry in the predoctoral curriculum, contributing to achieving clinical competence and self-assessment ability (Graham, Zubiaurre Bitzer, & Anderson, 2013; Lele, 2011; Schoonheim-Klein et al., 2006; Yu, Pagni, Park, & Karimbux, 2020). Teaching students how to accurately self-assess can contribute to improved performance, and an assessment method that incorporates self-assessments could be designed into the OSCE format to meet these needs (Langendyk, 2006).

The purpose of this study was to evaluate whether student self-assessment on OSCE performance was related to faculty assessment, and whether the progression through the clinical program had any effect on student self-assessment skills in the predoctoral dental program. We sought to determine whether students had the same opinion of their own performance as faculty had, and whether students improved their self-assessment skills over the course of the program.

2. Methods

The retrospective study was approved by the Institutional Review Board at XXX and XXXX (IRB15-3835). Informed consent was waived by the Institutional Review Board at XXX and XXXX as the data did not contain any personal identifiers. The tests were designed and are administered by the Office for Dental Education at XXXX, which also keeps the test results and self-evaluations.

The OSCE was organized as a multi-station event based on a standardized clinical scenario, during which students progress through discipline-specific stations where they must perform tasks or answer questions. Each station had a different examiner based on the area being assessed, as opposed to the traditional method of clinical examinations where a student would be assigned to an examiner for the entire examination. Evaluators are oriented and calibrated prior to the examination in an in-person session led by the Office for Dental Education, which goes over the case and the questions for each station. In two sessions held on the same morning, students rotated through all of the ten stations. Students take an OSCE appropriate to their expected skill level in Years 2, 3, and 4 of their predoctoral program.

At the end of each six-minute station, the examiners completed an evaluation form and graded students on a scale of eight points overall and two points for problem solving and communication skills for a total of ten points. The students were given a self-assessment form after the examination, and rated themselves on a four-point scale (Answered all questions correctly, Answered some questions with prompting, Answered all questions with prompting, Not correct/Could not answer) on their performance in each of the ten stations and also on a three-point scale (Excellent, Good, Poor) on their overall problem solving/analytical skills and their communication skills. Students were offered opportunities to review their performance and to reflect on the learning with each discipline faculty.

The structure of the OSCE was designed to test clinical skill performance and competence in skills such as communication, diagnosis, treatment planning, prescription writing, clinical techniques, radiographic image evaluation and interpretation of results. The OSCE consisted of ten stations covering endodontics, periodontics, pediatrics, orthodontics, oral radiology/pathology (OR/OP), oral surgery (OMFS), treatment planning, prosthodontics, oral health policy/epidemiology (OHPE), operative, and treatment planning. Each station was six minutes in total time, with students being examined on a one-to-one basis with faculty and postdoctoral examiners.

Analysis began by utilizing characteristic data about gender, overall undergraduate grade point average (GPA), undergraduate science GPA, Dental Aptitude Test (DAT) score, Perceptual Ability Test (PAT) score and dental school admissions interview score from the predoctoral classes of 2016 through 2021. This data was averaged and compiled along with each average OSCE score.

Data was analyzed using Stata software (Stata/MP 16.1 (StataCorp, College Station, TX) to determine the level of agreement between students' and examiners' assessment on OSCE performance. In addition, the correlation between students and examiners' assessments over the course of students' clinical education was evaluated. The Statistical difference in correlation over time was examined as well, and alpha was set at .05.

3. Results

The study included six predoctoral classes for a total of 198 students. One hundred and five of them were females (53.03%), and 93 male students (46.97%) (Table 1).

Table 1. Characteristics of students in classes of 2016-2021 by class (N=198)

Class	Female		Overall GPA[b]		Science GPA		DAT[c] score		PAT[d] score		Interview score		OSCE score	
	Frequency	(%)	Mean	SD[a]	Mean	SD[a]	Mean	SD[a]	Mean	SD[a]	Mean	SD[a]	Mean	SD[a]
Overall (N=198)	105	53.03	3.85	0.14	3.84	0.16	23.27	1.74	21.82	2.11	1.44	0.13	85.32	4.92
Class 2016 (N=31)	18	58.06	3.87	0.13	3.86	0.15	23.19	2.01	22.42	2.42	1.46	0.15	86.97	4.38
Class 2017 (N=34)	12	35.29	3.85	0.15	3.82	0.18	23.00	1.61	21.97	1.71	1.45	0.14	85.84	4.79
Class 2018 (N=32)	16	50.00	3.84	0.14	3.80	0.18	23.03	1.56	21.53	2.20	1.39	0.12	84.15	5.22
Class 2019 (N=34)	23	67.65	3.82	0.18	3.80	0.21	23.21	1.65	20.88	2.23	1.46	0.11	85.38	5.35
Class 2020 (N=34)	20	58.82	3.87	0.10	3.87	0.10	23.41	2.00	22.06	1.98	1.45	0.11	85.07	4.65
Class 2021 (N=33)	16	48.48	3.91	0.10	3.90	0.11	23.79	1.54	22.12	1.85	1.45	0.13	86.11	3.70

[a] Standard Deviation

[b] Grade Point Average

[c] Dental Admission Test

[d] Physical Abilities Test

Table 2. Linear Regression Analysis of the average difference in OSCE score by rater and students' characteristics between 2016-2021

	OSCE 1		OSCE 2		OSCE 3		Overall	
	Crude difference* (95%CI)	Adjusted difference* (95%CI)	Crude difference (95%CI)	Adjusted difference (95%CI)	Crude difference (95%CI)	Adjusted difference (95%CI)	Crude difference (95%CI)	Adjusted difference (95%CI)
Rater Faculty score Self-assessment	Ref. -7.07 (-8.20, -5.94)	Ref. -7.16 (-8.31, -6.02)	Ref. -5.61 (-6.94, -4.28)	Ref. -5.59 (-6.94, -4.24)	Ref. -7.04 (-8.20, -5.87)	Ref. -7.08 (-8.27, -5.89)	Ref. -6.69 (-7.62, -5.76)	Ref. -6.74 (-7.67, -5.80)
Gender Male Female		Ref. 0.63 (-0.91, 2.17)		Ref. -0.10 (-1.64, 1.43)		Ref. -0.59 (-1.99, 0.81)		Ref. -0.20 (-1.14, 1.10)
Overall GPA		2.87 (-3.42, 9.15)		0.20 (-6.39, 6.78)		2.66 (-2.75, 8.06)		1.71 (-2.95, 6.57)
DAT score		0.27 (-0.24, 0.77)		0.49 (-0.03, 1.01)		0.04 (-0.40, 0.47)		0.18 (-0.13, 0.61)
PAT score		-0.29 (-0.69, 0.11)		-0.25 (-0.62, 0.12)		0.06 (-0.24, 0.36)		-0.12 (-0.41, 0.14)
Interview score		0.97 (-5.11, 7.06)		4.48 (-1.61, 10.57)		-5.10 (-10.67, 0.48)		-1.25 (-4.84, 4.22)

95%CI: 95% confidence interval

*Crude difference is the difference between self-assessment score and faculty score without adjusting for students'

gender or admission factors. The adjusted difference is the difference taking into account the other factors in the table.

The class of 2019 had the highest proportion of female students (67.65%), while the class of 2017 had the lowest proportion of female students (35.29%). The average overall GPAs, Science GPAs, DAT scores, PAT scores, and admission scores were similar over the years in our population. After subtracting the student self-assessment score from faculty score in the crude difference in Table 2, the data was analyzed taking into account students' gender and admission score were confounding variables. Table 2 shows the adjusted difference after accounting for these items.

After compiling class characteristics, the average student self-assessment as a percentage compared to the average faculty assessment as a percentage was analyzed and is shown in Figure 1 for each OSCE.

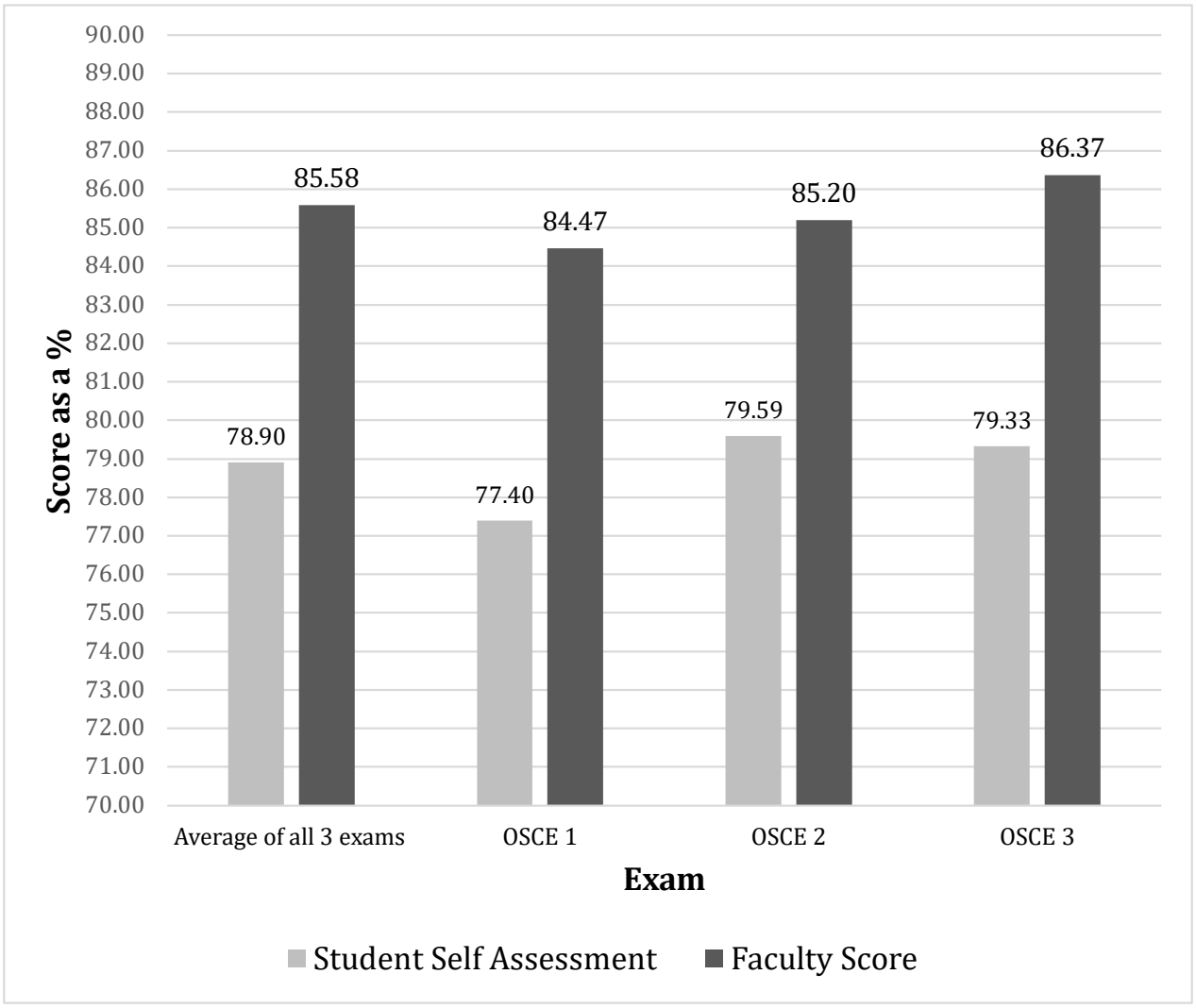

Figure 1. Student self-assessment vs faculty score for all 3 OSCE Exams

On average, students scored themselves 78.90 points out of 100 on all OSCEs, which is lower than that of faculty who scored the students 85.58 points out of 100. The closest difference was by 5.61 points in OSCE 2. For each successive OSCE, faculty scores increased starting from 84.47 out of 100 in OSCE 1 to 86.37 out of 100 in OSCE 3. While faculty scores increased for each OSCE, student scores did not, maxing out at 79.59 out of 100 in OSCE 2.

Students tended to rate themselves 6.68 points lower on an OSCE than faculty did (95%CI=-7.60, -5.76), and when adjusting for the characteristics of students mentioned in Table 1, the adjusted difference after stratifying by these items showed no changes larger than 10% were found. Therefore, none of these characteristics are considered confounding variables (Table 2).

A further analysis was run to determine not only the correlation between student and faculty assessment for each of the three OSCEs, but also the correlation for each discipline. This is summarized in Table 3 and visualized in Figure 2. The average correlation for student to faculty scores was poor at an ICC of 0.40 with a 95% confidence interval of 0.20 – 0.54 (Table 3).

Table 3. Measures of interrater consistency between faculty score and self-assessment for each OSCE and disciplines

Exam	OSCE 1		OSCE 2		OSCE 3		Average of 3 exams	
	Average Intraclass Correlation*	95% Confidence interval	Average Intraclass Correlation	95% Confidence interval	Average Intraclass Correlation	95% Confidence interval	Average Intraclass Correlation	95% Confidence interval
Overall	0.46	(0.27, 0.60)	0.39	(0.15, 0.57)	0.43	(0.23, 0.58)	0.40	(0.20, 0.54)
By discipline								
OMFS[a]	0.52	(0.35, 0.65)	0.58	(0.41, 0.70)	0.59	(0.44, 0.70)	0.55	(0.40, 0.66)
OR/OP[b]	0.56	(0.40, 0.68)	0.52	(0.32, 0.65)	0.46	(0.27, 0.60)	0.53	(0.38, 0.64)
Operative	0.45	(0.26, 0.60)	0.40	(0.16, 0.57)	0.30	(0.06, 0.49)	0.29	(0.06, 0.46)
Prosthodontics	0.23	(-0.24, 0.53)	0.39	(0.14, 0.56)	0.51	(0.34, 0.64)	0.40	(0.19, 0.56)
Endodontics	0.67	(0.55, 0.76)	0.58	(0.41, 0.70)	0.50	(0.32, 0.63)	0.50	(0.35, 0.62)
Periodontics	0.58	(0.43, 0.69)	0.56	(0.38, 0.69)	0.52	(0.35, 0.65)	0.57	(0.43, 0.67)
OHPE[c]	0.60	(0.46, 0.71)	0.31	(0.04, 0.51)	0.60	(0.46, 0.71)	0.39	(0.19, 0.54)
Pediatric Dentistry	0.57	(0.39, 0.69)	0.36	(0.11, 0.55)	0.58	(0.43, 0.69)	0.60	(0.47, 0.69)
Orthodontics	0.63	(0.49, 0.72)	0.52	(0.33, 0.66)	0.68	(0.56, 0.76)	0.63	(0.51, 0.72)
Treatment Planning	0.55	(0.39, 0.67)	0.30	(0.02, 0.50)	0.51	(0.33, 0.64)	0.53	(0.39, 0.65)

[a] Oral and Maxillofacial Surgery

[b] Oral Radiology and Oral Pathology

[c] Oral Health Policy and Epidemiology

*Reliability index for the similarity between students' self-assessment and faculty score. It ranges between 1 (perfect agreement) to 0 (perfect disagreement)

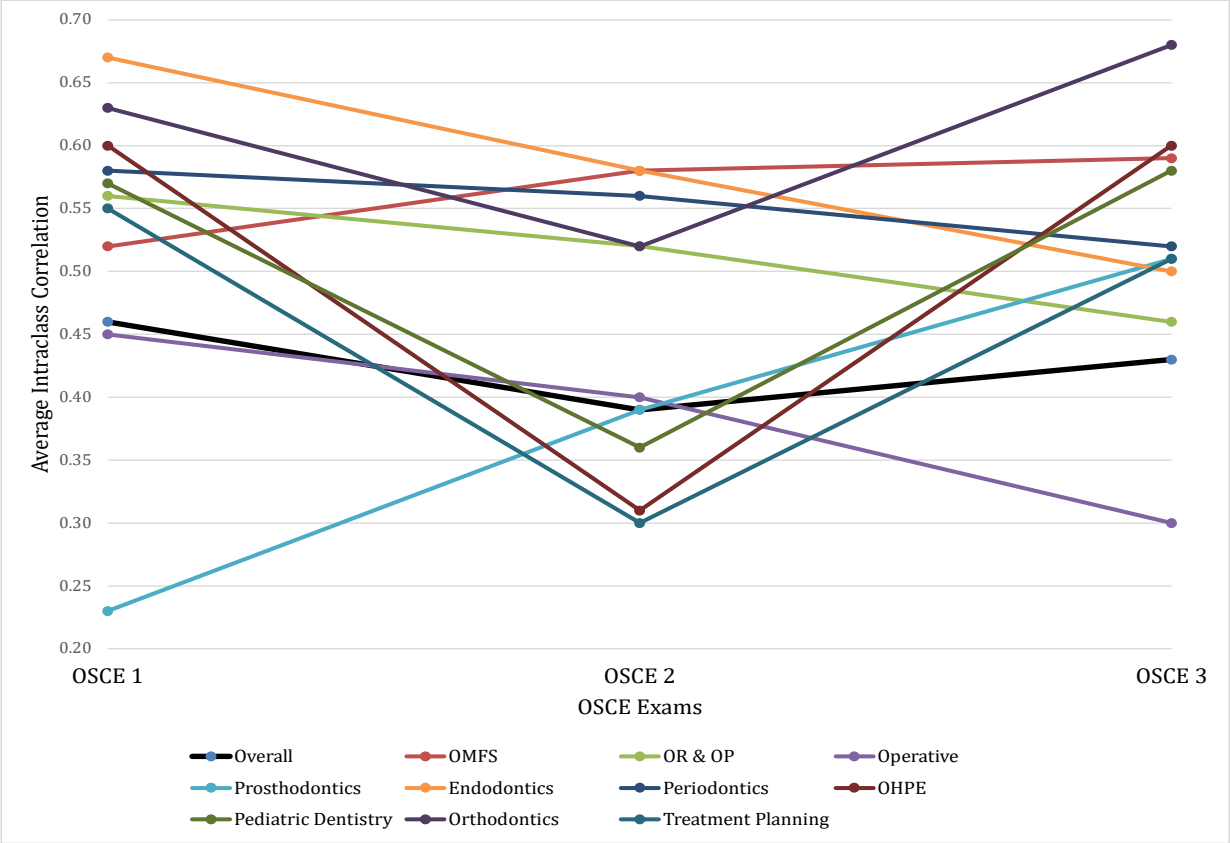

Figure 2. Correlation by discipline between student self-assessment and faculty score over three OSCE exams

The highest overall correlation for an OSCE was seen during OSCE 1 at an ICC of 0.46 (0.27-0.60) while the lowest correlation was seen in OSCE 2 at 0.39 (0.15-0.57). There was an overall trend that the correlation dipped at OSCE 2, however, the difference in correlation over time was not statistically significant (Figure 2). In the first OSCE, endodontics, orthodontics, OHPE, periodontics, pediatric dentistry, treatment planning, OR & OP, and OMFS have moderate correlation,

which stayed the same by OSCE 3 for every category, except OR & OP falling from 0.56 (0.40-0.68) to 0.46 (0.27-0.60). While moderate correlation remained, the actual ICC values fell from OSCE 1 to OSCE 3 in OR &OP, endodontics, periodontics, and treatment planning. Operative had poor correlation in OSCE 1 and the ICC value fell even lower by OSCE 3 from 0.45 (0.26-0.60) to 0.30 (0.06-0.49). On the other hand, while prosthodontics had a poor reliability in OCSCE 1, it rose to moderate reliability in OSCE 3 from 0.23 (-0.24-0.53) to 0.51 (0.34-0.64). The only statistically significant difference between OSCE 1 and OSCE 3 was seen in endodontics, which fell from 0.67 (0.55-0.76) to 0.50 (0.32-0.63).

Finally, a correlation analysis was run by splitting the OSCE data by those students who scored lower or equal to the median faculty score and those students who scored higher than the median faculty score. The results are shown in Table 4.

Table 4. Correlation between overall student-faculty score, stratified by students' median performance

	Lower Performing Students[d]		Higher Performing Students[e]	
	Average Intraclass Correlation	95% Confidence Interval	Average Intraclass Correlation	95% Confidence Interval
Overall Average	0.15	(-0.30, 0.45)	0.33	(-0.04, 0.56)
OMFS[a]	0.54	(0.33, 0.69)	0.60	(0.41, 0.74)
OR/OP[b]	0.52	(0.29, 0.67)	0.60	(0.41, 0.73)
Operative	0.41	(0.13, 0.60)	0.05	(-0.42, 0.35)
Prosthodontics	0.47	(0.21, 0.65)	0.23	(-0.22, 0.52)
Endodontics	0.38	(0.08, 0.58)	0.61	(0.43, 0.74)
Periodontics	0.56	(0.35, 0.70)	0.54	(0.31, 0.69)
OHPE[c]	0.61	(0.42, 0.73)	0.33	(-0.00, 0.55)
Pediatric Dentistry	0.61	(0.42, 0.73)	0.43	(0.17, 0.62)
Orthodontics	0.67	(0.51, 0.77)	0.56	(0.34, 0.70)
Treatment Planning	0.56	(0.36, 0.70)	0.56	(0.35, 0.71)

[a] Oral and Maxillofacial Surgery

[b] Oral Radiology and Oral Pathology

[c] Oral Health Policy and Epidemiology

[d] Students who scored lower than or equal to the median faculty score

[e] Students who scored higher than the median faculty score

When splitting the data based upon those students above and those students below the median faculty score. The overall average for lower performing students had a lower ICC at 0.15 (-0.30-0.45) compared to 0.33 (-0.04-0.56) for higher performing students (Table 4). This trend was consistent for OMFS, OR/OP, and endodontics. On the other hand, lower performing students had a of higher ICC for operative, prosthodontics, periodontics, OHPE, pediatric dentistry, and orthodontics. Treatment planning was the only discipline in which both sets of students shared the exact same ICC value. While these are trends, there were some significant findings. For operative, lower performing students had a higher correlation 0.41 (0.13-0.60) than higher performing students 0.05 (-0.42-0.35). For OHPE, lower performing students had a higher correlation 0.61 (0.42-0.73) than higher performing students 0.33 (-0.00-0.55). For endodontics, higher performing students had a higher correlation 0.61 (0.43-0.74) than lower performing students 0.38 (0.08-0.58).

5. Discussion

The current study aimed to determine whether student self-assessment on OSCE performance was related to faculty assessment, and whether the progression through the clinical program had any effect on self-assessment skills. Overall, there was a poor to fair level of correlation between faculty and students for all OSCEs based upon Koo and Li's guidelines (Koo & Li, 2016) and Cicchetti's guidelines (Cicchetti, 1994) respectively. When taking the average of all three OSCEs, the greatest amount of agreement was found for orthodontics, pediatrics, and periodontics. Based upon the low ICCs for the average of all three OSCEs there was poor reliability between faculty and student for operative, OHPE, and prosthodontics. While that is true for the average of all three OSCEs, by the third OSCE, the data shows only poor reliability for operative while oral health policy and epidemiology actually has one of the highest correlations. Therefore, upon review of the OSCE 3 data by discipline, most disciplines show a moderate correlation between students and faculty assessments.

In the current study, despite expectations as students progressed through the clinical program, their self-assessments did not correspond more closely to faculty assessments as seen in Figure 2. Over the course of the three OSCEs only prosthodontics and oral surgery have an increase in correlation for each OSCE. A previous study (Park, Cox, Susarla, Da Silva, & Howell, 2007) demonstrated that in the predoctoral year 3 prosthodontics course, student exam

performance was not affected by whether or not postdoctoral teaching assistants had had previous teaching experience; however, the results of this study appear to indicate that participation in the course provides students with a more solid basis for self-evaluation of their OSCE performance in this particular subject. On the other hand, endodontics, periodontics, OR/OP, and operative all have decreases in correlation over the course of each OSCE.

It is interesting to note that the rest of the subject areas (OHPE, pediatrics, orthodontics, and treatment planning) all drop lower at OSCE 2 and then return to a similar correlation for OSCE 3. This trend is consistent with the overall average data that shows a fall in correlation from OSCE 1 to OSCE 2 and then a rise from OSCE 2 to OSCE 3. This overall trend may be explained as students realize they have more to learn in the coming years when taking their first OSCE, so their scores correlate more to faculty assessment. Then during their second OSCE, they may not have an accurate assessment of their skills, as students may fear they are falling behind or have had different clinical experiences than their peers. Finally during OSCE 3 they may have a better understanding of how to assess themselves after spending more time in clinic, but may still be unsure of their clinical skills.

Contrary to expert learners, dental students exposed to early preclinical and clinical learning experience in particular require structure and depend on frequent feedback because they are not competent in accurately assessing their own work and identifying areas of deficiencies in need of improvement. (Kruger & Dunning, 1999) Rubrics can be an important tool for students to use to measure and monitor their progress and to reflect on their learning. Developing a rubric could help guide the students and faculty as evaluators and can serve as a communication tool to achieve calibration. (Stoddard, Labrecque, & Schonfeld, 2016; Donato & Harris, 2013) It would be useful for assessment criteria and grading rubrics to include specific details to establish performance standards for both the students and faculty evaluators.

There were some limitations to this study including the fact that the examiners were separated into two groups, each assessing different students and there is the possibility that the examiners differed in their scoring. Previous work by Park et al. (Park, Kim, Kristiansen, & Karimbux, 2015) found that part-time faculty members gave higher scores than full-time faculty members or postgraduate residents. However, since all data for this study was de-identified, it was not possible to determine which students had been scored by part-time or full-time faculty examiners in any of the ten disciplines. All faculty examiners participate in at least one calibration session before participating in an OSCE; however, further studies could investigate the interrater reliability of faculty OSCE scoring and relationship, if any, between full-time and part-time faculty examiner scores and student self-assessment scores.

Another limitation stems from the fact that since the data was de-identified, it was not fully possible to determine whether stronger and weaker students differed in their self-assessment skills. Although this is true, the data was slit based upon the OSCE itself to determine if those who performed better than the median faculty score on an OSCE had a higher correlation than those who scored lower or equal to the median faculty score. Table 4 shows that while these stronger performing students have over double the ICC value of weaker students, both sets of data fall into the poor reliability category. When reviewing the disciplines individually, there was a difference seen in operative and oral health policy and epidemiology where higher performing students had a lower correlation to faculty than lower performing students. The inverse was noted for endodontics, where lower performing students had a lower correlation to faculty scores. This may be explained that students who were higher performing may be overly critical about certain areas of operative compared to lower performing students. The same may be true for OHPE, where higher performing students may think that due to the complexities of this field, they may have less knowledge and rate themselves lower than faculty. As for endodontics, higher performing students may see this as a more straightforward discipline than those of lower performance.

It was the observations of the faculty evaluators, that weaker students had a tendency to assess themselves more highly than the stronger learners, consistent with other studies that showed that poor performing students were deficient in self-assessment skills (Kruger & Dunning, 1999; Hodges, Regehr, & Martin, 2001; Austin & Gregory, 2007; Colthart et al., 2008; Edwards, Kellner, Sistrom, & Magyari, 2003). It was noted that weaker students had a tendency to overestimate their understanding of knowledge and their clinical abilities. To the contrary, stronger performing students were often more cautious of overestimating their performance in their self-assessments. This finding also coincides with evidence shown by Hodges el al. that the low performers in the study were inaccurate in their self-assessments and the high performers generally underestimated their skills initially. The high performers were able to modify their assessments after observing the performance of their peers, whereas the weak performers were not able to re-adjust their assessments (Hodges et al., 2007). This may mean that accurate self-assessing will need to be stressed more in dental education as ensuring an accurate assessment of one's work is paramount in the field of dentistry. It may also show a need for students to recalibrate their self-assessment skills, especially amongst those weaker students. It is also interesting to note that Figure 1 shows students consistently scored themselves lower on average than faculty did for every OSCE. This may stem from the difference in scoring scales for faculty assessments and student assessments.

The current study showed a moderate correlation between students and faculty assessments throughout the program and

whether a relationship exists between reflective assessments and student performance is a topic for future research. Few studies are available on the effect of assessment criteria and rubrics on outcomes measures, and this prospective data is being investigated as part of our future study to measure outcomes assessment data from this program. It could contribute knowledge toward understanding student clinical performance with the new focus on reflection as a learning skill to overall student performance.

6. Conclusion

Designing and applying appropriate evaluation instruments can contribute to effective student learning. Understanding that teaching dental students how to accurately self-assess can contribute to improved performance overall, the objective structured clinical examination (OSCE) can be designed to incorporate self-reflection and provide an opportunity for students develop self-assessment skills.

References

Austin, Z., & Gregory, P. A. (2007). Evaluating the accuracy of pharmacy students' self-assessment skills. *American Journal of Pharmaceutical Education, 71*(5), 89. https://doi.org/10.5688/aj710589

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.

Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education, 22*(4), 444-57. https://doi.org/10.1080/0969594X.2014.996523

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284-290. https://doi.org/10.1037/1040-3590.6.4.284

Colthart, I., Bagnall, G., Evans, A., Allbutt, H., Haig, A., Illing, J., & McKinstry, B. (2008). The effectiveness of self-assessment on the identification of learner needs, learning activity, and impact on clinical practice: BEME guide no. 10. *Medical Teacher, 30*(2), 124-45. https://doi.org/10.1080/01421590701881699

Commission on Dental Accreditation. (2012). Self-study guide for dental education programs. Chicago, IL: American Dental Association, Section 2-10.

Donato, A. A., & Harris, I. (2013). Use of portfolios for assessment of resident teaching skills. *Journal of Graduate Medical Education*, *5*(3), 476-480. https://doi.org/10.4300/JGME-D-12-00309.1

Edwards, R. K., Kellner, K. R., Sistrom, C. L., & Magyari, E. J. (2003). Medical student self-assessment of performance on an obstetrics and gynecology clerkship. *American Journal of Obstetrics and Gynecology, 188*(4), 1078-82. https://doi.org/10.1067/mob.2003.249

Eva, K., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation and research agenda. *Academic Medicine, 80*(10 Suppl), S46-54. https://doi.org/10.1097/00001888-200510001-00015

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395-430. https://doi.org/10.3102/00346543059004395

Graham, R., Zubiaurre Bitzer, L. A., & Anderson, O. R. (2013). Reliability and predictive validity of a comprehensive preclinical OSCE in dental education. *Journal of Dental Education, 77*(2), 161-167. https://doi.org/10.1002/j.0022-0337.2013.77.2.tb05458.x

Hodges, B., Regehr, G., & Martin, D. (2001). Difficulties in recognizing one's own incompetence: novice physicians who are unskilled and unaware of it. *Academic Medicine, 76*(10 suppl), S87-S89. https://doi.org/10.1097/00001888-200110001-00029

Illeris, K. (2003). Toward a contemporary and comprehensive theory of learning. *International Journal of Lifelong Education, 22*(4), 396-406. https://doi.org/10.1080/02601370304837

Kim, A., Chutinan, S., & Park, S. E. (2015). Assessment skills of dental students as peer evaluators. *Journal of Dental Education*, *79*(6), 653-57. https://doi.org/10.1002/j.0022-0337.2015.79.6.tb05937.x

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*; *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kramer, G. A., Albino, J. E. N., Andrieu, S. C., Hendricson, W. D., Henson, L., Horn, B. D., Neumann, L. M., & Young, S. K. (2009). Dental student assessment toolbox. *Journal of Dental Education, 73*(1), 12-35. https://doi.org/10.1002/j.0022-0337.2009.73.1.tb04636.x

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121-34. https://doi.org/10.1037/0022-3514.77.6.1121

Krupat, E., Richards, J. B., Sullivan, A. M., Fleenor, T. J., & Schwartzstein, R. M. (2016). Assessing the effectiveness of case-based collaborative learning via randomized controlled trial. *Academic Medicine*, *91*(5), 723-29. https://doi.org/10.1097/ACM.0000000000001004

Langendyk, V. (2006). Not knowing that they do not know: self-assessment accuracy of third-year medical students. *Medical Education, 40*(2), 173-179. https://doi.org/10.1111/j.1365-2929.2005.02372.x

Lele, S. M. (2011). A mini-OSCE for formative assessment of diagnostic and radiographic skills at a dental college in India. *Journal of Dental Education, 75*(12), 1583-1589. https://doi.org/10.1002/j.0022-0337.2011.75.12.tb05218.x

McKeachie, W. J., Pintrich, P. R., Lin, Y. G., & Smith, D. A. F. (1986). Teaching and Learning in the College Classroom: A Review of the Literature. Ann Arbor, MI: University of Michigan Press.

National Research Council. (2000). How People Learn: Brain, Mind, Experience, and School: Expanded Edition. Washington, DC: The National Academies.

Park, S. E., & Howell, T. H. (2015). Implementation of a flipped classroom educational model in a predoctoral curriculum. *Journal of Dental Education,79*(5), 563-70. https://doi.org/10.1002/j.0022-0337.2015.79.5.tb05916.x

Park, S.E., Anderson, N., & Karimbux, N. (2016). OSCE and case presentations as active assessments of student performance. *Journal of Dental Education*, *80*(3), 334-38. https://doi.org/10.1002/j.0022-0337.2016.80.3.tb06089.x

Park, S.E., Cox, C.K., Susarla, S.M., Da Silva, J.D., & Howell, T.H. (2007). Do tutor expertise and experience influence student performance in a problem-based curriculum? *Journal of Dental Education*, *71*(6), 819-24. https://doi.org/10.1002/j.0022-0337.2007.71.6.tb04338.x

Park, S. E., Kim, A., Kristiansen, J., & Karimbux, N. (2015). The influence of examiner type on student OSCE scores. *Journal of Dental Education, 79*(1), 89-94. https://doi.org/10.1002/j.0022-0337.2015.79.1.tb05861.x

Park, S. E., Price, M. D., & Karimbux, N. Y. (2018). The dental school interview as a predictor of predoctoral OSCE performance. *Journal of Dental Education*, *82*(3), 269-76. https://doi.org/10.21815/JDE.018.026

Prober, C. G., & Heath, C. Lecture halls without lectures – a proposal for medical education. (2012). *New England Journal of Medicine, 366*, 1657-1659. https://doi.org/10.1056/NEJMp1202451

Regehr, G., & Eva, K. (2006). Self-assessment, self-direction, and the self-regulating professional. *Clinical Orthopaedics and Related Research, 449*, 34-8. https://doi.org/10.1097/01.blo.0000224027.85732.b2

Schoonheim-Klein, M.E., Habets, L.L., Aartman, I.H., Van der Vleuten, C.P., Hoogstraten, J., & van der Velden, U. (2006). Implementing an Objective Structured Clinical Examination (OSCE) in dental education: Effects on students' learning strategies. *European Journal of Dental Education, 10*(4), 226-235. https://doi.org/10.1111/j.1600-0579.2006.00421.x

Stoddard, H. A., Labrecque, C. A., & Schonfeld, T. (2016). Using a scoring rubric to assess the writing of bioethics students. *Cambridge Quarterly of Healthcare Ethics*, *25*(2), 301-311. https://doi.org/10.1017/S0963180115000602

Tain, M., Schwartzstein, R., Friedland, B., & Park, S. E. (2017). Dental and medical students' use and perceptions of learning resources in a human physiology course. *Journal of Dental Education*, *81*(9), 1091-97. https://doi.org/10.21815/JDE.017.063

White, C. B., Ross, P. T., & Gruppen, L. D. (2009). Remediating students' failed OSCE performances at one school: The effects of self-assessment, reflection, and feedback. *Adademic Medicine*, *84*(5), 671-4. https://doi.org/10.1097/ACM.0b013e31819fb9de

Yu, A., Pagni, S. E., Park, S. E., & Karimbux, N. (2020). Early clinical exposure in US dental schools and correlation with earlier competencies evaluation. *Journal of Dental Education*, *84*(2), 151-56. https://doi.org/10.21815/JDE.019.169