

Evaluating Student Self-Assessment through Video-Recorded Patient Simulations

Tammy R. Sanderson, RDH, MSDH; Rachel C. Kearney, RDH, MS; Denise Kissell, RDH, MPH; Jessica Salisbury, RDH, MS

Abstract

Purpose: The purpose of this pilot study was to determine if the use of a video-recorded clinical session affects the accuracy of dental hygiene student self-assessment and dental hygiene instructor feedback.

Methods: A repeated measures experiment was conducted. The use of the ODU 11/12 explorer was taught to students and participating faculty through video and demonstration. Students then demonstrated activation of the explorer on a student partner using the same technique. While faculty completed the student assessment in real time, the sessions were video recorded. After completing the activation of the explorer, students and faculty completed an assessment of the student's performance using a rubric. A week later, both students and faculty viewed the video of the clinical skill performance and reassessed the student's performance using the same rubric. The student videos were randomly assigned a number, so faculty reassessed the performance without access to the student's identity or the score that was initially given.

Results: Twenty-eight students and 4 pre-clinical faculty completed the study. Students' average score was 4.68 ± 1.16 on the first assessment and slightly higher 4.89 ± 1.45 when reviewed by video. Faculty average scores were 5.07 ± 2.13 at the first assessment and 4.79 ± 2.54 on the second assessment with the video. No significant differences were found between the differences in overall scores, there was a significant difference in the scores of the grading criteria compared to the expert assessment scores ($p=0.0001$).

Conclusion: This pilot study shows that calibration and assessment without bias in education is a challenge. Analyzing and incorporating new techniques can result in more exact assessment of student performance and self-assessment.

Keywords: student self-assessment, clinical assessment, dental education, dental hygiene, faculty calibration

This study supports the NDHRA priority area, **Professional Education and Development:** Validate measures that assess continued clinical competency.

INTRODUCTION

The importance of self-assessment in health care related occupations has been well established in the literature and the ability to self-assess is vital to ensure professional growth and development.^{1,2} One challenge that arises in the implementation of self-assessment is students' understanding of self-assessment and its purpose. Students have reported that they are unfamiliar with the concept of self-assessment or are unclear of the expectations especially in the clinical environment.^{1,2} In some cases it is due to a lack of exposure to self-assessment. In a dental hygiene study, Mould et al reported that 52% of students had no previous experience with self-assessment. An additional 24% of students reported minimal experience with self-assessment.² Medical literature shows that students do not accurately self-assess and that there is a need to provide self-assessment instruction during clinical education.¹

In dental hygiene, self-assessment is a necessary

component of the curriculum, but can be difficult to cultivate. The Commission on Dental Accreditation (CODA) Standards for Dental Hygiene Programs outlines the responsibilities of institutions, putting the onus on educators to incorporate the self-assessment process into the curriculum.³ Critical thinking and self-assessment expectations are noted in standard two for educational programs stating "Graduates must be competent in the application of self-assessment skills to prepare them for life-long learning."³ Consequently, dental hygiene programs must have evaluation mechanisms in place to examine student understanding as well as ability to apply self-assessment skills.³ CODA accreditation standards advise dental hygiene programs to demonstrate that students are competent in the application of self-assessment. Programs can demonstrate compliance through "evaluation mechanisms designed to monitor knowledge and performance."³

The challenge of teaching the skills of effective self-assessment is one seen in other health professions. In addition, health profession studies have shown that the evaluation scores from student self-evaluation do not correlate with the scores given by faculty members.^{4,5} In providing instruction to students, educators must be able to show students discrepancies in their performance so that the student can accurately self-assess in alignment with the instructor or expert assessment. The use of video has offered the ability to record student performance for later review and critique. Video may provide the opportunity for a student to re-examine the clinical skill performance to identify deficiencies in their performance thus having the potential to improve self-assessment practices. This is already a standard educational practice in some medical programs that could be applied to dentistry and dental hygiene as well. A study involving physiotherapy students reported that a recorded simulated exam allowed students to reflect on and evaluate their performance against established standards.⁶ Based on those study findings, students can use video to compare their perceived performance to the actual recorded performance. This process of video review may provide the student with an opportunity to recognize areas of adequate and deficient performance. In addition, the process of video review could provide a measurable objective to discern whether or not self-assessment practices are in alignment with instructor assessment.

Clinical instructors could benefit from the utilization of video review as well. Observation is the most frequently utilized type of clinical assessment technique among dental hygiene educators⁷ but direct observation "is limited by high interrater score variability."⁸ The use of video will provide an alternative assessment technique for clinical instructors or even enhance the technique of the direct observation. The use of video as an alternative method for clinic assessment may help to reduce rater bias. In addition, instructors can use the video-recorded performances to determine if the feedback provided is complete and identify missed opportunities to provide corrective feedback. The feedback offered during a direct observation experience may differ from the feedback offered after a video review of the clinical performance.

Dental hygiene students must learn to evaluate their clinical performance to ensure that safe and appropriate care is provided to patients, however there is currently not a systematic method of self-assessment training for these students. The utilization of a video recorded clinical simulated patient learning experience is an alternative teaching method that can be investigated for its usefulness in the training of dental hygiene students and its effect on dental hygiene clinical instructor feedback. Other

health professions have used instructional media such as simulation and video to provide students with opportunities to reflect on their clinical performance.^{4,9,10} Students from other allied health professions have reported that the use of videos allowed them to more accurately self-assess.^{5,11} A medical study found that the implementation of video exercises allows for improved calibration between student self-assessment and faculty feedback.⁵ Although video is a widely used form of technology, its most effective use in the training of dental hygiene students has yet to be determined. Further, the use of a video-recorded clinical session used for student self-assessment has not been evaluated for dental hygiene. Because dental hygiene has a unique psychomotor skill set, it must be established whether or not the use of video is a relevant tool to evaluate the accuracy of student self-assessment practices. The purpose of this study is twofold in that it seeks to determine if the use of a video-recorded clinical session affects the accuracy of dental hygiene student self-assessment, if the use of a video-recorded clinical session influences feedback provided to students by dental hygiene clinical instructors.

METHODS AND MATERIALS

This study was a repeated measures experiment that received expedited approval from an institutional review board. In preparation for the study, reading assignments, a video demonstrating the technique for using an ODU 11/12 explorer, and a self-assessment orientation were completed. First year dental hygiene students in the pre-clinical course were invited to participate in this pilot study. Clinical instructors viewed the same video that the students watched demonstrating the technique for using an ODU 11/12 explore. Review of this benchmark video before the laboratory exercise was intended to calibrate both students and clinical instructors. Students then demonstrated the proper technique, indicated in the benchmark video for using the ODU/11/12 explorer on a simulated patient. There were 4 criteria used to assess the clinical skill performance of exploring with the ODU 11/12 explorer. The first criterion assessed the choice of the correct working end, the second criterion assessed adaptation of the instrument to the teeth, the third criterion assessed instrument insertion into the gingival tissue, and the fourth criterion assessed instrument activation. These scores were compared to a standardized score established for each video based on criteria taught from the textbook by the pre-clinical faculty and video review by all faculty prior to the real time assessment. The clinical instructor observed the techniques of the student while one of the investigators video-recorded the clinical performance. Immediately following the performance, the student completed a self-assessment and the clinical instructor completed a faculty

assessment of student performance. The data collected from this clinical session was classified as the pre-video assessment.

The post-video assessment was completed a week later using a secure streaming server and email contact entailing evaluation instructions and a link to the video. This link connected viewers to the video-recorded clinical performance and an electronic version of the rubric used for the pre-video assessment. The post-video assessment was completed through a link to a secure electronic evaluation form in a survey software program (Qualtrics, Provo, Utah). In addition to the faculty score, the primary investigator and the co-investigator along with 2 pre-clinic course directors established a score for each video performance. The score established by the investigators and course directors is referred to as the "expert assessment" for each video performance. The standardized or expert assessment was determined by a negotiated approach. The expert assessment was the comparison for other scores completed by the students and the faculty.

Both descriptive and inferential statistics were used to analyze the data. Statistical analysis involved a total sum score and sub categorical scores assigned to each of the rubrics received from the students, instructors and investigators. These scores were totaled from the assessment rubric using the 4 criteria described previously. The criteria on the rubric were used for students and instructors to rate the performance of the clinical skill. The criteria were each scored on a 0 to 2 scale; a score of 0 represented major errors, a score of 1 represented minor errors in performance or inability to complete the skill, and a score of 2 represented the ability to perform the skills without errors. Scores were totaled by rater (instructor or student) and by time (pre- and post-video review). A mixed model for repeated measures was used to account for the association of scores between the student pre- and post-video reviews (intra-rater comparison), clinical instructor pre and post video reviews (intra-rater comparison), as well as associations between students, instructors and the expert assessment score (inter-rater comparison). The following parameters were estimated based on the mixed model: the difference of total score and criterion scores at different time intervals for the student self- assessment, the difference of total score and criterion scores at different time intervals by the instructor, difference between instructor and student self-assessment at different time intervals, and difference of scores evaluated by the student or instructor as compared to the expert assessment scores. A power analysis of 85% was determined and an alpha level of 0.01 was used for statistical testing.

RESULTS

Thirty-two first year dental hygiene students were invited to participate in the study. Twenty-eight of the students provided consent for an 87.5% participation rate. Due to an incomplete data set, data analysis was completed for 27 subjects (n=27). Pair-wise comparisons using both parametric and non-parametric methods were analyzed. In Table I, the data was summarized by count. Students scored their performance as a 0, 1 or a 2. For criterion one, 2 students scored their performance as a zero before the video was reviewed. After the video was reviewed, 5 students scored their performance as a 0 for criteria one. Instructors assigned a 0 score to 7 students for criteria one for both pre and post-video review assessments. There were 5 zeroes assigned for criteria one by the expert assessment group. The counts for the other criteria can be reviewed in Table I. Data analysis showed an increase in mean student self-assessment scores post-video review. Average instructor scores decreased post-video review. Overall scores assessed by the instructors and by the students were rated higher as compared to the expert assessment mean scores.

Differences in scores from the expert assessment scores by rater, time and criterion can be reviewed in Table II. Analysis of variance from the summary data can be examined in Table III. No statistically significant difference was found between pairs after adjustments were made using the Tukey-Kramer method. Overall scores were statistically different by criteria from the expert assessment scores. The pair-wise comparisons of scores can be reviewed in Table IV.

DISCUSSION

This study investigated the effect of video on assessment by comparing differences in student self-assessment scores as well as differences in instructor scores. The ratings by the student and by the instructor were compared to an expert assessment rating. There were 4 criteria used to assess the clinical skill performance of exploring with the ODU 11/12 explorer. The first criterion assessed the choice of the correct working end, the second criterion assessed adaptation of the instrument to the teeth, the third criterion assessed instrument insertion into the gingival tissue, and the fourth criterion assessed instrument activation.

There was an interesting trend observed when comparing students' scores before and after video review. Student scores were higher than the expert assessment scores before review of the video and lower after review of the video for criterion one. It is possible that the students were more critical of their performance after the video review and recog-

Table I: Data Summary Count by Rubric Category and Overall Score

Rubric Criteria	Criterion 1			Criterion 2			Criterion 3			Criterion 4			Total			
	0	1	2	0	1	2	0	1	2	0	1	2	0 to 3	4 to 6	7 to 8	Average
Student																
Pre-Video Review	2	5	21	1	25	2	5	21	2	2	22	4	5	22	1	4.68±1.16
Post-Video Review	5	3	20	1	22	5	3	19	6	4	18	6	4	21	3	4.89±1.45
Instructor																
Pre-Video Review	7	2	19	3	15	9	4	17	7	4	12	12	7	10	10	5.07±2.13
Post-Video Review	7	2	19	6	12	10	9	11	8	6	9	13	8	11	9	4.79±2.54
Expert Assessment																
Post-Video Review	5	1	21	15	8	4	16	5	6	7	4	16	12	9	6	4.15±2.55

nized areas for improvement. A physiotherapy study lends support to this trend where students reported that the experience helped them to see how much they needed to work on.⁶ Conversely, students rated their performance slightly higher for criterion two after they reviewed their video-recorded performance. The students may have observed their performance on video to be better than they had perceived during the actual clinical performance. This tendency has been noted in medical literature whereby higher performing students underrate their scores.¹ A similar trend related to assessment of the insertion of the instrument was seen with criterion three. One theory for this trend is that previous assessment experiences and feedback from the clinical instructors may not have been consistent with the established standards as evidenced by the differences in scores from the expert assessment means.

Although not statistically significant, there were some numeric changes in scores by instructors after they reviewed the video. For criteria two and three the instructors decreased the scores from the ratings they would have given pre-video review (during the direct observation/real-time experience). Similar trends were reported in a study by Benson et al where higher scores were assessed when students were evaluated in real-time compared to scores assessed in a videotaped evaluation.¹² The decrease in scores could also be due to the fact that instructors were blinded to which students' video they were scoring. Grading bias during the direct observation could have accounted for the higher scores. The expert assessment scores derived from blinded review of the videos by the investigators and preclinical course directors. In other words, the raters did not know which students they were assessing. It is possible that the instructors tend to rate higher because of their close interaction with the students in the clinical setting. Although the instructors intend to evaluate the students to a competent standard, the instructors may be recognizing the students as novice clinicians. In medical education, positive bias has been referred to as "generosity error."¹³

Table II: Mean differences of Scores from Expert Assessment for Rater, Time and Criteria

Rater and Time	Criterion 1	Criterion 2	Criterion 3	Criterion 4
Student Pre-Video	0.074	0.444	0.259	-0.259
Student Post-Video	-0.074	0.556	0.481	-0.259
Instructor Pre-Video	-0.185	0.593	0.444	-0.037
Instructor Post-Video	-0.185	0.519	0.296	-0.111

Table III: Summary Statistics of the 3-Way Analysis of Variance

Effect	Num DF	F Value	Pr>F
Criteria	3	25.74	<0.0001
Rater	1	0.04	0.8367
Time	1	0.04	0.8367
Criteria*Rater	3	1.30	0.2728
Criteria*Time	3	0.14	0.9338
Rater*Time	1	0.80	0.3719
Criteria*Rater*Time	3	0.65	0.5847

Another factor that could have influenced the instructor evaluations is a problem noted in medical education. Instructors do not want to show favoritism so they assess higher scores when evaluating students in small group.¹³ Without realizing, instructors may take into account the students' attitude and personality when evaluating the skill performance.^{12,14} For criterion one, there was no change with instructors' scores between pre- and post-video review. For criterion four, instructors increased their ratings after video review. For clinical instructors, the use of video could allow for a more accurate assessment of the student's clinical skill performance.¹² Some of the differences between instructor ratings

Table IV: Pairwise Comparisons of Scores from Expert Assessment

Pairs	Estimate	Standard Error	DF	t Value	Pr> t	Adjusted p
1 to 2	-0.6204	0.09521	405	-6.52	<0.0001	<0.0001
1 to 3	-0.4630	0.09521	405	-4.86	<0.0001	<0.0001
1 to 4	0.07407	0.09521	405	0.78	0.4370	0.8644
2 to 3	0.1574	0.09521	405	1.65	0.0991	0.3500
2 to 4	0.6944	0.09521	405	7.29	<0.0001	<0.0001
3 to 4	0.5370	0.09521	405	5.64	<0.0001	<0.0001

and expert assessment ratings could be attributed to the need for additional calibration. Although a video and rubric were provided before the study to demonstrate the assessment parameters, a reliability quotient was not established. In addition, the accuracy of assessment by the instructors could be improved through training of junior faculty with senior faculty. Nursing research reports that novice faculty are hesitant to assess lower grades as this may in turn effect their evaluations by the students.¹⁵ The patterns noted in this study may indicate a need for an unbiased method for grading students in the clinical setting.

The comparison of instructor to student scores showed varied trends. For criterion one, student mean scores decreased after video review while the instructor mean scores remained consistent. Conversely, student scores remained consistent and instructor scores decreased after video review for criterion four. With regard to criteria two and three, the students increased their scores while the instructors decreased the scores post-video review. Overall means between instructor and student scores post video review differed by 2.07%. The relationship between clinical instructor scores and student self-assessment scores was discussed by Geissler who reported a difference of 5% between student and faculty scores.¹⁶

Overall means by criteria are significantly different from the expert assessment means. This difference suggests that criteria one and four were well understood by the students and instructors. It is also possible that the assessment criteria for two and three required additional clarification for the students and instructors. Other limitations discovered through this pilot study were the small sample size and the need for additional calibration among instructors.

Even though the use of video did not have a statistically significant effect on student self-assessment scores, it may still provide value as a teaching tool. Today's generation of students are expecting technology to be incorporated into their education.¹⁷ In-

structors can consider using video to review a clinical performance with a student and compare self-assessment ratings to instructor ratings. This method could allow instructors to help students more accurately assess by reviewing performance deficiencies as well as proficiencies.¹⁸ The results of this pilot study can be used as a foundation for a full-scale study. Additional research related to the use of video-recorded patient simulations as a method for evaluating student self-assessment is warranted.

CONCLUSION

Self-assessment has been defined broadly as the involvement of learners in judging whether or not learner-identified standards have been met.¹⁸ Although there is not currently a systematic method for self-assessment training of dental hygiene students, accreditation requires dental hygiene programs to have evaluation mechanisms in place to examine students' understanding as well as ability to apply self-assessment skills.³ This pilot study aimed to bridge the gap by providing a basis for future investigation into the use of video to aid in the self-assessment training of dental hygiene students. In addition, the self-assessment strategy could be used in a continuum of time to indicate progression of skill and student acknowledgement of their strengths and weaknesses. Moreover this strategy could be applied to assessment of other instruments used for clinical performance such as curets or scalers. The data collected in this study also evokes a need for inquiry into the use of video for calibration of dental hygiene faculty. It is necessary to discover a valid method for self-assessment training of dental hygiene students.

Tammy R. Sanderson, RDH, MSDH is the senior clinic coordinator at the College of Southern Nevada. Rachel C. Kearney, RDH, MS is the graduate program director at The Ohio State University College of Dentistry Division of Dental Hygiene. Denise A. Kissell, RDH, MPH, is a Clinical Assistant Professor, Dental Hygiene, The Ohio State University College of Dentistry. Jessica Salisbury, RDH, MS, is an advisor for Press Ganey Associates, Inc.

REFERENCES

1. Evans AW, McKenna C, Oliver M. Self-assessment in medical practice. *Journal of the Royal Society of Medicine*. 2002;95(10):511-3. PubMed PMID: 12356978; PubMed Central PMCID: PMC1279183.
2. Mould MR, Bray KK, Gadbury-Amyot CC. Student self-assessment in dental hygiene education: a cornerstone of critical thinking and problem-solving. *Journal of dental education*. 2011;75(8):1061-72. PubMed PMID: 21828300.
3. American Dental Association Commission on Dental Accreditation. Accreditation Standards for Allied Dental Education Programs (Dental Hygiene). Chicago: ADA, 2013. Accessed 9-12-15, <http://www.ada.org/en/coda/current-accreditation-standards/>
4. Watts WE, Rush K, Wright M. Evaluating first-year nursing students' ability to self-assess psychomotor skills using videotape. *Nursing education perspectives*. 2009;30(4):214-9. PubMed PMID: 19753853.
5. Hawkins SC, Osborne A, Schofield SJ, Pournaras DJ, Chester JF. Improving the accuracy of self-assessment of practical clinical skills using video feedback--the importance of including benchmarks. *Medical teacher*. 2012;34(4):279-84. doi: 10.3109/0142159X.2012.658897. PubMed PMID: 22455696.
6. Corrigan R, Hardham G, Cant R, Mort JR. Use of technology to enhance student self evaluation and the value of feedback on teaching... ..including commentary by Cant R and Mort JR. *International Journal of Therapy & Rehabilitation*. 2011;18(10):579-90. PubMed PMID: 2011352847. Language: English. Entry Date: 20111202. Revision Date: 20111216. Publication Type: journal article.
7. Navickis MA, Bray KK, Overman PR, Emmons M, Hessel RF, Cowman SE. Examining clinical assessment practices in U.S. dental hygiene programs. *Journal of dental education*. 2010;74(3):297-310. PubMed PMID: 20203330.
8. Yeates P, O'Neill P, Mann K, K WE. 'You're certainly relatively competent': assessor bias due to recent experiences. *Medical education*. 2013;47(9):910-22. doi: 10.1111/medu.12254. PubMed PMID: 23931540.
9. Zick A, Granieri M, Makoul G. First-year medical students' assessment of their own communication skills: a video-based, open-ended approach. *Patient education and counseling*. 2007;68(2):161-6. doi: 10.1016/j.pec.2007.05.018. PubMed PMID: 17640843.
10. Seif GA, Brown D. Video-recorded simulated patient interactions: can they help develop clinical and communication skills in today's learning environment? *Journal of allied health*. 2013;42(2):e37-44. PubMed PMID: 23752242.
11. Gadbury-Amyot CC, Purk JH, Williams BJ, Van Ness CJ. Using tablet technology and instructional videos to enhance preclinical dental laboratory learning. *Journal of dental education*. 2014;78(2):250-8. PubMed PMID: 24489032.
12. Benson AA, Bobo LS, Green MS. Comparison of video and real-time scoring techniques. *Journal of allied health*. 2012;41(3):118-22. PubMed PMID: 22968773.
13. Albanese MA. Challenges in using rater judgments in medical education. *Journal of evaluation in clinical practice*. 2000;6(3):305-19. PubMed PMID: 11083041.
14. Yaphe J, Street S. How do examiners decide?: a qualitative study of the process of decision making in the oral examination component of the MRCGP examination. *Medical education*. 2003;37(9):764-71. PubMed PMID: 12950938.
15. Isaacson JJ, Stacy AS. Rubrics for clinical evaluation: objectifying the subjective experience. *Nurse Educ Pract*. 2009;9(2):134-40. doi: 10.1016/j.nepr.2008.10.015. PubMed PMID: 19083270.
16. Geissler PR. Student self-assessment in dental technology. *Journal of dental education*. 1973;37(9):19-21. PubMed PMID: 4517642.
17. Bloomfield JG, Jones A. Using e-learning to support clinical skills acquisition: Exploring the experiences and perceptions of graduate first-year pre-registration nursing students - A mixed method study. *Nurse education today*. 2013;33(12):1605-11. doi: DOI 10.1016/j.nedt.2013.01.024. PubMed PMID: WOS:000327002600028.
18. O'Connor AB. *Clinical Instruction and Evaluation*. 3rd ed: Jones & Bartlett Learning; 2015.