The Validity of Using E4D Compare’s “% Comparison” to Assess Crown Preparations in Preclinical Dental Education

Richard S. Callan, DMD, EdS; Van B. Haywood, DMD; Jeril R. Cooper, DMD; Alan R. Furness, DMD; Stephen W. Looney, PhD

Abstract: When a dental school is deciding which technology to introduce into a curriculum, it is important to identify the educational goals for the system. The authors’ primary goal for the use of a computer-aided resource was to offer students another way to assess their performance, to enhance their learning, and to potentially decrease their learning curve in the preclinical environment prior to using the technique in clinical patient care. The aim of this study was to examine the validity of the “% Comparison” numbers derived from the E4D Compare software program. Three practical examinations were administered to a class of 82 students at one U.S. dental school over a six-week period. The grading of the practical examinations was performed with individual faculty members being responsible for evaluating specific aspects of each preparation. A digital image of each student’s practical examination tooth was then obtained and compared to the digital image of an ideal preparation. The preparations were compared, and the “% Comparison” was recorded at five tolerance levels. Spearman’s correlation coefficient (SCC) was used to measure the agreement in rankings between the faculty scores on practical exams 1-3 and the scores obtained using E4D Compare at the different tolerance levels. The SCC values for practical exams 2 and 3 were all between 0.2 and 0.4; for practical exam 1, the SCC values ranged from 0.47 to 0.56. There was no correlation between the faculty scores and the numbers given by the “% Comparison” of the software.

Dr. Callan is Associate Professor and Chair, Department of General Dentistry, College of Dental Medicine, Georgia Regents University; Dr. Haywood is Professor, Department of Oral Rehabilitation, College of Dental Medicine, Georgia Regents University; Dr. Cooper is Associate Professor, Department of General Dentistry, College of Dental Medicine, Georgia Regents University; Dr. Furness is Assistant Professor, Department of Oral Rehabilitation, College of Dental Medicine, Georgia Regents University; and Dr. Looney is Professor, Department of Biostatistics and Epidemiology, Medical College of Georgia, Georgia Regents University. Direct correspondence to Dr. Richard S. Callan, College of Dental Medicine, Georgia Regents University, Room GC 3080, 1430 John Wesley Gilbert Drive, Augusta, GA 30912-1290; 706-721-3881; rcallan@gru.edu.

Keywords: dental education, assessment, grading, preclinical simulation, CAD/CAM

Submitted for publication 1/30/15; accepted 5/27/15

Virtually all dental schools use some type of simulation in the preclinical training of their students. Most schools train their students with dentoforms mounted on poles or with manikins with which students practice basic preparation designs on plastic teeth. Many schools include computer-aided simulation in their curricula. Manufacturers of each computer-aided system available today highlight aspects of their technology that they believe can be beneficial to the learner. The DentSim unit (DenX Ltd., Tel Aviv, Israel) offers many feedback options and the ability for self-assessment. The Simodont Dental Trainer (ACTA, Amsterdam, Netherlands) enables students to “develop their manual skills in a realistic virtual world, while getting feedback on their decisions and abilities,” thus “changing the way that students learn and the way in which they master competencies.” Romexis Compare (Planmeca USA, Roselle, IL, USA) uses “powerful analytical and reporting tools” to provide students with detailed objective feedback” with the intent of “perfecting their clinical skills.” Sirona (Salzburg, Austria) describes its prepCheck as an easy-to-learn application offering objective assessment of preparations and fillings: “The analysis software provides various interactive tools in order to assess the result compared to the predefined values or the master preparation.”

When a dental school is deciding which technology to introduce into a curriculum, it is important to identify the educational goals for the system. Our primary goal for the use of a computer-aided resource at the College of Dental Medicine, Georgia Regents University was to offer our students another way to assess their preclinical performance, to enhance their learning, and to potentially decrease their learning curve in the preclinical environment. Furthermore, we wanted to use a technology that our students would later be able to directly apply to a clinical setting. Learning how to scan, design, and fabricate...
indirect restorations in a preclinical setting with the same technology they will be using in the clinical setting is of considerable benefit. With the exception of Sirona prepCheck and E4D Compare, the use of the simulators listed above does not carry over into actual clinical application.

Prior studies have identified potential problems with the implementation of technology for educational purposes. Arnetzl and Dornhofer have commented that “the manufacturer still has much to do with regard to the user friendliness and ergonomics of the software” with regards to the PREPassistant (KaVo Dental, Biberach, Germany). In regards to the Visio-Haptic Integrated Dental Training Simulation system (Novint Technologies, Albuquerque, NM, USA), Konukseven et al. reported above average responses to the “usability, clarity, effectiveness, help/support provided, and satisfaction” from the dentists tested, but highlighted the need for “further improvements, especially in the clarity category. Some institutions, as a whole, are unreceptive to the inclusion of computer simulation in their established cultures, while others struggle to determine the correct technology for the desired application. An earlier study concluded that the full instructional potential of this technology would not be realized until it was incorporated into the school’s grading rubric. Once students realize they have the ability and opportunity to obtain an objective assessment of their performance and that the same assessment will be used in grading their work, that study postulated that they will be more likely to utilize the resource as an instructional tool. Providing ongoing, real-time feedback will ultimately help students better evaluate their own work. Students have long lamented, and many articles have been published about, the inherent shortcomings of subjective grading. The inclusion of a more objective means of evaluation has been a subject of considerable debate for quite some time.

In order to use this technology for assessment purposes, it is imperative to prove its accuracy, reliability, and validity. Previous studies have confirmed the accuracy and reliability of using computer-assisted design/computer-assisted manufacturing (CAD/CAM) technology for the assessment of crown preparations. For a system to be reliable means the assessment tool produces stable and consistent results, regardless of the number of times the assessment is made. For those results to be valid means that the score received from the computer scan would measure the preparation in the same manner as that of calibrated dental faculty. The aim of this study was therefore to examine the validity of the “% Comparison” numbers in the E4D Compare software program (Richardson, TX, USA) with regards to the overall grades given by experienced dental faculty members on three practical examinations in which second-year dental students prepared tooth #30 for a full gold crown in a preclinical environment. The E4D Compare software provides a three-dimensional comparison of the digital images of two objects. The hypothesis of the study was that when ranked from lowest to highest, the results from the conventional faculty grading would correlate within an acceptable range to the “% Comparison” numbers derived with E4D Compare.

Methods

This study received approval from the Institutional Review Board of Georgia Regents University. The study was conducted during the second-year students’ initial fixed prosthodontics course, traditionally taken during the fall term of the second year. During this course, three practical examinations (tooth #30 full gold crown preparation) were administered to the class of 82 students (48 men, 34 women) over a six-week period, with approximately two weeks practice time allotted between each of the exams. The practical exams were performed in a simulation laboratory utilizing Kilgore (Coldwater, MI, USA) typodonts (made specifically for Georgia Regents University) mounted in A-dec (Newberg, OR, USA) simulators. The students were given two hours to prepare tooth #30 to receive a full gold crown, the ideal criteria of which is described on the Faculty Evaluation Sheet (available from the corresponding author).

The practice of performing multiple practical exams on the same tooth over a relatively short period of time has been used in this course for a number of years. Once the grades are recorded, the students receive a copy of their grading form indicating which student's progress and are in need of remediation. During this course, three practical examinations were taken during the fall term of the second year. The practice was performed in a simulation laboratory utilizing Kilgore (Coldwater, MI, USA) typodonts (made specifically for Georgia Regents University) mounted in A-dec (Newberg, OR, USA) simulators. The students were given two hours to prepare tooth #30 to receive a full gold crown, the ideal criteria of which is described on the Faculty Evaluation Sheet (available from the corresponding author).
Faculty members responsible for evaluating specific criteria of each preparation in one of the four categories (occlusal reduction, proximal reduction, facial-lingual reduction, and margins and draw). These faculty members undergo calibration every year prior to the beginning of the course. The same four faculty members used the same criteria to grade all three practical exams. Each of the four criteria on the evaluation form was given an individual numerical grade from 1 to 100 based on the faculty member’s subjective evaluation of individual items under each criterion. The four criteria were weighted equally, and the individual items under each criterion were evaluated according to the degree each item deviated from the stated ideal. An average of the four criteria grades was calculated, resulting in the final grade.

Once the faculty grading was completed, a digital image of each student’s practical exam tooth was obtained and compared to the digital image of an ideal preparation (Figure 1). In order to use the auto-align feature of the E4D Compare software, the digital image of the student’s preparation must first be manually aligned with the digital image of the ideal preparation. Once these two images were closely approximated, the auto-align feature was activated, and the software completed the process. After the auto-align feature was used, a margin related to the preparation was indicated on the student’s model. In order to compare the entire tooth and not just the portion above the actual preparation margin, the comparison margin of the preparation was drawn at the juncture of the tooth and the tissue, not the actual preparation margin (Figure 2). Marking the actual preparation margin makes it impossible to assess its distance from the gingival crest, one of the criteria listed on the grading rubric. Once the models were aligned and the comparison margin indicated, the preparations were compared, and the “% Comparison” was recorded at five tolerance levels (0.1 mm, 0.2 mm, 0.3 mm, 0.4 mm, and 0.5 mm). The “% Comparison” feature of E4D Compare indicates the percentage of the surface area of the student’s preparation that is within 0.1 mm (or other tolerances 0.2 mm, 0.3 mm, 0.4 mm, 0.5 mm) of the surface.

Figure 1. Student’s tooth compared to ideal tooth: tolerance level 0.3 mm
E4D Compare score was used. The criteria proposed by Morton and McCarter were used to interpret the magnitude of the SCC: values between 0.00 and 0.20 in absolute value were classified as negligible, values between 0.20 and 0.50 as weak, values between 0.50 and 0.80 as moderate, and values between 0.80 and 1.00 as strong. Steiger’s method was used to compare the SCC values obtained for each of the three competency exams, separately for each tolerance. Since three comparisons of the SCC values were performed at each tolerance (exam 1 vs. exam 2, exam 1 vs. exam 3, exam 2 vs. exam 3), a Bonferroni adjustment was used to control the family-wise error rate at the 0.05 level for each tolerance.

**Results**

The SCC values, along with 95% confidence intervals and p-values for testing the null hypothesis that the true value of the SCC was equal to 0,
are shown in Table 2. The SCC values for practical exams 2 and 3 were all between 0.20 and 0.40, which would be classified as weak according to the Morton and McCarter criteria. For practical exam 1, the SCC values ranged from 0.47 to 0.56, which would be classified as weak to moderate. None of the upper 95% confidence limits reached the 0.8 criterion for strong. Steiger’s method with a Bonferroni adjustment indicated no significant differences between the SCC values for exam 1, as compared with the SCC values for exams 2 and 3. At best, these SCC values indicated only weak to moderate agreement in ranks between practical exam scores and scores obtained using E4D Compare.

Discussion

The statistical results of this study lead to the rejection of the hypothesis. When ranked from lowest to highest, the results from the conventional grading by the faculty did not correlate within an acceptable range to the “% Comparison” numbers derived using E4D Compare software.
One reason for this lack of correlation may be that many factors on the grading criteria form, which are thus assessed by the faculty, were not within the current capability of the technology used. These considerations included items such as damage to adjacent teeth and/or presence of dirt and debris on the articulator. If infractions such as these were evident, a student may receive a lower grade from the faculty than may be recognized by the “% Comparison” score (see numbers with double asterisks in Table 1). Since those factors are only a limited portion of each grade and since, due to the blindness of the grading process, we have no way of determining how many infractions there were of this type, it is difficult to estimate the impact of these items on the overall analysis of the data.

This study did not attempt to analyze the ability of this technology to accurately identify and appropriately measure aspects of the preparations that would be considered critical errors. Critical errors are defined as features of a preparation that would make it impossible to fabricate a clinically acceptable restoration: errors such as gross underreduction of the occlusal surface, severe undercuts along the axial walls, and thin fragile margins that could not survive the crown fabrication process. When using the tolerances described, the measurements would be differential in their impact on the preparation grade depending on where the deviations occur. For instance, to a grader, a deviation of +0.5 mm on an occlusal reduction would have a much lower deduction than a deviation of +0.5 mm or -0.5 mm on a finish line width (perhaps a critical error). The machine ascribes the same percentage penalty for both.

Another factor is the potential for differential penalties that are not present in the E4D Compare “% Comparison” feature. Errors in one area of the preparation could lead to compounding deductions in “% Comparison” that may not occur in faculty grading with the rubric. For instance, if the student prepared a margin that was 0.3 mm too wide circumferentially but had ideal taper to the axial walls, the “% Comparison” feature would recognize that all of the axial walls were outside of a 0.2 mm tolerance and score a huge drop at the 0.2 mm tolerance level. A faculty grader, using the rubric, might penalize the student for a wide margin but not for the categories of proximal reduction and facial-lingual reduction.

Previous studies confirmed the reliability of CAD/CAM technology in assessing crown preparations in a preclinical setting as well as a high inter- and intrarater reliability when using E4D Compare. These studies provide confidence that the comparisons of the student’s preparations to the ideal are accurate and reproducible. Another study confirmed the ability to compare scans taken from different dentoforms of the same make and model. Different dentoforms of the same make and model were utilized with this study, which made the process much less cumbersome and time-consuming. The “% Comparison” numbers derived in this study were both accurate and reproducible. The questions now become: How do we best utilize these results? How do we apply these numbers to provide an objective assessment to the students in order to encourage students to use the technology in self-paced learning?

A limitation of this study was that the “% Comparison” values obtained from E4D Compare were compared to the overall scores on the practical examinations and not to scores based on only the crown preparations. The structure of our current grading model did not permit a clean separation of these individual elements. Studies are currently under way to evaluate grading models that will maximize the potential of the technology (objective evaluation) while simultaneously taking advantage of the expertise of the faculty (subjective evaluation). Furthermore, since this study was undertaken in only one year at one dental school, its findings may not be generalizable to other schools or groups of students.

Conclusion

The study results did not confirm a direct correlation between a rank ordering of “% Comparison” scores and faculty grades, perhaps because the “% Comparison” feature does not assess numerous criteria that are part of the grading rubric used in faculty grading. Additional studies are necessary to determine the proper application of these computer-generated numbers and the development of an appropriate grading model that uses scanning technology.

Acknowledgments

We would like to thank Henry Schein, Inc., and E4D Technologies for the use of the Nevo scanner (version 5.0.1.6) and E4D Compare (version 2.0).

REFERENCES


