Dental Students’ Ability to Evaluate Themselves in Fixed Prosthodontics


Abstract: Self-evaluation is an essential skill for dental professionals for lifelong learning and improvement through the course of their careers. Students taking a preclinical fixed prosthodontics course were studied. The students were asked to assign themselves a grade upon completion of their timed preparation examination (teeth preparations and provisional restorations), and these were compared with grades given by the faculty. The poorer performing students tended to be less critical with their examination and to overrate their performance whereas the higher performing students were more critical of themselves and underrated their performance.

Dr. Cho is Associate Professor, Co-Director of Advanced Prosthodontics, and Predoctoral Director of Implant Dentistry, Herman Ostrow School of Dentistry, University of Southern California; Dr. Chee is Ralph and Jean Bleak Professor of Restorative Dentistry, Associate Professor, Co-Director of Advanced Prosthodontics, and Director of Implant Dentistry, Herman Ostrow School of Dentistry, University of Southern California; and Dr. Tan is Clinical Assistant Professor, Herman Ostrow School of Dentistry, University of Southern California. Direct correspondence and requests for reprints to Dr. George C. Cho, Herman Ostrow School of Dentistry, University of Southern California, 925 W. 34th Street, Room 4378, Los Angeles, CA 90089-0641; 213-740-1542; gccho@usc.edu.

Keywords: self-evaluation, dental student, preclinical project
Submitted for publication 4/21/10; accepted 6/3/10

Dental education is different from other professions because it requires that the individual understand the biology, physiology, and pathology of oral structures while also developing the skills and judgment needed to restore calcified tooth structure that has no capacity to heal. This requires that hand-to-eye coordination, spatial awareness, and the ability to visualize three-dimensional objects in fine detail are developed in the curriculum. This aspect, though mechanistic, is an important part of a clinical dentist’s career.

Most students have no difficulty with the didactic part of their dental education as it is familiar to them from their earlier educational experiences. However, the practical exercises that will prepare them to treat patients clinically are a new experience for most students. From reproducing teeth in three dimensions on a wax block to forming a provisional restoration, these exercises require hand-to-eye skill development. These skills are easier to acquire for those with good three-dimensional perception, hand and eye coordination, and some artistry.

In our experience of teaching students newly introduced to these practical exercises, students often ask their instructors to assess their projects without first self-evaluating their work. However, the ability to self-evaluate is essential for learning and improving with each exercise. Understanding the evaluation criteria, being able to visualize the ideal, and being able to evaluate against the ideal are skills that need to be developed. In our study, students taking a clinical course designed to teach them concepts of fixed prosthodontics in the anterior region of the oral cavity were sampled. This course was taught in the second year of a four-year dental curriculum. Five years of data comprising self-evaluation scores and faculty-determined scores were analyzed. The average difference between faculty- and student-determined grades were computed in order to determine if mean average differences in grade differed by grade categories, as this would indicate differences between over- or underestimation of scores.

Methodology

The faculty- and student-determined scores were calculated from preclinical examinations of tooth preparations for porcelain fused to metal anterior single unit and three-unit fixed partial denture restorations, as well as single-unit and three-unit fixed partial denture provisional restorations that were performed on artificial casts (Viade Inc., Camarillo, CA). The students were formally instructed on the grading procedures for the tooth preparations and provisional restorations and were then given grading
sheets that outlined the specific categories and ranges of acceptability/unacceptability. The supervising faculty members were also familiarized with the grading procedures and were calibrated by the course director. Calibration of faculty is of paramount importance for the evaluation process as described by Joe Moffa from the U.S. Public Health Service (USPHS) in the early 1970s. This self-evaluation scoring process had been introduced in previous preclinical courses and was familiar to the students.

This type of grading system for nonparametric parameters was originally described by Cvar and Ryge, adopted for the USPHS guidelines in 1971, and originally used to evaluate materials in non-clinical research. The grading system was further modified to adapt to evaluating clinical data with ratings of R, S, T, and V. Categories of Outline Form, Internal Form, Retention, and Cavosurface Margin Finish were defined for evaluation of the tooth preparations. Categories of Surface, Marginal Integrity, Occlusion, and Axial and Proximal Contours were defined for evaluation of the provisional restorations. In each category, the level of acceptability consisted of a grade of excellent (identified arbitrarily as an R with a value of 25), satisfactory (S valued at 20), and acceptable (M valued at 17). The level of unacceptability consisted of a grade of unacceptable (T valued at 8) and irreparable damage (V valued at 0). The lower limit grade of acceptability (M) would indicate that the treatment given would be adequate without any alteration of the treatment. The unacceptability grade of T indicates that some alteration to the treatment is necessary in order to continue with treatment, and V indicates irreparable damage has occurred and some significant changes to the treatment are necessary. A grade of M or T indicates an assessment that falls between acceptability and unacceptability. The two faculty members’ grades must be in agreement within the same category. The grading value of R, S, M, T, and V letters were also designed so that discussion of treatment evaluation could occur in the presence of a patient, without allowing the patient to discern the quality of the grade being assigned.

After each tooth preparation or provisional restoration examination had been completed, the dental students had ten minutes to evaluate their examination based on the criteria for tooth preparation and provisional restorations and generate a student-determined grade. The faculty members would then evaluate the examination independent of what the students had scored for themselves and provide a faculty-determined grade. Each examination was identified by student identification number; however, the faculty members were not aware of the students’ and their identification number to help eliminate any bias.

The faculty had been standardized by the course director for the evaluation of each category. Since there were four categories for evaluation, eight faculty members (one pair each) were responsible for evaluating that specific category for the entire class to maintain standardization throughout the class. Each pair of faculty members graded the specific category independently of one another. However, if the grade differed by more than one grade level in any category, the two faculty members would have to consult with one another until they reached an agreement of not more than one grade level difference (i.e., one faculty member would have to come down one grade level or the other would have to come up one grade level). In addition, if one faculty member graded criteria as an acceptable range and the other faculty member graded that same category as an unacceptable range, then the two faculty members had to again consult one another until they reached an agreement to either be acceptable or unacceptable in that category.

The self-evaluation score was calculated by a formula between the faculty-determined grade and student-determined grade, which was then computed towards the final grade. This was the incentive for the students to attempt to accurately assess themselves according to the grading criteria. The self-evaluation grade was determined by a correlation between the faculty-determined grade and student-determined grade. If a student had evaluated a category exactly the same as the faculty grade in that same category, a score of 25 would be awarded. If the student-determined grade differed from the faculty-determined grade by one level of acceptability, then a score of 19 was given. If the score differed by two levels of acceptability, then a score of 13 was given. If the score differed between acceptable M and unacceptable T, a value of 0 was given. For each evaluation there were four categories, giving a possible maximum score of 100 and a possible minimum score of zero. This self-evaluation score was weighted 10 percent of the overall course grade.

Table 1 illustrates how the self-evaluation score was computed between the faculty-determined grade and student-determined grade for a tooth preparation examination. The faculty-determined grade value was a score of 70, while the student-determined grade value was a score of 82. The self-evaluation
score was calculated to be 57, which meant that
the student-determined grade was not similar to
the faculty-determined grade. In fact, the student-
determined grade value indicated that the student
thought he or she had performed better than the
faculty's evaluation.

The ability of students to accurately evaluate
tooth preparations or provisional restorations would
give the dental student a higher student self-evalua-
tion score, thereby improving his or her overall course
average. Even if the students performed poorly during
the examination and gave themselves all unacceptable
grades T for their student-determined grades, if this
corresponded correctly with the faculty-determined
grades of all unacceptable grades T, then the score for
self-evaluation would be 100 since the faculty grades
and the student self-evaluation grades were the same
T for all categories, which would give a value of 25
x 4=100. The faculty-determined grade if given T
in all four categories would be a value of 8 x 4=32.

Average faculty-determined, student-deter-
mined, and self-evaluation scores were determined
for each student based on the number of exams
completed (generally seven exams in 1996 and
Students received grades of A, B, C, D, or F in the
course. Grades of D and F were uncommon and
were combined for these analyses. Mean scores by
letter grade were calculated, and analysis of covari-
cance was used to determine if mean scores adjusted
for year were significantly different by letter grade category.

In addition to letter grade categories, faculty
score categories were determined for each student
by averaging his or her faculty-determined scores
and categorizing them according to quartiles (66.7
and below, 66.8 to 71.8, 71.9 to 76.7, and 76.8 and
above). Mean scores by quartile were calculated, and
analysis of covariance was used to determine if mean
scores, adjusted for year, were significantly different
by quartile.

A general linear mixed model was used to
determine if student self-evaluation scores improved
over time. Since self-evaluation may improve at
differing rates for stronger or weaker students, this
model was adjusted for student letter grade. All data
were analyzed using SAS, Version 9.00.

### Results

Table 2 shows average faculty-determined,
student-determined, and self-evaluation scores and
mean differences between faculty- and student-deter-
dined scores by letter grade category. The faculty-
determined, student-determined, and self-evaluation
scores differ significantly (p<.0001) by letter grade
category with higher scores associated with higher
letter grades. The difference between faculty- and
student-determined scores also differs significantly
(p<.0001) by letter grade category, with A students
more likely to underestimate their work, and B, C,
and D/F students increasingly likely to overestimate
their work.

<table>
<thead>
<tr>
<th>Student-Determined Grade (Value)</th>
<th>Faculty-Determined Grade (Value)</th>
<th>Self-Evaluation Score</th>
<th>Corresponding Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outline R (25)</td>
<td>M (17)</td>
<td>13</td>
<td>Two-level evaluation difference</td>
</tr>
<tr>
<td>Internal S (20)</td>
<td>S (20)</td>
<td>25</td>
<td>Same-level evaluation</td>
</tr>
<tr>
<td>Retention S (20)</td>
<td>R (25)</td>
<td>19</td>
<td>One-level evaluation difference</td>
</tr>
<tr>
<td>Cavo finish M (17)</td>
<td>T (8)</td>
<td>0</td>
<td>Acceptable/unacceptable evaluation difference</td>
</tr>
<tr>
<td>Total score 82</td>
<td>70</td>
<td>57</td>
<td></td>
</tr>
</tbody>
</table>

Grading Scale:

- ACCEPTABLE: R, S
- M
- T
- UNACCEPTABLE: V

### Table 1. Self-evaluation score between student-determined grade, faculty-determined grade, and corresponding self-evaluation score for a tooth preparation examination
Table 2. Average scores by letter grade: faculty-determined, student-determined, and self-evaluation scores and mean differences between faculty and student-determined scores

<table>
<thead>
<tr>
<th>Letter Grade</th>
<th>N</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
<th>Mean</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>105</td>
<td>78.1</td>
<td>89.9</td>
<td>58.9</td>
<td>77.9</td>
<td>89.7</td>
<td>64.0</td>
<td>1.13</td>
<td>77.4</td>
<td>89.3</td>
<td>49.3</td>
</tr>
<tr>
<td>B</td>
<td>240</td>
<td>73.1</td>
<td>85.8</td>
<td>55.5</td>
<td>73.7</td>
<td>85.7</td>
<td>53.7</td>
<td>0.60</td>
<td>73.8</td>
<td>88.4</td>
<td>49.9</td>
</tr>
<tr>
<td>C</td>
<td>218</td>
<td>68.5</td>
<td>79.5</td>
<td>53.8</td>
<td>71.8</td>
<td>91.0</td>
<td>58.0</td>
<td>3.32</td>
<td>70.3</td>
<td>85.6</td>
<td>50.7</td>
</tr>
<tr>
<td>D/F</td>
<td>47</td>
<td>61.6</td>
<td>85.0</td>
<td>47.7</td>
<td>67.3</td>
<td>83.8</td>
<td>50.4</td>
<td>5.73</td>
<td>66.5</td>
<td>86.5</td>
<td>46.8</td>
</tr>
</tbody>
</table>

p-value: <.0001

Table 3. Average scores by letter grade: faculty-determined, student-determined, and self-evaluation scores and mean differences between faculty and student-determined scores

<table>
<thead>
<tr>
<th>Letter Grade</th>
<th>N</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
<th>Mean</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>105</td>
<td>78.1</td>
<td>89.9</td>
<td>58.9</td>
<td>77.9</td>
<td>89.7</td>
<td>64.0</td>
<td>1.13</td>
<td>77.4</td>
<td>89.3</td>
<td>49.3</td>
</tr>
<tr>
<td>B</td>
<td>240</td>
<td>73.1</td>
<td>85.8</td>
<td>55.5</td>
<td>73.7</td>
<td>85.7</td>
<td>53.7</td>
<td>0.60</td>
<td>73.8</td>
<td>88.4</td>
<td>49.9</td>
</tr>
<tr>
<td>C</td>
<td>218</td>
<td>68.5</td>
<td>79.5</td>
<td>53.8</td>
<td>71.8</td>
<td>91.0</td>
<td>58.0</td>
<td>3.32</td>
<td>70.3</td>
<td>85.6</td>
<td>50.7</td>
</tr>
<tr>
<td>D/F</td>
<td>47</td>
<td>61.6</td>
<td>85.0</td>
<td>47.7</td>
<td>67.3</td>
<td>83.8</td>
<td>50.4</td>
<td>5.73</td>
<td>66.5</td>
<td>86.5</td>
<td>46.8</td>
</tr>
</tbody>
</table>

p-value: <.0001

Table 3. Average scores by faculty score quartiles: faculty-determined, student-determined, and self-evaluation scores and mean differences between faculty- and student-determined scores

<table>
<thead>
<tr>
<th>Quartile</th>
<th>N</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
<th>Mean</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥76.8</td>
<td>153</td>
<td>80.1</td>
<td>89.9</td>
<td>76.7</td>
<td>78.8</td>
<td>91.0</td>
<td>66.0</td>
<td>1.35</td>
<td>78.7</td>
<td>89.3</td>
<td>65.4</td>
</tr>
<tr>
<td>71.9 to 76.7</td>
<td>152</td>
<td>74.1</td>
<td>76.7</td>
<td>71.9</td>
<td>74.8</td>
<td>85.9</td>
<td>59.0</td>
<td>0.65</td>
<td>74.2</td>
<td>86.4</td>
<td>51.7</td>
</tr>
<tr>
<td>66.8 to 71.8</td>
<td>154</td>
<td>69.2</td>
<td>71.8</td>
<td>66.7</td>
<td>72.0</td>
<td>82.0</td>
<td>60.9</td>
<td>2.79</td>
<td>71.0</td>
<td>82.0</td>
<td>55.0</td>
</tr>
<tr>
<td>≤66.7</td>
<td>151</td>
<td>62.1</td>
<td>66.7</td>
<td>47.7</td>
<td>67.4</td>
<td>79.3</td>
<td>50.4</td>
<td>5.31</td>
<td>66.4</td>
<td>80.1</td>
<td>46.8</td>
</tr>
</tbody>
</table>

p-value: <.0001

Discussion

Friedson has stated that the most distinguishing characteristic of a profession is its ability for self-regulating “legitimate autonomy.” This autonomy involves members being able to honestly and competently assess themselves. Accurate self-assessment is a skill requiring deliberate training and practice.

Gordon noted that the ability to accurately assess one’s own strengths and weaknesses is fundamental to self-directed lifelong learning and to continued competence in the health professions. He also noted that although health profession trainees do acquire professional behaviors, they do not appear to be able to accurately apply to themselves the evaluative criteria and standards of their supervisors.

From the results of this study, it appeared that poorer performing students (lower half of the class) were unable to realistically evaluate their own projects as they perceived their work to have met the grading criteria. Better performing students (upper half of the class) tended to be more critical of themselves and therefore would continue to practice and work for self-improvement. These findings are consistent with the published literature. Gordon concluded that superior students would persist in the belief that they were inferior performers, whereas inferior performers would persist in the belief that they were superior.

Woolliscroft et al. studied third-year medical stu-
dents in an internal medicine clerkship and found that the bottom quartile of students consistently rated their knowledge and skills highest whereas the top quartile of students rated themselves lower than would be anticipated. This was despite objective feedback to the contrary from multiple sources. Edwards et al.\textsuperscript{9} came to a similar conclusion sampling medical students in an obstetrics and gynecology clerkship. Arnold et al.\textsuperscript{10} presented similar findings. There was a tendency for conservative self-evaluation amongst the higher performing students in their study, and they interpreted this to mean that the higher achievers held themselves to more stringent standards and may be assessing themselves against their own potential rather than the performance of their peers. Coutts and Rogers\textsuperscript{11} speculated that there is a benefit to this underestimation as students who underestimate their own capabilities would be motivated to strive harder and ultimately become higher achievers. Conversely, a tendency towards overestimation could be detrimental for learners, as students who overestimate their performance ultimately performed less well compared to their peers. Edwards et al.\textsuperscript{9} suggested that below-average students had the potential to derive the most benefit from more formal feedback.

Self-directed learning is based on the assumption that adult learners can identify and remedy deficits in their knowledge and skills.\textsuperscript{10} High performers have the ability to recalibrate their self-assessments more accurately when presented with benchmarks and feedback. However, even if this feedback is provided, the individual must possess the capacity to process this feedback and come to an accurate understanding of why the failure had occurred.\textsuperscript{11,12} Incompetent individuals not only perform poorly, but are unable to recognize that they have performed poorly. In fact, Kruger and Dunning\textsuperscript{13,14} noted that participants in the bottom percentile not only overestimated themselves, but thought they were above average. Poor performers showed lesser metacognitive skills than top performers, and these deficiencies mediated the link between performance and miscalibration. They noted that it was paradoxical that the only way to make individuals recognize their incompetence was to make them competent and that competence would lead to calibration. Erhlinger et al.\textsuperscript{15} confirmed this, noting that poor performers demonstrated little insight into the depth of their deficiencies relative to their peers and that these poor performers did not seem to know how poorly they were doing. These findings could explain the observations of Koper,\textsuperscript{16} who felt that the failure of prosthetic service could be attributed to the failure of the dentist to understand his or her own limitations, due to insufficient professional skill. This would also account for the findings of Milgrom et al.,\textsuperscript{17} who noted that less experienced dentists tended to rate themselves higher than experienced colleagues. Felson\textsuperscript{18} postulated that individuals may evaluate themselves favorably to maintain their self-esteem and that these individuals are subsequently likely to see what they want or expect to see, based on their self-esteem.

Self-representation influences how an individual responds to and processes information about himself or herself.\textsuperscript{19} Some students have well-developed internal representations of their abilities that do not meet the reality of their performance. Over time, these self-representations may become increasingly resistant to change when feedback that is inconsistent or contradictory to the self-representation is obtained.\textsuperscript{19} Curtis et al.\textsuperscript{20} suggested that creating an environment in which students are encouraged to understand their mistakes is an integral part of the educational process and that this could lead to improved self-assessment skills that would then lead to improved learning. Abrams and Kelly\textsuperscript{21} felt that incorporating a student self-evaluation situation in a pediatric operative dentistry technique course might encourage decision making and critical analysis.

### Conclusions

A self-evaluation exercise was conducted as part of a preclinical fixed prosthodontic course. The results of the exercise showed that poorer performing students tended to overrate their performance. Conversely, the higher performing students were more critical of themselves and underrated their performance. The results were consistent with similar

<table>
<thead>
<tr>
<th>Test Number</th>
<th>N</th>
<th>Mean Self-Evaluation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>609</td>
<td>73.7</td>
</tr>
<tr>
<td>2</td>
<td>610</td>
<td>75.2</td>
</tr>
<tr>
<td>3</td>
<td>609</td>
<td>68.7</td>
</tr>
<tr>
<td>4</td>
<td>610</td>
<td>77.1</td>
</tr>
<tr>
<td>5</td>
<td>610</td>
<td>73.7</td>
</tr>
<tr>
<td>6</td>
<td>610</td>
<td>69.6</td>
</tr>
<tr>
<td>7</td>
<td>608</td>
<td>71.6</td>
</tr>
<tr>
<td>8</td>
<td>361</td>
<td>71.1</td>
</tr>
<tr>
<td>9</td>
<td>361</td>
<td>69.3</td>
</tr>
</tbody>
</table>
In addition, it was found that the students did not improve in their ability to self-evaluate their work. The ability to carry out accurate self-evaluation and self-critique is necessary in any profession, such as dentistry where lifelong learning is essential. Perhaps the largest paradox, as stated by Kruger and Dunning, is that the way to calibrate an individual’s self-evaluation, enabling those less competent to recognize their inadequacies and accurately evaluate themselves, is to educate them and make them competent and that this competence would lead to calibration and accurate self-assessment.

**Acknowledgments**

The authors extend special thanks to Dr. Tony Gaurino for providing the statistical analysis for this article.

**REFERENCES**