Short-Answer Questions and Formula Scoring Separately Enhance Dental Student Academic Performance

R. Neal Pinckard, Ph.D.; C. Alex McMahan, Ph.D.; Thomas J. Prihoda, Ph.D.; John H. Littlefield, Ph.D.; Anne Cale Jones, D.D.S.

Abstract: In this study, numerical course scores of second-year dental students in four successive classes in an oral and maxillofacial pathology course were compared. While the course content and teaching methods were essentially unchanged throughout the four years, two modest departures from the sole use of multiple-choice format questions were made in the assessment of student achievements. The modifications consisted of creating a more challenging examination procedure through the inclusion of un-cued short-answer format questions and the institution of correction-for-guessing scoring on multiple-choice examinations. Academically, the students in the four classes were comparable, as indicated by their respective numerical course score distributions in a prerequisite general pathology course in which the course content was unchanged, and all multiple-choice format questions were used to assess student academic achievements. This four-year study demonstrated that two qualitative changes in the educational environment—utilization of un-cued short-answer questions and correction for guessing scoring of multiple-choice questions—separately resulted in significant improvements in student course scores. Our results support the notion that, without any changes in curricular content or emphasis, combinations of qualitative changes in the assessment procedures alter student behavior and, as a consequence, appreciably improve their academic achievements.

Keywords: dental education, educational measurement, educational methodology, assessment, aptitude-treatment interaction, short-answer questions, multiple-choice questions, formula scoring, oral and maxillofacial pathology

Submitted for publication 3/10/11; accepted 9/23/11

The long-term goals of our educational research program are to identify aspects of the educational environment that foster improved student learning behaviors and then to implement appropriate modifications to enhance students’ academic achievements. This research was conducted in an oral and maxillofacial pathology course for second-year dental students at the University of Texas Health Science Center at San Antonio (UTHSCSA). Initially, we concentrated on modifying two of the many aspects of the educational environment: the examination question format and the scoring algorithm for multiple-choice examinations. We first observed that, by retrospectively rescoring multiple-choice examinations using a correction-for-guessing algorithm, there was significantly better agreement of these corrected scores with the scores on short-answer examinations (50 percent of the examination) testing the same body of knowledge.¹ Our interpretation of this improved agreement was that correction for guessing increased the validity of multiple-choice examination scores. In a follow-up study, we documented the anticipated increased validity of the multiple-choice examination scores when correction for guessing was implemented prospectively (the students knew that their multiple-choice examinations would be scored with correction for guessing).² Moreover, we observed significant improvements in student scores on both the multiple-choice and short-answer portions of the examinations by lower achieving students as identified by their previous course scores in a prerequisite general pathology course. Thus, this one alteration in the academic environment had a dramatic effect on student course scores.² Subsequently, we found that the utilization of short-answer format questions in oral and maxillofacial pathology also significantly improved student scores on the multiple-choice portion (50 percent of
course score) of examinations. However, in this case the improved scores were observed in students at all levels of previous academic achievement. Of note, we also observed that student performance was not further enhanced when 100 percent short-answer examinations were used instead of only 50 percent short-answer examinations.

The findings and implications of the preceding studies made it imperative that we conduct additional statistical analyses to investigate the combined effects of utilizing short-answer examinations and correction-for-guessing scoring of multiple-choice examinations on student academic achievements. The analyses confirmed and extended our previous observations that a modest qualitative change in the educational environment, using 50 percent short-answer questions, was associated with improved student performance in oral and maxillofacial pathology. However, a further quantitative increase to 100 percent short-answer questions did not result in additional improvements. These analyses also confirmed that a second qualitative change in the assessment procedure—implementation of correction-for-guessing scoring on multiple-choice examinations—was associated with significantly improved student performance. Most importantly, these analyses documented that, relative to the use of all multiple-choice examinations, employing these two qualitatively different changes in assessment separately enhanced student achievements. The collective results support the notion that combinations of appropriate changes in the educational environment will promote beneficial changes in student learning behavior and, as a consequence, significantly improve their academic achievements.

### Methods

Student scores in an oral and maxillofacial pathology course for second-year dental students were compared for four classes (2004–05 to 2007–08). The educational environment including course content, clinical cases, and lecturers was essentially unchanged over the four years. However, there were differences in the question formats (short-answer [SA] or multiple-choice [MC]) and in the multiple-choice examination scoring algorithm (number-correct or standard correction for guessing [MC*]). (See Table 1 for a summary of question formats, scoring methods, and abbreviations used to identify the classes.) We previously detailed the history and reasons for changing examination procedures. Course scores in the prerequisite general pathology course were used to determine whether the four classes were academically comparable and to classify students differing in their previous achievements.

The general pathology course at the UTHSCSA Dental School is presented in the fall semester of the second year and is a prerequisite for the oral and maxillofacial pathology course in the following spring semester. The oral and maxillofacial pathology course consisted of fifty hours of didactic lectures and four two-hour examinations. Students were informed of examination procedures, scoring procedures, and calculation of final course scores in the written syllabus as well as orally by the course director in the first class. Students were told explicitly, in the syllabus and during the first class, that they were expected to learn and understand the following characteristics for each of the pathologic processes discussed in the course: etiology, pathogenesis, age

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>Number of Students</th>
<th>General Pathology</th>
<th>Oral and Maxillofacial Pathology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fraction of Course Score</td>
<td>Correction for Guessing</td>
</tr>
<tr>
<td>2004–05 (SA/MC)</td>
<td>88</td>
<td>100%</td>
<td>Retrospective†</td>
</tr>
<tr>
<td>2005–06 (SA/MC*)</td>
<td>83</td>
<td>100%</td>
<td>Retrospective†</td>
</tr>
<tr>
<td>2006–07 (MC)</td>
<td>88</td>
<td>100%</td>
<td>Retrospective†</td>
</tr>
<tr>
<td>2007–08 (SA)</td>
<td>86</td>
<td>100%</td>
<td>Retrospective†</td>
</tr>
</tbody>
</table>

*Correction for random guessing applied prospectively.
†Correction for random guessing not used in computation of official grades; number-correct scoring used for official grades.
SA=short-answer questions; MC=multiple-choice questions; MC*=multiple-choice questions with prospective correction for guessing.
and sex predilection, most common anatomic location, distinguishing features (clinical, radiographic, microscopic), diagnostic aids and laboratory tests, treatment options, and prognosis. Moreover, students were told explicitly they would be examined only on these characteristics. Both the short-answer and multiple-choice examinations were constructed to test students’ knowledge and understanding of these specific disease characteristics. For example, a typical test question might be posed as follows: “A twenty-two-year-old male presents with a two-day history of ulcerations involving the labial mucosa, dorsal tongue, and vermilion surface of the lips; circular, erythematous skin lesions are also present. The patient states that the week before he had an outbreak of herpes labialis. What is the most likely clinical diagnosis?” In the short-answer format, the answer should be “erythema multiforme.” In the multiple-choice format, five options were given: “A. Mucous membrane pemphigoid, B. Erythema multiforme, C. Erythema migrans, D. Lupus erythematosus, E. Lichen planus.”

Each of the four examinations in the oral and maxillofacial pathology course covered between eleven and thirteen hours of lecture material. The questions on both short-answer and multiple-choice examinations were equally weighted to the topics that were presented prior to each of the four examinations. A disease process was tested only in the short-answer component or the multiple-choice portion. This approach ensured that the examinations covered a broad range of topics and that a given topic was not stressed more than another.

In the 2004–05 (SA/MC) and 2005–06 (SA/MC*) classes, each of the four examinations was divided into two one-hour examinations. The first hour of each was based on twenty-five clinical cases, each of which consisted of a brief written clinical history and projected gross, microscopic, and/or radiographic findings. Students were advised to respond succinctly to the two short-answer questions for each case; appropriate answers typically consisted of one or more sentences or key words. The course director graded all short-answer examinations by identifying key words delineated when the examination was constructed. Points were not deducted for spelling errors, as long as words were phonetically correct. If a student gave several answers, only the first was evaluated; no partial credit was awarded. The short-answer examinations were collected at the conclusion of this portion. The second hour of each examination in the 2004–05 (SA/MC) and 2005–06 (SA/MC*) classes consisted of fifty multiple-choice questions, each of which had only one correct answer and four plausible distractors. At the end of the second hour, the answer sheets were collected and graded electronically.

In the 2005–06 (SA/MC*) class, the students were informed in the written syllabus and orally by the course director during the first class that a scoring procedure employing a correction for random (no knowledge) guessing would be applied to the multiple-choice examinations. For other classes of the oral and maxillofacial pathology course and for the general pathology course, the correction for guessing was applied retrospectively and thus had no effect on student behavior.

In the 2006–07 (MC) class, each of the four examinations was comprised of seventy-five multiple-choice questions of construction similar to that described for the 2004–05 (SA/MC) and 2005–06 (SA/MC*) classes. Similarly, in the 2007–08 (SA) class, each of the four examinations was comprised of seventy-five short-answer questions similar to those described for the 2004–05 (SA/MC) and 2005–06 (SA/MC*) classes. The additional multiple-choice questions required in 2006–07 (MC) were developed from the short-answer questions used in 2004–05 (SA/MC) and 2005–06 (SA/MC*) to test the same topics. Similarly, the additional short-answer questions required in 2007–08 (SA) were developed from previously used multiple-choice questions. The clinical cases included in the examinations in 2006–07 (MC) and 2007–08 (SA) were the same as the cases used in the examinations in 2004–05 (SA/MC) and 2005–06 (SA/MC*).

The general pathology course and its examination methods were the same for all academic years included in this report. The course consisted of sixty-one hours of didactic lectures, four two-hour review sessions, and four two-hour examinations. Each of the examinations consisted of seventy-five multiple-choice questions with one correct answer and four distractors; test construction was similar to that described for the oral and maxillofacial pathology course. Each of the four examinations covered between thirteen and nineteen hours of lecture material. The multiple-choice questions covered information presented in the lectures and reading assignments in the period immediately preceding each examination. When the multiple-choice examinations were constructed, the questions were equally weighted to the topics presented in the lectures prior to each of the four examinations. Examination questions were
selected from a large, secure question bank. Examination scores in the prerequisite general pathology course were used to determine whether the four classes were academically comparable.

A comprehensive final examination was not given in either the general pathology course or the oral and maxillofacial pathology course. Students received a final course grade in both courses based on the averages calculated from the four two-hour examinations; that is, each two-hour examination comprised 25 percent of the final course grade. Examinations were not returned to students in either the general pathology course or the oral and maxillofacial pathology course. The examinations were stored securely, and students were allowed and encouraged to review their graded examinations under supervision.

The reliabilities of the examinations were measured by the Cronbach’s alpha statistic in both the general pathology and oral and maxillofacial pathology courses (Table 2). Carmines and Zeller describe the Cronbach’s alpha statistic as an estimate of the expected correlation between one test and a hypothetical alternative form containing the same number of items. All reliabilities were greater than 0.80, the desirable level advocated by Carmines and Zeller.

The correction for guessing formula utilized in these studies was a modification of a common scoring method for multiple-choice examinations (number-correct or number-right scoring) in which no point is assigned for an incorrect answer and +1 point (full credit) is given for a correct answer. In the multiple-choice examinations that we investigated, each multiple-choice question had five possible answers. Therefore, the standard formula for correction for guessing consisted of awarding -¼ point for an incorrect answer, 0 for a question not answered, and +1 point for a correct answer. Assuming a random selection, the probability of guessing the correct answer was 1/5 (0.20), and the probability of guessing an incorrect answer was 4/5 (0.80). Therefore, using the standard correction for guessing, the expected value of the number of points gained due to random guessing was zero [(0.20)(1)+(0.80)(-¼)=0].

We retrospectively applied the standard correction for guessing formula to rescore the multiple-choice examination scores from the 2004–05 (SA/MC) and 2006–07 (MC) classes in the oral and maxillofacial pathology course. This not only ensured comparability of multiple-choice examination scores with those scores from the 2005–06 (SA/MC*) class in which prospective correction for guessing was used but also improved validity of the scores from the multiple-choice examinations. We also retrospectively applied the correction for guessing to the scores from general pathology; this improved the validity of the multiple-choice examinations in general pathology and provided the same scale for comparison with the subsequent performance in oral and maxillofacial pathology. The retrospective application of the correction for guessing represented application of a linear transformation because students answered all multiple-choice examination questions. Such a linear transformation of the data does not affect the significance levels. We analyzed scores by students completing all four examinations in oral and maxillofacial pathology and all four examinations in general pathology. This study was approved by the Institutional Review Board of UTHSCSA.

Mean numerical scores of students in an academic class were compared between academic classes using analysis of variance, and standard deviations (variances) were compared using Levene’s test. When variances were unequal, Satterthwaite’s modification was used to compare means. Cumulative relative frequency distributions were used to graphically show distributions of numerical scores.

Table 2. Cronbach’s alpha statistics for examinations in the general pathology and oral and maxillofacial pathology courses, by academic year

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>General Pathology Multiple-Choice</th>
<th>Oral and Maxillofacial Pathology Short-Answer</th>
<th>Oral and Maxillofacial Pathology Multiple-Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006–07 (MC)</td>
<td>0.92</td>
<td>—</td>
<td>0.93</td>
</tr>
<tr>
<td>2004–05 (SA/MC)</td>
<td>0.91</td>
<td>0.89†</td>
<td>0.89†</td>
</tr>
<tr>
<td>2007–08 (SA)</td>
<td>0.92</td>
<td>0.94</td>
<td>—</td>
</tr>
<tr>
<td>2005–06 (SA/MC*)</td>
<td>0.87</td>
<td>0.87†</td>
<td>0.87†</td>
</tr>
</tbody>
</table>

*Correction for random guessing applied prospectively.
of individual students among the four academic classes in the general pathology course and the oral and maxillofacial pathology course. Students were placed into one of three grade categories based on their numerical course score (80–100 percent, 70–79 percent [the precise definition of this category was 70 percent to <80 percent], and <70 percent). Frequencies of students in grade categories were analyzed using the chi-square test for independence. When low expected frequencies were encountered, an exact procedure was used to obtain the significance level (p-value). To investigate Aptitude-Treatment Interactions, we classified students’ achievement based on their previous grade category in general pathology (which we utilized as an estimate of student aptitude) and then analyzed their subsequent numerical scores in oral and maxillofacial pathology by these classifications and class year in a two-way analysis of variance.

The relationship between the numerical scores of students in oral and maxillofacial pathology and their respective numerical scores in general pathology were examined by using principal component lines estimated from the variance-covariance matrix. The first principal component is the linear combination (straight line) of the variables that has the maximum variance among all (normalized) linear combinations. Also, the first principal component is the line through the means \((\bar{X}, \bar{Y})\), which minimizes the sum of the squared perpendicular distances of the data points to the line. The first principal component line more accurately estimated the linear relationship between these \(X\) and \(Y\) variables than would ordinary linear regression analysis. The improved accuracy from the use of principal component analysis was because both the \(X\) and \(Y\) variables were random variables; in ordinary linear regression analysis, only the \(Y\) variable is considered to be a random variable and the estimator of the line is biased if \(X\) also is a random variable.

The equation of the principal component line that we used was \(Y = \beta_{70} + \beta(X - 70)\) where \(Y\) is the score in oral and maxillofacial pathology, \(X\) is the score in general pathology, \(\beta\) is the slope, and \(\beta_{70}\) is the value of \(Y\) at \(X=70\). The slope (\(\beta\)) of the principal component line describes the predicted change in the score in oral and maxillofacial pathology if the score in general pathology were increased by one point. The intercept parameter (\(\beta_{70}\)) is the score in oral and maxillofacial pathology predicted from the principal component line for a general pathology score of 70 percent. We present the value of \(Y\) for \(X=70\) instead of the usual intercept parameter that represents the value of \(Y\) for \(X=0\). The usual \(Y\) intercept represents an extrapolation to values that are not observed in these data, whereas the predicted value we present is well within the range of the observed data. We believe that this presentation is more meaningful to educators because it enables us to show the effects of the examination format and scoring procedure on marginal students, who in this study were students scoring a 70 percent (our minimum acceptable score). Thus, the intercept parameter (\(\beta_{70}\)) quantifies the effects of changes in the assessment procedures on the lower achieving students. A principal component line with \(\beta=1\) and \(\beta_{70}=70\) corresponds to the line of equality.

A bootstrap procedure with 1,000 samples was used to estimate confidence intervals for the parameters of the first principal component lines. The bootstrap is a nonparametric procedure and thus does not depend on any particular probability distribution, such as the normal distribution. The statistic of interest is calculated in bootstrap samples, of the same size as the original, that are generated by random sampling with replacement from the original data. If the resampling is repeated a large number of times, the empirical distribution of the statistic generated from many bootstrap samples approximates the actual distribution.

We estimated the effect of regression toward the mean on the differences between scores in oral and maxillofacial pathology and general pathology. Regression toward the mean is the tendency for a variable that is extreme on its first measurement to be closer to the center of the distribution at a later measurement. In order to distinguish between the effects of the interventions and the effect of regression toward the mean, we estimated the magnitude of the regression toward the mean using a statistical model. In this model a student’s score was comprised of a combination of a true knowledge score and a random error component. Based on the statistical analysis of student scores in general pathology and oral and maxillofacial pathology in the 2006–07 (MC) class, where both courses used only multiple-choice examinations, the true knowledge component of the model for a student was identical in the two courses. Thus, the differences between the scores in the two courses represented only random within-student variation. If the true knowledge score and the random variation component have independent normal distributions, then the course scores in general pathology and oral and maxillofacial pathology jointly have a bivariate normal distribution. Using these assumptions and
the parameter estimates computed from data in the 2006–07 (MC) class, we calculated the expected regression toward the mean.\textsuperscript{14-16}

Statistical significance was defined as $p \leq 0.05$. All calculations were carried out using SAS 9.1.3 (SAS Institute, Cary, NC).

## Results

### Overall Course Scores

The cumulative frequency distributions for numerical course scores during four successive years in the general pathology course for second-year dental students are shown in Figure 1A. There were no significant differences in mean scores among the four classes in the prerequisite general pathology course taken in the semester immediately preceding the oral and maxillofacial pathology course ($p=0.2334$, Table 3) nor in the course grade category (80–100 percent, 70–79 percent, <70 percent) distributions ($p=0.3290$). Thus, these four classes were academically comparable.

In contrast to general pathology, there were significant differences in the frequency distributions of the numerical course scores for these same students in oral and maxillofacial pathology (Figure 1B). The mean course scores in oral and maxillofacial pathology for the 2004–05 (SA/MC), 2005–06 (SA/MC\textsuperscript{*}), and 2007–08 (SA) classes were significantly higher than the mean course score of the 2006–07 (MC) class (Table 3). Thus, changes in the examination question format alone or combined with changes in the scoring procedure for multiple-choice examinations resulted in significant improvements in student course scores relative to examinations using all multiple-choice questions that were scored using number-correct scoring.

The relationships between numerical course scores in oral and maxillofacial pathology and course scores in general pathology are shown in Figure 2 for the four academic classes; the estimated parameters of the principal component lines are given in Table 3. The slope of the principal component line for the 2006–07 (MC) class was not significantly different from 1.00 (Table 3), the slope of the line of equality. In addition, the 70.3 score in oral and maxillofacial pathology predicted for a 70 percent score in general pathology was not significantly higher than a 70 percent score in the line of equality (Table 3). This similarity of the principal component line to the line

---

### Figure 1. Cumulative frequency distributions of student scores in the general pathology and oral and maxillofacial pathology courses, by academic year

**Note:** Cumulative frequency distributions of student course scores in the general pathology course, which utilized all multiple-choice examinations, appear in Panel A. Cumulative frequency distributions of student course scores in the oral and maxillofacial pathology course appear in Panel B. All multiple-choice examination scores have been corrected for random guessing. Cumulative frequency distributions for the general pathology course in 2004–05 (SA/MC) and 2005–06 (SA/MC\textsuperscript{*}) reported in Prihoda TJ, Pinckard RN, McMahan CA, Littlefield JH, Jones AC. Prospective implementation of correction for guessing in oral and maxillofacial pathology multiple-choice examinations: did student performance improve? J Dent Educ 2008;72(10):1149–59.

- ○ 2004–05 (SA/MC) class
- ▲ 2005–06 (SA/MC\textsuperscript{*}) class
- ● 2006–07 (MC) class
- △ 2007–08 (SA) class

SA=short-answer questions; MC=multiple-choice questions; MC\textsuperscript{*}=multiple-choice questions with prospective correction for guessing
of equality indicated that students of all abilities, when tested solely with multiple-choice examinations in oral and maxillofacial pathology, performed in the same manner as they did in general pathology where assessment of academic achievement was also all multiple-choice examinations.

When the examinations in oral and maxillofacial pathology were comprised of 50 percent short-answer questions and 50 percent multiple-choice questions (2004–05 [SA/MC]) or all short-answer questions (2007–08 [SA]), student performance was significantly better than their respective performance in general pathology (Figures 2B and 2C, Table 3). In both classes, the course scores in oral and maxillofacial pathology predicted for a 70 percent score in the prerequisite general pathology course were significantly higher than 70 percent (Table 3), indicating significant improvement. The slopes in the 2004–05 (SA/MC) and 2007–08 (SA) classes were not significantly different from 1.00; that is, the principal component lines were parallel to the line of equality. The finding that the principal component lines were parallel to the line of equality indicated that students of all abilities, as assessed by their scores in general pathology, responded similarly to the changes in examination question format. Furthermore, there were no significant differences in the principal component lines (either the slopes or the predicted score in oral and maxillofacial pathology for a score in general pathology of 70 percent) of the 2004–05 (SA/MC) and 2007–08 (SA) classes (Figures 2B and 2C, Table 3). This finding indicates that no additional improvement in student achievement occurred as a result of using 100 percent short-answer questions compared to 50 percent short-answer questions.

When correction for guessing on the multiple-choice examinations was implemented prospectively, as well as the utilization of half short-answer examinations, there was an additional significant increase in student achievement (Figure 2D, Table 3). In the 2005–06 (SA/MC*) class, the slope of the principal component line was significantly different from 1.00, indicating that students varying in their previous achievement responded differently—an Aptitude-Treatment Interaction. In the 2005–06 (SA/MC*) class, the predicted course score in oral and maxillofacial pathology for a 70 percent score in general pathology was significantly higher than 70 percent, as well as significantly higher than the predicted course scores for the other three classes (Table 3). Thus, the additional improved achievement of the 2005–06 (SA/MC*) class occurred particularly among the lower achieving students.

We next investigated how the changes in assessment procedures affected students differing in their previous achievement. We classified students into grade categories using common cutoff values. For each of the three categories (80–100 percent, 70–79 percent, <70 percent) in general pathology, Figure 3 shows the differences between the means

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>General Pathology Mean±SE</th>
<th>Oral and Maxillofacial Pathology Mean±SE</th>
<th>Principal Component Line†</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004–05 (SA/MC)</td>
<td>77.2±0.8†</td>
<td>82.2±0.8†</td>
<td>74.5 (74.3–96.6)†</td>
</tr>
<tr>
<td>2004–05 (SA/MC*)</td>
<td>76.1±0.8</td>
<td>84.1±0.6</td>
<td>76.0 (74.6–77.3)</td>
</tr>
</tbody>
</table>


*aCorrection for random guessing applied prospectively.
*bEquation of the principal component line is Y = β0 + β1X. β0 is the score in oral and maxillofacial pathology course predicted for a score of 70 percent in the general pathology course. β1 is the slope of the principal component line.
*cSignificantly (p≤0.05) different from 70.0.
*dSignificantly (p≤0.05) different from 1.00.
*eSignificantly (p≤0.05) different from 2006–07 (MC).
*fSignificantly (p≤0.05) different from 2005–06 (SA/MC*).
in oral and maxillofacial pathology and the means in general pathology for each of the four classes. For the 2006–07 (MC) class, the decrease in mean scores in oral and maxillofacial pathology relative to general pathology for students in the 80–100 percent grade category and the increase in course scores for students in the <70 percent grade category simply represent regression toward the mean. This interpretation is supported by the similarity of the observed and expected changes between the two courses. For the 80–100 percent students, we calculated a regression toward the mean of -1.39 percent and observed -1.74 percent; for the 70–79 percent students, we calculated 0.57 percent and observed 0.49 percent; and for the <70 percent students, we calculated 2.34 percent and observed 2.71 percent. Thus, the observed differences are consistent with the conclusion shown in Figure 2 that students’ achievement in the 2006–07 (MC) class in oral and maxillofacial pathology was almost identical to their achievement in general pathology.

For all three grade categories established in general pathology, the three classes in oral maxillofacial pathology in which short-answer questions were used (2004–05 [SA/MC], 2005–06 [SA/MC*], and 2007–08 [SA]) had significantly higher scores than in the class in which all multiple-choice examinations were used (2006–07 [MC]) (Figure 3). There were no significant differences between the 2004–05 (SA/MC) class and the 2007–08 (SA) class, indicating that utilizing all short-answer examinations was no better or worse than utilizing half short-answer examinations (Figure 3).

Figure 2. Scatter diagrams and principal component lines relating course scores in the oral and maxillofacial pathology and general pathology courses, by academic year

Note: The solid line represents the first principal component line, the long dashed line represents equality, and the short dashed lines indicate the predicted score in the oral and maxillofacial pathology course for a score of 70 percent in the general pathology course. All multiple-choice examination scores have been corrected for random guessing. The course scores for the 2006–07 (MC) class are shown in Panel A, the 2004–05 (SA/MC) class in Panel B, the 2007–08 (SA) class in Panel C, and the 2005–06 (SA/MC*) class in Panel D.
For the students scoring <70 percent in general pathology, the mean numerical course score for the 2005–06 (SA/MC*) class in oral and maxillofacial pathology was significantly higher than in the 2004–05 (SA/MC) and 2007–08 (SA) classes (Figure 3). These findings indicate that prospective utilization of the correction for guessing scoring in multiple-choice examinations resulted in improved scores by lower performing students beyond the improvement obtained by use of short-answer examinations. That is, the response to the correction for guessing scoring was different for students of different abilities—a reflection of the Aptitude-Treatment Interaction.

Regression toward the mean, of similar magnitude as that which occurred in the 2006–07 (MC) class, was expected to occur in the other three classes. However, the changes between scores in general pathology and oral and maxillofacial pathology for the 2004–05 (SA/MC), 2007–08 (SA), and 2004–05 (SA/MC*) classes were all significantly higher than the changes in the 2006–07 (MC) class tested using all multiple-choice examinations (Figure 3). These significantly higher differences were observed in all grade categories in the general pathology course. Thus, the magnitudes of increased academic achievement observed were far larger than the magnitude of the regression toward the mean and support genuine effects due to changes in the assessment procedures.14 Similarly, the significantly enhanced achievement observed for the lower performing students, when both short-answer questions and prospective correction for guessing were used, also supports genuine effects due to these interventions.

Instructors usually report course performance using grade categories, so it is important to see how the examination question format and scoring procedure affected grade category distributions. The grade category distributions in oral and maxillofacial pathology are shown in Figure 4 for each of the three grade categories previously established from general pathology. For all three general pathology grade categories, the frequencies of students in the three grade categories in the three classes of oral and maxillofacial pathology in which short-answer examinations were used (2004–05 [SA/MC], 2005–06 [SA/MC*], and 2007–08 [SA]) were significantly (p≤0.0281) im-
proved relative to the grade category distribution in 2006–07 (MC). The improvements in grade category distributions reflected a greater percentage of students improving their academic achievement beyond their previous achievement in general pathology and/or a lesser percentage of students declining in their achievement in oral and maxillofacial pathology relative to their achievement in general pathology. There were no significant differences in the grade category distributions in oral and maxillofacial pathology between the use of 50 percent short-answer examinations (2004–05 [SA/MC]) and 100 percent short-answer examinations (2007–08 [SA]).

For students scoring either 70–79 percent or <70 percent in general pathology, the fractions of students improving their grade category in oral and maxillofacial pathology were significantly higher in 2005–06 (SA/MC*) than in 2004–05 (SA/MC) (p≤0.0486). These results again show that academic achievement was improved when examinations included short-answer questions and that lower performing students accomplished additional higher achievement, above that accomplished with the use of short-answer examinations, when correction for guessing was included in the scoring algorithm for multiple-choice examinations.

**Short-Answer and Multiple-Choice Examination Scores**

The foregoing results for overall course scores from oral and maxillofacial pathology indicate that examination question format and scoring procedure used for the multiple-choice examinations affected student performance. Although overall course score is a prime concern of dental educators, we also examined how the different examination formats and/or scoring procedures separately affected scores on short-answer examinations and on multiple-choice examinations. The findings from these analyses, which showed that the effects were similar to those observed on overall course score, provided essential validation of the conclusions obtained using overall course scores.

---

**Figure 4. Distributions of course grade category in the oral and maxillofacial pathology and general pathology courses, by academic year**

*Note:* Column of pie charts represents distribution (percentage of students) of grade category in oral and maxillofacial pathology course based on overall score by academic year; and row of pie charts represents course grade category in the general pathology course. All multiple-choice examination scores have been corrected for random guessing. Pie charts for 2006–07 (MC) and 2007–08 (SA) classes reported in Pinckard RN, McMahan CA, Prihoda TJ, Littlefield JH, Jones AC. Short-answer examinations improve student performance in an oral and maxillofacial pathology course. J Dent Educ 2009;73(8):950–61.

- Students in the 80–100 percent course grade category in the general pathology course.
- Students in grade category 70–79 percent in the general pathology course.
- Students in grade category <70 percent in the general pathology course.
Short-answer examination scores. The relationships between the student scores on the short-answer examinations in oral and maxillofacial pathology and their scores in general pathology are shown graphically for the 2004–05 (SA/MC) (Figure 5A), 2005–06 (SA/MC*) (Figure 5B), and 2007–08 (SA) (Figure 2C) classes (see Table 4 for the estimated parameters of the principal component lines). The slopes were not significantly different from 1.00, nor were there significant differences among the slopes for the 2004–05 (SA/MC), 2005–06 (SA/MC*), and 2007–08 (SA) classes. However, the scores in oral and maxillofacial pathology predicted from the principal component lines for a 70 percent score in general pathology were significantly greater than 70 percent for all three classes; these significant differences indicated improved achievement relative to general pathology in the classes in which short-answer examinations were used in oral and maxillofacial pathology. Of note, the short-answer examination scores predicted for a 70 percent score in general pathology were significantly higher for the 2005–06 (SA/MC*) class than for the 2004–05 (SA/MC) and 2007–08 (SA) classes. Thus, prospectively implementing correction-for-guessing scoring of the multiple-choice examinations also enhanced scores on the short-answer examinations.

Figure 5. Scatter diagrams and principal component lines relating scores in the oral and maxillofacial pathology and general pathology courses, by academic year and examination question format

Note: The solid line represents the first principal component line, the long dashed line represents equality, and the short dashed lines indicate the predicted score in the oral and maxillofacial pathology course for a score of 70 percent in the general pathology course. All multiple-choice examination scores have been corrected for random guessing. Data for scatter diagrams reported in Prihoda TJ, Pinckard RN, McMahan CA, Littlefield JH, Jones AC. Prospective implementation of correction for guessing in oral and maxillofacial pathology multiple-choice examinations: did student performance improve? J Dent Educ. 2008;72(10):1149–59. Scores from the short-answer examinations for the 2004–05 (SA/MC) class are shown in Panel A and for the 2005–06 (SA/MC*) class in Panel B; scores from the multiple-choice examinations for the 2004–05 (SA/MC) class are shown in Panel C and for the 2005–06 (SA/MC*) class in Panel D.
Table 4. Estimated parameters of principal component lines relating course scores in the oral and maxillofacial pathology and general pathology courses, by examination format and academic year

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>Short-Answer Examinations</th>
<th>Multiple-Choice Examinations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated $\beta_{70}$</td>
<td>Estimated $\beta$</td>
</tr>
<tr>
<td></td>
<td>(95% CI)</td>
<td>(95% CI)</td>
</tr>
<tr>
<td>2006–07 (MC)</td>
<td>—</td>
<td>70.3 (68.5–71.9)$^b$</td>
</tr>
<tr>
<td>2004–05 (SA/MC)</td>
<td>75.1 (73.5–76.9)$^{ab}$</td>
<td>74.8 (73.6–76.0)$^{abc}$</td>
</tr>
<tr>
<td>2007–08 (SA)</td>
<td>76.0 (74.6–77.3)$^{ab}$</td>
<td>—</td>
</tr>
<tr>
<td>2005–06 (SA/MC$^*$)</td>
<td>79.5 (77.6–81.4)$^a$</td>
<td>78.4 (76.8–79.9)$^a$</td>
</tr>
</tbody>
</table>

*$^a$Correction for random guessing applied prospectively.
*$^b$Significantly (p≤0.05) different from 70.0.
*$^c$Significantly (p≤0.05) different from 2005–06 (SA/MC$^*$).
*$^d$Significantly (p≤0.05) different from 2006–07 (MC).

Figure 6A summarizes graphically the differences between the means of the short-answer examinations in oral and maxillofacial pathology and the means in general pathology (based on multiple-choice examinations). Regardless of grade categories established in general pathology, all three classes (2004–05 [SA/MC], 2007–08 [SA], and 2005–06 [SA/MC$^*$]) performed significantly better on the short-answer examinations than predicted from their scores in the general pathology course. There was no significant difference in scores between utilizing 100 percent short-answer examinations (2007–08 [SA]) and 50 percent short-answer examinations (2004–05 [SA/MC]). However, in the class (2005–06 [SA/MC$^*$]) in which prospective correction for guessing on multiple-choice examinations was utilized, there was an Aptitude-Treatment Interaction observed in the short-answer examination scores, with the lower performing students attaining higher scores.

Multiple-choice examination scores. The relationship of the student scores on the multiple-choice examinations in oral and maxillofacial pathology with the scores in general pathology are shown graphically for the 2004–05 (SA/MC) (Figure 5C), 2005–06 (SA/MC$^*$) (Figure 5D), and 2006–07 (MC) (Figure 2A) classes; the estimated parameters of the principal component lines are shown in Table 4. The slopes of the principal component lines were not significantly different from 1.00, nor were there significant differences among the slopes for the 2004–05 (SA/MC), 2005–06 (SA/MC$^*$), and 2006–07 (MC) classes. The predicted score in the 2004–05 (SA/MC) class in oral and maxillofacial pathology for a 70 percent score in general pathology was significantly higher than 70 percent, indicating improved achievement on multiple-choice examinations due to inclusion of a short-answer component. Of note, the score predicted for a 70 percent score in general pathology in the 2005–06 (SA/MC$^*$) class not only was significantly higher than for the 2006–07 (MC) class, but also was significantly higher than the predicted scores for the 2004–05 (SA/MC) class. This higher predicted score demonstrated additional improved achievement of the lower performing students in response to prospective implementation of correction-for-guessing scoring.

Discussion
By analyzing the numerical course scores of students in four successive academic years of an oral and maxillofacial pathology course, we have shown
that, relative to using 100 percent multiple-choice examinations, the inclusion of a 50 percent short-answer component to the examinations significantly improved student achievement. This improvement was observed in students at all prior levels of achievement as determined by their course scores in a prerequisite general pathology course. Using 100 percent short-answer examinations did not further increase student achievements over and above that observed with 50 percent short-answer examinations. We also found that, by prospectively applying a scoring algorithm for correction for random (no knowledge) guessing to multiple-choice examinations, there was an additional improvement over and above the improvement from using short-answer examinations. However, this improvement applied only to the lower scoring students in the general pathology course. The statistical analysis techniques employed in this study not only confirmed our previously reported results but extended those findings by showing that the use
of short-answer questions and correction for guessing on multiple-choice examinations each separately enhanced student academic achievement. Thus, simply modifying the educational environment by changes in examination question format and scoring procedure had dramatic effects on student behaviors as reflected in their increased academic achievement.

At the time we decided to use the more challenging all short-answer examinations (2007–08 [SA]) in oral and maxillofacial pathology, we were concerned that this could adversely affect student course grades. However, our concern was unfounded as student course scores in fact were significantly enhanced (Figure 1, Table 3). Even the uncorrected scores for the multiple-choice examinations (data not shown), which are known to be inflated, did not exceed the all short-answer examination scores. Indeed, students often are able to eliminate one or more incorrect answers to a multiple-choice question with four or more options. Thus, the reported scores from multiple-choice examinations in this study likely are still inflated due to guessing and the benefits of using short-answer examinations relative to multiple-choice examinations likely are greater than reported here. Therefore, we concur with a previous study that suggested that short-answer examinations made the students study differently, thereby reaching higher levels of academic achievement.

The various changes made in the oral and maxillofacial pathology course were those that the course directors thought at the time were appropriate to provide a better educational experience for the students; however, even with good intentions, changes instituted in the academic environment can have negative consequences. For example, the decision to implement all multiple-choice examinations in the oral and maxillofacial pathology course in 2006–07 was made anticipating that these students would outperform the preceding two classes (2004–05 [SA/MC] and 2005–06 [SA/MC*]), in which more challenging short-answer examinations and correction for guessing on the multiple-choice questions were used. However, unfortunately, the scores of the 2006–07 (MC) class declined significantly when compared with the 2004–05 (SA/MC) and 2005–06 (SA/MC*) classes and, subsequently, relative to the 2007–08 (SA) class as well. Thus, the nature of the academic environment can foster detrimental, as well as beneficial effects, on student learning behaviors and academic achievement. This points to the need for all aspects of the educational environment to be carefully assessed for positive or negative influences on student learning.

We believe that students’ preparation for the more challenging un-cued short-answer examinations beneficially changed their overall study approach and perhaps their learning behavior since their scores on the multiple-choice components also improved. The students in the classes in which examinations were comprised of 50 percent short-answer questions did not know what material would be tested in those questions and what material would be tested in multiple-choice questions. Thus, these students had to prepare for all material to be tested using the short-answer examinations, and it is not surprising that the 100 percent short-answer examinations resulted in essentially the same performance as the half short-answer examinations. These similar outcomes suggest that quantitatively augmenting an assessment procedure, even a procedure with established benefits, may not produce additional improvements.

Undoubtedly, students pass on examination information to students of succeeding classes, and this could have affected our conclusions. However, because of the changes in question format that were made during the four years, we feel that this passing on of examination information would be minimized. Indeed, the significant decrease in student performance by changing to all multiple-choice questions in the 2006–07 (MC) class from half multiple-choice and half short-answer questions in the 2005–06 (SA/MC*) class supports this contention. While the 2007–08 (SA) class outperformed the 2006–07 (MC) class, the higher scores of the 2007–08 (SA) class were made in spite of facing the more challenging, un-cued short-answer examinations. A much stronger case for passing on examination information could be made relative to the 2004–05 (SA/MC) and 2005–06 (SA/MC*) classes, when the formats were identical. However, we previously documented that comparable improvements between these two classes were observed for the two subsets of the same questions and different questions in both the short-answer and multiple-choice portions of the examinations. Thus, while we cannot completely rule out the passing on of examination information, it is our contention that this did not play an important role in our investigations.

We do not have an assessment of the effects of changes in the educational environment on long-term student knowledge of oral and maxillofacial pathology relative to their performance on the course examinations. An external assessment, such as score from a subset of board examination questions, would be valuable in this regard. Currently, there is no such
score available, and the lack of such represents a limitation of our study.

Another limitation in our study is that we do not know which aspects of the students’ learning behavior might have been changed nor do we know what psychological factors might have been affected by the changes in examination question format and scoring procedure. We can only say with certainty that, with a more challenging question format and scoring algorithm, students’ performance significantly improved. Further study elucidating the underlying factors that influenced the improvement would be helpful to faculty members trying to implement changes that would most improve student learning behavior. However, even without information on underlying factors, educators can use our findings to guide their changes in assessment procedures and anticipate improvements such as we observed.

Conclusion

Examining students to certify the level of knowledge they have attained is a required and important faculty function. This function can be carried out using many different examination question formats and scoring procedures. Our studies show that the appropriate selection of examination question format and scoring procedure not only can accomplish this certification function, but also can have positive effects on student learning behaviors to enhance their academic achievement.

Our goal is for all second-year dental students to attain a course score of 80 percent or higher in the oral and maxillofacial pathology course. We believe the competitive selection process used for student admission to our dental school makes this goal realistic. As we have demonstrated, including a short-answer question component to examinations and a more demanding scoring algorithm has moved us towards our goal. It is important to point out that these positive changes in student performance were accomplished in an academic setting in which there were no changes in either curricular content or emphasis during the four-year time span of our study. Moreover, it must be emphasized that we have assessed only two of many possible changes that could have been made in the educational environment to foster improved student learning and academic achievement. Thus, we are encouraged by the results so far and intend to continue modifying the educational environment for the betterment of student learning behavior and, thereby, enhancing their academic achievement. Finally, we hope that others will consider implementing similar as well as additional beneficial changes in the educational environment to foster better student learning during the four years of dental education.

Acknowledgment

The authors thank William D. Hendricson for his helpful and valued review of our manuscript.

REFERENCES