Empowering Student Learning Through Rubric-Referenced Self-Assessment*

Xiaohua He, MD, MS, and **Anne Canty**, PhD, Palmer College of Chiropractic Florida

Purpose: The purpose of this study was to investigate the effect of rubric-referenced self-assessment on performance of anatomy assignments in a group of chiropractic students. **Methods:** Participants ($N = 259$) were first-quarter students who were divided into a treatment group ($n = 130$) and a comparison group ($n = 129$). The intervention for both groups involved the use of rubrics to complete the first draft of assignments. General feedback was given by the instructor, and then the students had the opportunity to amend the assignments before resubmission (second draft). The treatment group, however, was also asked to perform rubric-referenced self-assessment of their assignments during their second draft. Although the comparison group was also provided with the identical rubrics for the assignments, the students in this group did not perform rubric-referenced self-assessment. **Results:** The results revealed that the students in the treatment group who used a rubric-referenced self-assessment learning tool received statistically significant higher scores than the comparison group, who did not use this rubric-referenced self-assessment tool. **Conclusion:** This study suggests that practicing rubric-referenced self-assessment enhances student performance on assignments. However, educators continue to face the challenge of developing practical and useful rubric tools for student self-assessment. (J Chiropr Educ 2012;26(1):24–31)

Key Indexing Terms: Chiropractic; Education; Learning; Self-Assessment

INTRODUCTION

Although student self-assessment has been shown as a very effective and integrative tool in student learning,[1] the application of student self-assessment, including the development of related tools for student self-assessment (such as rubrics), remains a challenge for classroom instructors. Other challenges, which include misconceptions about student self-assessment, difficulty finding time to give all students the feedback, and lack of knowledge necessary for creating assessment tools, may limit the application of student self-assessment in the classroom. Recognizing the potential benefits of student self-assessment in student learning, we conducted a pilot study to explore opportunities for students and teachers to collectively act and apply student self-assessment to student learning.[2] As part of this effort, we interviewed 37 students who participated in the study of self-assessment during which the students examined their attitudes toward self-assessment and developed a means through which they could effectively support the use of such assessment in their learning. During the self-assessment, students reflected on the quality of their work on the assignments, judged the degree to which it reflected explicitly stated goals or criteria, and revised their work accordingly. Students reported that their attitudes toward self-assessment became more positive as they gained experience with it.[2]

One of the ways to support thoughtful self-assessment is to provide a rubric to students. A rubric is a document that lists criteria and describes various levels of quality, from excellent to poor, for

* This paper was selected as a 2011 Association of Chiropractic Colleges Research Agenda Conference Prize Winning Paper. The award is funded by the National Board of Chiropractic Examiners

The Journal of Chiropractic Education

Vol. 26, No. 1. Printed in U.S.A.
1042-5055/$22.00

a specific assignment.[3] Rubrics are often used to grade student work, but they can also serve another more important role as well; that is, rubrics can be used to teach as well as to evaluate.[4–6] The claim that rubrics can promote learning and achievement has global appeal, but there is only limited empirical evidence to support it. Our previous study, which utilized a small sample size ($N = 74$) found that students who used rubrics for self-assessment received higher scores on written examinations than students who did not use the rubrics.[2] This study revealed that students who had received rubrics tended to identify more of the criteria by which their assignments were evaluated. This suggests that the students were developing an understanding of what was required for successful completion of assignments as defined by the rubrics they received.

Like rubric-referenced evaluation, student self-assessment with its potential to increase learning and achievement has received a lot of attention, much of it rhetorical.[7,8] A study of the influence of self-assessment training on students' scores on external exams[9] showed a consistently positive effect from the application of rubrics. These results suggest a link between rubric-referenced self-assessment and student achievement. Given the recent surge in interest in creation of rubrics at our college, there is need for more research. This study was designed to extend the previous pilot study utilizing a small sample size to a larger sample group. The purpose of this study was to test popular claims about the relationship between rubric-referenced assessment, in particular self-assessment, and students' performance on anatomy assignments. Two research questions were addressed:

1. Did the utilization of rubric-referenced self-assessment affect scores assigned to students' classroom work?
2. Which of the following correlated more closely to scores received on student assignments: the amount of time students spent on classroom assignments or the implementation of rubric-referenced self-assessment in completion of the assignments?

METHODS

This study was approved by the Institutional Review Board of Palmer College of Chiropractic.

Participants

The study employed a convenience sample of 259 volunteer participants. One hundred thirty participants were used as the treatment group, and the remaining 129 participants served as the comparison group. First-quarter students from five different academic terms were recruited during their anatomy class at Palmer College of Chiropractic, Florida Campus. The allocation of the groups was based on the class. The demographic information of each class was similar: 97% of these students had college-level anatomy before enrolling in the chiropractic program. Among the participants, 96 (37.1%) were females and 163 (62.9%) were males.

Assignments

Each group was asked to do four anatomy assignments based on the materials learned in lectures and laboratories. These assignments included structural comparison, concept description, material summarization, and clinical relevance of anatomical knowledge. The time allowed for each assignment was 120 minutes. The actual time spent on the assignments was recorded and correlations were made between the time spent on the assignments and the scores assigned to each assignment. A means of leveling the students' anatomy knowledge was included in the rubrics. Four different scoring levels were chosen in order to increase discrimination and measure performance quality on the assignment (see Appendix). The scoring rubrics were tested by students from previous terms and other teachers who were not involved in the study. The rubrics were repeatedly revised until ambiguity was minimized.

Procedures

The assignment process in each group was as follows: The instructor gave directions for the assignments to both groups. For example, when students were asked to do the structural comparison, the instructor gave the example of structures for comparison. Students ultimately decided how many structures they would compare based on their understanding of the importance of these structures. They were provided with rubrics, wrote the first drafts, submitted the assignments, and got general feedback from the teacher. They then made further amendments to their assignments.

The treatment group differed from the comparison group in two ways: The students in the treatment group (1) discussed their own strengths and weaknesses in performing the assignments and (2) used the rubrics to self-assess their first drafts. The students in the comparison group did not discuss their strengths and weaknesses and did not perform self-assessment on the assignments, although they were asked to review their first drafts and note possibilities for improvement in the final draft. For the purposes of research, the rubrics given to the students in the treatment and comparison groups were identical to each other.

Self-Assessment

During the self-assessment performed by students in the treatment group, students were asked to identify key points in the rubric, then underline or circle in their first submission the evidence of having met the standard articulated by the key points. If they found that they had not met the standard, they were asked to make improvements in their final submission. Students were given class time (120 minutes) to complete each step of the assignments. Students in the comparison group were also given the same amount of time for their assignment revision. However, self-assessment was not required of this group. We compared the time spent on assignments between the treatment and comparison groups, as well as the correlation of time spent on assignments to assignment scores.

Grading

The time spent on the assignments was recorded. The total score (10 points for each assignment, maximum 40 points) for the assignments was applied based on the accuracy of individual criteria exhibited in the assignments. These criteria included appropriate anatomical terminology (6 points), number of common anatomical features (6 points), number of specific anatomical features (6 points), accuracy of structural relationships (6 points), accuracy of functional relationships (6 points), clinical relevance (5 points), and the time to complete the assignment (5 points).

We compared the scores between the treatment and comparison groups. The main effects of the variables on the scores for individual criteria were examined using a general linear model multivariate test. The correlation between time spent on the assignment and scores was evaluated using Pearson correlation and regression analysis. The comparisons of assignment scores and time spent on assignments between groups and individual criteria were analyzed using independent sample *t* tests. All statistics were performed using SPSS 15.0 software (SPSS Inc., Chicago, IL).

RESULTS

Total Assignment Score and Time Spent on Assignment

The total score for the final submission was derived from the sum of the four assignments (10 points each, maximum 40 points). The average assignment score for the entire sample was 34.25 (SD = 5.2), with a range of 24 to 40. On average, the treatment group's assignment scores (M = 36.45, SD = 3.6) were higher than the comparison group's scores (M = 32.5, SD = 5.7). Independent *t* sample tests showed a statistical significance between the two groups for scores achieved on assignments ($p < .001$) (Fig. 1). On average, the comparison group's time spent on assignments (M = *70.5*, SD = *11.5*) was less than that of the treatment group (M = 111.3, SD = 19). Independent *t* sample tests showed a statistical significance between two groups for time spent on the assignments ($p < .001$) (Fig. 1).

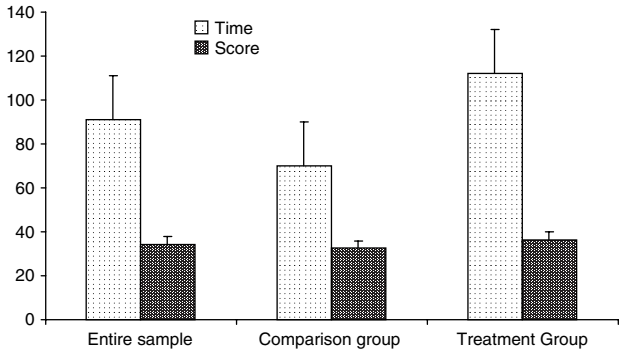

Figure 1. Bar graph shows mean ± SD of time and scores of each group. There are statistically significant differences for time and scores between comparison and treatment groups ($p < .001$).

Scores on Individual Criteria

Figure 2 contains the means and standard deviations for scores on each criterion of the rubrics by groups. The general linear model (GLM) results show that there was a statistical significance across criteria scores between the treatment group and the comparison group ($p < .001$). A statistical significance was determined across the scores for the following assignment criteria: terminology ($p < .05$), common features ($p < .01$), specific features ($p < .01$), structural relationship ($p < .01$), and functional relationship ($p < .01$). The only criteria that did not show statistical significance were time of submission ($p > .05$) and clinical relevance ($p > .05$).

Correlation Between Time Spent on Assignments and Scores

The amount of time devoted to each assignment varied by individuals, from 43 to 110 minutes for the comparison group and 91 to 120 minutes for the treatment group. The mean time across the entire sample was 91 minutes. When considering both groups collectively, time spent on assignments significantly correlated with assignment scores [analysis of variance (ANOVA): $p < .01$; Pearson: $r = 0.63$] (Fig. 3). However, when the time spent on the assignment and the assignment scores within individual groups were compared, although there was a significant difference on students' time spent on the assignment (ANOVA: $p < .01$), there was no significant correlation between time spent on the assignment and scores ($r = 0.15$ for treatment group; $r = 0.16$ for comparison group) (Figs. 4 and 5).

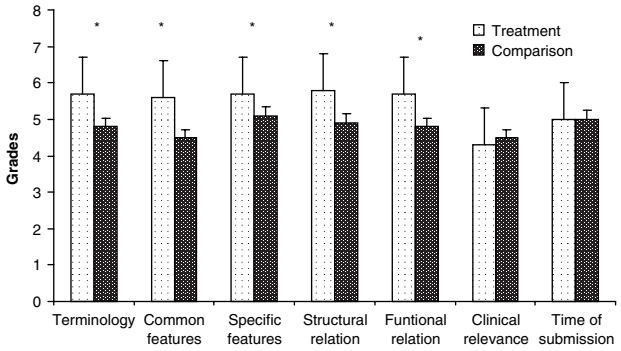

*Figure 2. Bar graph shows mean ± SD grades of individual criteria. Note that except for the clinical relevance and time of submission there are statistically significant differences between treatment and comparison groups. *, Statistical significance.*

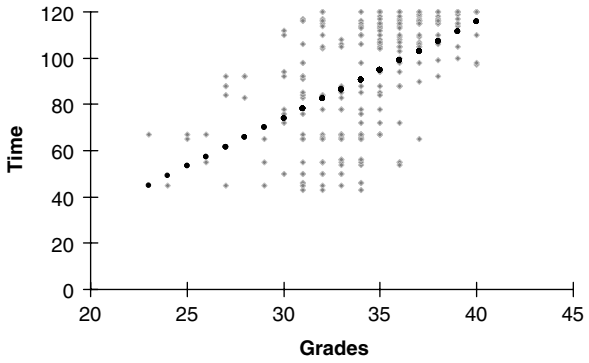

Figure 3. Diagram suggests the correlation between the time spent on each assignment and the scores on the entire sample ($r = 0.63$, $p < .01$).

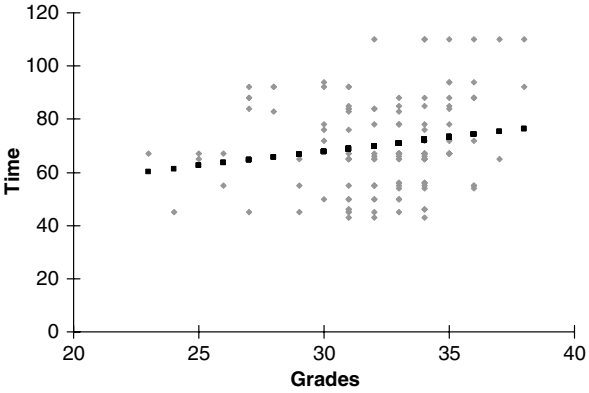

Figure 4. Diagram shows no correlation between the time spent on each assignment and the scores within the comparison group ($r = 0.15$), although there is a significant difference among students on time spent on each assignment ($p < .01$).

DISCUSSION

This study provides positive support for the hypothesis that the utilization of rubrics to perform self-assessment on assignments influences scores on the performance of such assignments. This raises the question, then, as to which factor contributed to such influences. Initially, it seemed that the amount of time spent on assignments influenced the scores on performance of the assignment because the correlation across the entire sample showed significance. If this was true, the time spent on assignments would

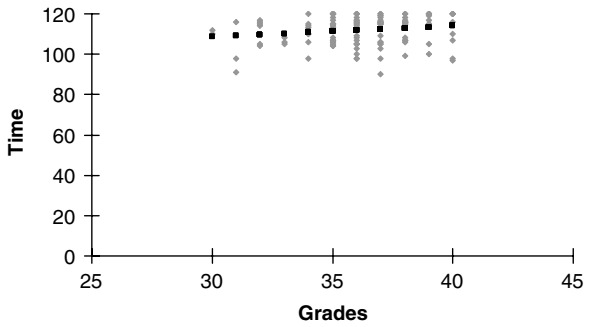

Figure 5. Diagram shows no correlation between the time spent on each assignment and the scores within the treatment group ($r = 0.15$), although there is a significant difference among students on time spent on each assignment ($p < .01$).

contribute to the higher scores for students in the treatment group, since students in this group spent an average of 40 more minutes on their assignments. However, when we examined the correlation within each group separately, we found, interestingly, that although there was a statistical significance on time spent on assignments among individuals, there was actually no correlation between time spent on assignment and assignment scores. This indicates that time is not a direct contributor for better scores for students in the treatment group. The fact that the treatment group spent more time on the assignments should not be surprising. The reason that the treatment group used more time on their assignments was mainly because of the incorporation of rubric-referenced self-assessment. Therefore, it was rubric-referenced self-assessment and not the time spent on the assignments that should be considered as a main contributor to better performance in the treatment group. The self-assessment requires students to discuss the strength and weaknesses of their first draft of the assignment, check the criteria, and find the key points. Although students in the comparison group were also provided with the same rubrics, they were not asked to perform self-assessment; hence, it is clear that utilizing rubrics for self-assessment has a statistically significant and positive effect on assignment scores.

Translation of the scores achieved on the rubric-referenced assignments to typical classroom grades (70%–100%) can be meaningful. For example, a score of 40 on each assignment would be translated to 100%, a score of 36 on each assignment would be translated to 90%, and so forth. The average grade for the treatment group resulted in a typical classroom grade in the lower 90% range, compared to the average comparison group grade in the lower 80% range.

In the analysis of the scores received on individual criteria (terminology, common features, specific features, structural relationship, functional relationship, clinical relevance, and time of submission), rubric-referenced self-assessment had a significant influence on all criteria except for clinical relevance and time of submission. Time of submission was the criterion not explicitly required during the rubric-referenced self-assessment performed by the treatment group. The reason that there was no statistical significance on clinical relevance was because these students were in their first quarter of study. The first quarter mainly focuses on the study of the basic sciences, and clinical contents are not as stressed as in the subsequent quarters. Because of the relatively limited clinical application in the first-quarter curriculum, it is reasonable to understand why there was no statistical significance for this criteria between groups. In regard to the time of submission, the total time allowed to complete the assignments was the same for both groups. All students from both groups did not exceed the required amount of time; therefore, no statistical difference was shown.

We interpret the phenomenon that students in the treatment group performed better on terminology, common features, specific features, structural relationship, and functional relationship than their counterparts in the comparison group as additional evidence of the effect of formal self-assessment on student learning. The fact that rubric-referenced self-assessment was associated with higher scores on these important content areas further testifies to the potential of such processes to help students master significant, meaningful subject matter.

LIMITATIONS

There are several limitations to this investigation. One limitation is the short treatment time. Students were asked to perform the assessment over one term only (11 weeks). Research on the long-term effects of rubric-referenced self-assessment implemented by students across multiple quarters could be illuminating. Another limitation was that we recommended the use of this learning strategy only to the teaching of anatomy because that is what we studied. Studies like this one are needed in other subject areas,

especially clinical sciences, which tend to involve students in qualitatively different kinds of work.

It is worthy to note that although it is generally acknowledged that self-assessment drives learning, self-assessment can have both intended and unintended consequences. Students study more thoughtfully when they perform the self-assessment and focus more on the details that instructors require them to know. Self-assessment with peers seems also to promote professionalism, teamwork, and communication. The unintended effects of assessment include the tendency for students to do the assessment just because they receive a grade. Also, students may only associate the criteria in the rubric with what they think they need to know for the entire course content. In this respect, it can be limiting for the student.

CONCLUSION

This study revealed that practicing rubric-referenced self-assessment enhanced students' performance on anatomy assignments.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

ACKNOWLEDGMENTS

The authors thank Ying Cao, a research staff member at Palmer College of Chiropractic Research Center, for his statistical assistance.

About the Authors

Xiaohua He and Anne Canty are both professors at Palmer College of Chiropractic Florida. Address correspondence to Xiaohua He, Palmer College of Chiropractic Florida, 4777 City Center Parkway, Port Orange, FL 32129 (e-mail: shawn.he@palmer.edu). This article was received April 22, 2011, revised July 20, 2011, and accepted August 9, 2011.

REFERENCES

1. Asadoorian J, Batty HP. An evidence-based model of effective self-assessment for directing professional learning. J Dent Educ 2005;69(12):1315–23.
2. He X, Canty A. Student self-assessment: what research says and what practice shows. J Chiropr Educ 2010; 24(1):95.
3. Andrade H. Using rubrics to promote thinking and learning. Educ Leadership 2000;57(2):13–18.
4. Arter J, McTighe J. Scoring rubrics in the classroom: using performance criteria for assessing and improving student performance. Thousand Oaks, CA: Corwin Press; 2001.
5. Quinlan A. Assessment made easy: scoring rubrics for teachers from K-college. Lanham, MD: Rowman and Littlefield Education; 2006.
6. Stiggins RJ. Student-involved classroom assessment. 3rd ed. Upper Saddle River, NJ: Merrill/Prentice-Hall; 2001.
7. Baldwin D. A guide to standardized writing assessment. Educ Leadership 2004;62(2):72–75.
8. Leonhardt A. Using rubrics as an assessment tool in your classroom. Gen Music Today 2005;19(1):10–16.
9. MacDonald B, Boud D. The impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations. Assess Educ 2003;10(2):209–20.

APPENDIX
SAMPLE RUBRICS USED FOR STUDY

Grading Rubric 1: Compare and contrast vertebrae of different segments

Grade	Criteria	Additional Considerations
F	➢ Compare and contrast the common features of vertebrae of different segments (eg, bodies, arches, processes) or less only.	
C	➢ Compare and contrast the common features of vertebrae of different segments (eg, bodies, arches, processes). ➢ Compare and contrast the typical and atypical vertebrae in each segment (eg, C_1C_2 vs. C_3; $T_{11}T_{12}$ vs. T_6; L_5 vs. L_2). ➢ Briefly mention the clinical relevance of above mentioned comparison.	◦ Did you work along or work with a group?
B	➢ Clearly compare and contrast the common features of vertebrae of different segments (eg, bodies, arches, processes). ➢ Clearly compare and contrast the typical and atypical vertebrae in each segment (eg, C_1C_2 vs. C_3; $T_{11}T_{12}$ vs. T_6; L_5 vs. L_2). ➢ Clearly compare and contrast the unique features of different segments (eg, extra structures on vertebrae of different segments). ➢ State clinical relevance of above comparison.	◦ Did you utilize your textbook, lecture notes, and atlas and lab materials as your reference? ◦ This homework is worth 10 points. The following point scale is used for grading: A = 9–10 B = 8–8.9 C = 7–7.9 F = 0–6.9
A	➢ Clearly compare and contrast the common features of vertebrae of different segments (eg, bodies, arches, processes, etc). ➢ Clearly compare and contrast the typical and atypical vertebrae in each segment (eg, C_1C_2 vs. C_3; $T_{11}T_{12}$ vs. T_6; L_5 vs. L_2). ➢ Clearly compare and contrast the unique features of different segments (eg, extra structures on vertebrae of different segments). ➢ Clearly compare and contrast the facet joint orientation (eg, different joint facet orientation of cervical, thoracic, and lumbar). ➢ Clearly explain the functional differences of above comparisons. ➢ Clearly state clinical relevance of above comparisons.	◦ The students work together as a group when doing self-grading. The instructor will use the same grading system.

Grading Rubric 2: List all the structures at the following three levels: C1; T5, and L3

Assessment Plan Elements	Grades			
	A	B	C	F
Students' learning outcomes	At least 22 structures are identified at C1 level. At least 18 structures are identified at T5 level. At least 16 structures are identified at L3 level. At least 4 structures have clinical significance.	At least 20 structures are identified at C1 level. At least 16 structures are identified at T5 level. At least 14 structures are identified at L3 level. At least 3 structures have clinical significance.	At least 18 structures are identified at C1 level. At least 16 structures are identified at T5 level. At least 14 structures are identified at L3 level. At least 2 structures have clinical significance.	Less than 15 structures are identified at C1 level. Less than 13 structures are identified at T5 level. Less than 11 structures are identified at L3 level. At least 1 structure has clinical significance.
Assessment method	At least 4 methods combined, eg: 1. lecture notes 2. textbook 3. atlas 4. cadaver laboratory	At least 3 methods used	At least 2 methods used	Only 1 method used
Groups are included	Group work	Group work	Individual work	Individual work
Timeline for completion	On time	On time	Not on time	Not on time