
ORIGINAL ARTICLE

A comparison of the efficacy of test-driven learning versus self-assessment learning

Xiaohua He, MD, MS and Anne Canty, PhD

Objective: We compared self-assessment and test-driven learning in two groups of students who studied the same subject.

Methods: This was a randomized comparative experimental study. The subjects were 259 first-quarter students who were divided into a test group and a self-assessment group based on the methods they used for their learning assessments. We measured the scores and difficulty levels of 3 formal written exams. Students' attitudes toward self-assessment or test-driven learning were surveyed.

Results: The mean scores of exam 1, exam 2, and a summative exam were 34 (± 6), 32 (± 8), and 44 (± 6) for the self-assessment group, respectively, with corresponding scores of 33 (± 6), 33 (± 7), 43 (± 6) for the test group. There were no significant differences in the mean scores on all 3 tests between the two groups ($p > .05$). Of the students in the self-assessment group, 64% scored at least 90%, whereas 47% of students in the test group answered at least 90% of the questions correctly ($p < .001$). For the survey, students expressed a positive attitude toward both learning strategies.

Conclusion: Both self-assessment and tests could have a significant impact on students' learning, but each offers different strengths and weaknesses.

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

J Chiropr Educ 2013;27(2):110–115 DOI 10.7899/JCE-13-6

INTRODUCTION

One of the major goals of teaching is to help students master new knowledge and skills and then help them develop strategies for long-term retention of the material.¹ To help students remember key facts, concepts, and knowledge, instructors may employ different strategies, including assignments, exams, tests, self-evaluation, and analysis. Among these strategies, tests (exams and quizzes) have been the most widely used strategy employed over time to assist students in their learning. Consequently, many educators are focusing on the development of testing as an evaluative technique and thus have been concerned with issues such as the effectiveness of the test, test reliability, and test validity.²

The benefits of testing are apparent. Numerous studies have shown that when students take a test on studied material, it results in better learning and retention than does additional studying of the target material.^{3,4} Other studies suggest that even when students are not provided with testing feedback, testing often offers a greater advantage than does restudying the material.^{5–7} This could be explained by the fact that testing does not merely enforce memory because of re-presentation of material, but rather because of a retrieval effect. Because of these benefits, testing-enhanced learning is currently the most popular method used in many schools, particularly in medical education.

The popularity of testing in health sciences education is obvious. Due to the nature of the board examinations of numerous health-care professions, tests and quizzes are indisputably the major means, if not the exclusive means, used by both instructors and students. Many major medical colleges provide online tests for students. Weekly tests have been adopted by many instructors as a major method of assessment. Students, on the other hand, also use tests as a means of learning. Research has already

This paper was selected as a 2012 Association of Chiropractic Colleges Research Agenda Conference prize-winning paper. The award funded by the National Board of Chiropractic Examiners.

© 2013 Association of Chiropractic Colleges

shown that in medical education, testing-enhanced learning can potentially strengthen clinical knowledge that will lead to improved expertise.⁸⁻¹⁰

However, it should be recognized that teacher-made informal and standardized tests may give the instructor information about student learning, but they do not provide all the information about how students learn. When students use test formats for their learning, it is considered passive learning. Teachers today are experimenting with alternatives to traditional tests because alternate forms of assessment can generate other information. For the last few years, we have been developing alternate forms of authentic student assessment strategies. The research evidence accumulating in our studies,¹¹ and the data produced by other researchers,¹²⁻¹⁴ make us optimistic about the impact of student self-assessment on students' academic learning and performance.

In contrast to test-enhanced learning, self-assessment allows students to judge the quality of their own work based on evidence and explicit criteria for the purpose of doing better work in the future. Our previous studies have shown that self-assessment is a potentially powerful technique because of its impact on student performance through enhanced self-efficacy and increased intrinsic motivation.^{11,15}

To answer our research question as to how these two methods could impact student learning, we started by comparing and contrasting the efficacy of self-assessment and test-driven learning in two groups of students who studied the same subject at our institution. We then moved to examine the difference between self-assessment and test-driven learning to explore the potential shortcomings and strengths of both self-assessment and test-driven learning. In addition, we surveyed students' attitudes toward either self-assessment or test-driven learning. We hope by addressing these issues we can better understand how different assessments could be used as a memory tool and how these assessment tools fit into the existing literature of medical education.

METHODS

Subjects

A convenience sample of 259 first-quarter students at Palmer College of Chiropractic, Florida Campus, Port Orange, Florida, participated in this study. The participants were from 6 different academic terms and were recruited during their anatomy class. All subjects were volunteers for this study. All students were informed that the results yielded from this study had no impact on their academic results. This study was approved by the institutional review board of Palmer College of Chiropractic.

Study Design

The design of this study was a randomized comparative experimental study. The subjects were assigned into 2 groups based on the academic terms: a self-assessment group (3 academic terms, 129 participants) and a test group (3 academic terms, 130 participants). The materials

used for the study incorporated content from the spinal anatomy course. The content of spinal anatomy included structural comparisons, concept description, material summarization, and clinical relevance of anatomical knowledge.

The instructor explained the method of practice to both groups. For example, when students in the self-assessment group were asked to do structural comparisons, they were given an example of structures for comparison. A rubric was also provided to the students for their reference (please see "Appendix," available online with this article). Similarly, when students in the test group were asked to take previously circulated tests to learn material, they were also provided these tests before taking the formal tests.

The students in the self-assessment group discussed their own strengths and weaknesses in learning anatomical materials and used the rubrics for self-assessment. The detailed process of the self-assessment group has been described in our previous studies,^{11,15} which, briefly, includes identifying key points in the rubric and underlining or circling 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 in the test group were provided tests corresponding to the same content used by the self-assessment group. They took the tests first, then were provided with the right answers, followed by permission to discuss the contents.

Clinical Case Discussions

The purpose of using clinical case discussions was to evaluate each student's ability to answer questions pertaining to clinical conditions based on their knowledge. This would demonstrate in-depth learning. Each group was assigned 32 case discussions. The cases were selected from the textbook *Clinical Oriented Anatomy*, 4th edition, by Keith Moore. The necessary knowledge for each case discussion was covered in the lectures. There were several questions listed at the end of each case discussion. Only the instructor had the standard answers, which were used to evaluate the students' understanding of anatomical structures and their clinical relevance. The percentage of correct answers was calculated after grading the answers to the questions for each case.

Mini Survey

Students in both groups were encouraged to take a mini survey at the end of study. The survey questionnaire is shown in Table 1. The purpose of the survey was to evaluate students' attitudes toward self-assessment or test-driven learning. The survey questions were validated by 3 faculty members and 3 students who did not participate in the study. The questions were revised upon the feedback from these members.

Statistical Method and Data Analysis

The first step of the data analysis was to investigate how well the convenience sampling produced groups with equivalent characteristics. The data groups were compared on the basis of race, gender, age distribution, and academic

Table 1 - The Percentage of Students in the Test Group vs the Self-Assessment Group (in Parentheses) Responding to the Survey

Questions	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
Overall, tests (self-assessments) are beneficial to my learning.	80 (82)	12 (8)	8 (10)	0 (0)	0 (0)
I am familiar with tests (self-assessments).	98 (16)	2 (24)	0 (40)	0 (30)	0 (0)
Tests (self-assessments) are convenient.	88 (3)	4 (15)	6 (12)	0 (63)	0 (13)
Tests (self-assessments) enhance my active learning.	21 (74)	31 (16)	24 (4)	12 (6)	6 (0)
Tests (self-assessments) enhance my passive learning.	77 (67)	19 (21)	2 (2)	2 (9)	0 (1)
Tests (self-assessments) help my grade.	68 (72)	13 (13)	11 (9)	8 (6)	0 (0)
Tests (self-assessments) help to gain knowledge.	76 (74)	8 (13)	12 (6)	3 (2)	1 (1)
Tests (self-assessments) help case discussions.	34 (77)	34 (10)	24 (12)	6 (1)	2 (0)
Tests (self-assessments) enhance critical thinking.	25 (76)	34 (16)	26 (6)	12 (2)	3 (0)
Tests (self-assessments) help me as an independent learner.	43 (70)	12 (26)	45 (4)	0 (0)	0 (0)

background. The Pearson χ^2 goodness of fit analysis was used.

The next step was to test the major research question of the study, which was whether the different learning mode had an effect on the students' academic performance. We measured the mean scores of 3 formal written exams (exam 1, exam 2, and the summative exam) as the result of the practice. The results of these exams were first analyzed and then compared between 2 groups. The questions on the exams, which had been given to several groups of students previously, were revised until there was minimal possible variation. There were 40 questions each on exam 1 and exam 2, and 50 questions on the summative exam, yielding a total score of 130 points. The exam questions incorporated three difficulty levels: terminology (level 1), concept (level 2), and application (level 3). Besides the comparison of total scores (percentage) of the exams, we also analyzed the potential disparity of difficulty levels of these exams between the two groups. The comparisons of exam scores and difficulty levels were analyzed using 1-way analysis of variance (ANOVA) with repeated measures to compare the results of the 3 written exams and case discussion. When significance was found, a post hoc analysis was performed for pairwise comparison. All statistics were performed using SPSS 15.0 software (SPSS Inc, Chicago, IL).

To obtain a psychometric instrument to measure the results from the mini survey, a Likert-type scale was chosen for the questions. This instrument was selected because it has been shown to be both reliable and valid, and interpretation of the results is straightforward.¹⁶ The Likert scale used in this study had five options: "Strongly agree," "Agree," "Not sure," "Disagree," and "Strongly disagree" (Table 1). Descriptive statistics were used to characterize percent response.

RESULTS

Demographics

The 2 groups were balanced on the basis of gender, age, ethnic group, and educational background distribution. The Pearson χ^2 showed there was no significant statistical difference between each category (Table 2).

Mean Scores of Exams

The 1-way ANOVA suggested that there were no significant differences in the mean scores on all 3 tests between the 2 groups ($p > .05$). The mean scores of exam 1, exam 2, and the summative exam are shown in Table 3.

Percentile Difference of Difficulty Levels

We pooled the questions of the 3 exams and calculated the percentage of students who answered on each difficulty level (Fig. 1). For example, the total number of level 3 questions in all 3 exams was 30 (100%). The number of students who answered 27 or above ($>90\%$) was calculated and compared between the groups. There were no statistically significant differences in performance of students on level 1 and 2 questions between the 2 groups ($p > .05$), whereas there was a statistically significant difference in the performance of students on the level 3 questions between two groups ($p < .001$; Fig. 1).

Table 2 - The Demographic Comparisons of Gender, Age, Race, and Educational Background Between Two Groups

Demographics	Test Group	Self-Assessment Group
Male	79	75
Female	51	54
Age between 20 and 30 years	115	119
Age above 30 years	15	10
Caucasian	104	105
Hispanic	8	7
African American	6	8
Other	12	9
Students with bachelor's degrees only	124	123
Students with master's degree	6	5
Students with doctorate degree	0	1
Prior anatomy course taken	125	124
Total	130	129

Chi-square (χ^2) showed no significant difference in gender, age, race, or educational background between the 2 groups (gender: $\chi^2 = 0.186$, $p = .666$; age: $\chi^2 = 1.065$, $p = .302$; race: $\chi^2 = 1.002$, $p = .80$; educational background: $\chi^2 = 1.091$, $p = .579$).

Table 3 - The Results of Exams Between the Test Group and the Self-Assessment Group

Group	Exam 1 (40)	Exam 2 (40)	Summative (50)
Test	34 (± 6)	32 (± 8)	44 (± 6)
Self-assessment	33 (± 6)	33 (± 7)	43 (± 6)
Statistics	$p > .05$		

There are no significant differences between the two groups.

Percentile Difference of Case Discussion

Each group had 32 case discussions. We used the standard answers of case discussions provided by the textbook as the assessment criteria. Overall, the self-assessment group met a higher percentage of criteria than did the quiz group (Fig. 2).

Mini Survey Results

The percentage of answers to the questions from the questionnaire are listed in Table 1. The majority of students in both groups agreed that the intervention they used in their studies helped them in their learning. In contrast to the students in the self-assessment group, students in the test group overwhelmingly agreed that tests were convenient. In contrast to the students in test group, the majority of students in the self-assessment group strongly agreed that self-assessment enhanced their critical thinking.

DISCUSSION

This study has shown several interesting findings. Students expressed a positive attitude toward both learning strategies as the survey results showed. The majority of students in both groups agreed that both self-assessment and quizzes helped their learning. This was supported by the lack of significant differences in exam results between groups. However, the study also revealed

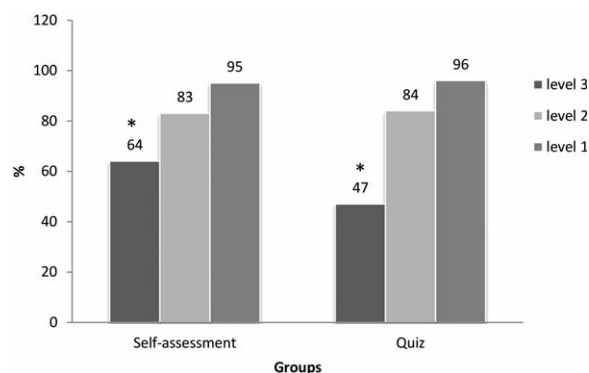


Figure 1 - Percentile difference of difficulty levels between groups. This bar graph shows the percentages of students in their academic performance. There were no significant differences between the two groups of either 80% and 90%. There was a significant difference between the two groups below 70%.

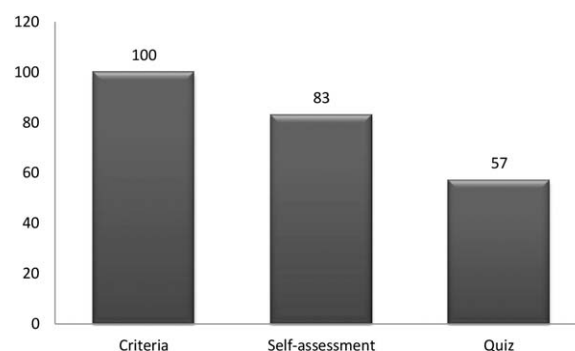


Figure 2 - This bar graph illustrates the percentage of students who accurately answered the questions in case discussions. The criteria are the answers provided by the textbook.

that the benefits of these two methods may have different strengths.

Convenience vs Time Consuming

One of the advantages of test-enhanced learning is convenience. From the survey of the students in the test group, 92% of the students strongly agree or agree that the tests are convenient. In contrast, only 18% of students in the self-assessment group agree that self-assessment is convenient. Self-assessment students expressed that it took time to become familiar with the format of self-assessment and also that it was time-consuming to perform the self-assessment.

It has been long recognized that tests are convenient and flexible. Almost 100% of students have experience in taking quizzes during their prior academic years. Because of this, tests have a huge advantage for students as one of their primary methods of learning. Most colleges even provide online test banks that are easily accessible to students. On the other hand, because most of the students in the self-assessment group did not have prior experience in formal self-assessment, some described it as “a big headache” initially.¹¹ From the results of this study, we can conclude that self-assessment is more complex and time-consuming and also more costly in terms of time and effort when compared to tests.

Superficial vs In-Depth Learning

It was noticed that although there were no significant differences in average scores on exams between the two groups, there was a significant difference in terms of difficulty level. When we analyzed and compared the difficulty levels, we found that more self-assessment students vs test students answered at least 90% of level 3 questions correctly. In addition, self-assessment students also performed better on case discussions. An abundance of evidence from literature has demonstrated that taking tests can improve one's knowledge.^{7,8} A similar amount of evidence from the literature has also demonstrated that self-assessment is a powerful learning tool in helping students gain knowledge. One of the advantages of self-assessment is that students develop critical thinking skills through the process of performing self-assessment. During

self-assessment, students reflect on the quality of their work on the assignments, judge the degree to which it reflected explicitly stated goals or criteria, and revise their work accordingly.¹¹ Students in the self-assessment group do not just learn the answers to the questions, they put the information into a larger context, making sure that they understand the connection between self-assessment and the goal of better learning and retention of course material.

On the other hand, when students use tests as a learning tool, they are more focused on the specific answers, the styles of questions, and the amount of time spent on the test. Test practices can also improve students' test-taking skills, such as choosing the best guess to answer the questions when the correct answers are not apparent. Such strategies may help students' test grades. However, under such a learning model, some may ignore the deeper content of the material, that is, the application of the knowledge. They may not develop as strong an ability as do students in the self-assessment group in terms of analyzing knowledge learned in relation to clinical application. The results of our study, particularly the discrepancy between the groups in results of case discussions, could be due to the different learning models used by the two groups.

Student-Centered Learning vs Passive Learning

The remarkable consistency in the survey regarding students' confidence in relation to self-oriented learning or passive learning is an indicator that the questionnaire did yield reproducible results of existing trends in student perceptions of their own abilities.

Our study showed that over 90% of the students in the self-assessment group developed confidence in their abilities to become independent learners after classroom teaching; yet, that number is only slightly over 50% for the students in the test group. The relatively high levels of confidence expressed by the students in the self-assessment group may be related to the improvement in self-assessment skills in their study. The lower confidence levels in the test group may be a reflection of the deficiency of practical experience in the assessment of learning materials.

Limitations

The short time duration of this study is one of the limitations. This study was carried out over a period of 11 weeks (1 academic quarter). Multiple quarter results could be more revealing. Another limitation was that this study was performed only as it pertained to the teaching of anatomy because that was the content chosen for this study. Multiple subject studies are needed, such as biochemistry and physiology, which may show a broader spectrum of results. Although the survey demonstrated that students in the self-assessment group had higher confidence in their self-learning, problems with inaccuracy in self-assessments should not be ignored. This includes students who inaccurately assess themselves and who tend toward overestimating their abilities.¹⁷ Not surprisingly, testing is still a more feasible way to assess students' academic performance than is self-assessment.

CONCLUSION

The authors conclude that both self-assessment and tests could have a significant impact on students' learning, but each offers different strengths and weaknesses. Utilizing a test bank of questions for studying is convenient and not time consuming; however, the development of critical thinking skills is sacrificed. Whereas, even though performing self-assessment takes more time and effort, it is an effective learning strategy because it offers students an opportunity to become critical thinkers, which will ultimately allow them to become independent learners.

CONFLICT OF INTEREST

There were no external sources of funding for this study, and no conflicts of interests were identified within this investigation.

About the Authors

Xiaohua He and Anne Canty are both professors at the Palmer College of Chiropractic, Florida. Address correspondence to Xiaohua He, 4777 City Center Parkway, Port Orange, FL 32129; shawn.he@palmer.edu. This article was received April 14, 2012, revised August 13, 2012 and December 31, 2012, and accepted January 13, 2013.

© 2013 Association of Chiropractic Colleges

REFERENCES

1. Pashler H, Bain PM, Bottge BA, et al. *Organizing Instruction and Study to Improve Student Learning*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Dept of Education; 2007. NCER 2007-2004. <http://ies.ed.gov/ncee/wwc/pdf/practiceguides/20072004.pdf>.
2. McDaniel MA, Roediger HL, McDermott KB. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev*. 2007;14(2):200–206.
3. Hogan RM, Kintsch W. Differential effects of study and test trials on long-term recognition and recall. *J Verbal Learning Verbal Behav*. 1971;10:562–567.
4. Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci*. 2006;17:249–255.
5. Carpenter SK, DeLosh EL. Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. *Mem Cognit*. 2006;34:268–276.
6. Carrie M, Pashler H. The influence of retrieval on retention. *Mem Cognit*. 1992;20:633–642.
7. McDaniel MA, Masson MEJ. Altering memory representations through retrieval. *J Exp Psychol Learn Mem Cogn*. 1985;11:371–385.

8. Larsen DP, Butler AC, Roediger HL. Test-enhanced learning in medical education. *Med Educ*. 2008;42(10):959–966.
9. Galvagno SM, Segal BS. Critical action procedures testing: a novel method for test-enhanced learning. *Med Educ*. 2009;43(12):1182–1187.
10. Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ*. 2004;38:974–979.
11. He XH, Canty A. Student self-assessment: what research says and what practice shows. *J Chiropr Educ*. 2010;24(1):95.
12. Weiss PM, Koller CA, Hess LW, Wasser T. How do medical student self-assessments compare with their final clerkship grades? *Med Teach*. 2005;27(5):445–449.
13. Andrade H, Du Y. Student responses to criteria referenced self-assessment. *Assess Eval High Educ*. 2007;32(2):159–181.
14. Wanigasooriya N. Student self-assessment of essential skills in dental surgery. *Br Dent J*. 2004;197:11–14.
15. He XH, Canty A. Empowering student learning through rubric-referenced self-assessment. *J Chiropr Educ*. 2012;26(1):24–31.
16. Likert R. A method of constructing an attitude scale. In: Maranell GM, ed. *Scaling: A Sourcebook for Behavioral Scientists*. Chicago, IL: Transactions Publishers; 2007:233–243.
17. Liu EZ, Lin S, Yuan S. Alternatives to instructor assessment: a case study comparing self and peer assessment with instructor assessment under a networked innovative assessment procedure. *Int J Instr Media*. 2002;29(4):395–404.