Association Between Dental Student-Developed Exam Questions and Learning at Higher Cognitive Levels

Carlos Gonzalez-Cabezas, DDS, MSD, PhD; Olivia S. Anderson, PhD; Mary C. Wright, PhD; Margherita Fontana, DDS, PhD

Abstract: New dental accreditation standards emphasize that graduates must be competent in the use of critical thinking (a high cognitive-level skill). Despite this new standard, most written assessments in dental school courses are still based on low cognitive-level questions. The aim of this study was to determine if an exercise that allows students to collaboratively write exam questions would help cultivate higher cognitive levels of learning. To evaluate this exercise at one U.S. dental school, the cognitive level (according to Bloom's taxonomy) of multiple-choice exam questions and students' scores across two cohorts in a cariology course were compared. This evaluation took place using a control group in which questions were instructor-generated and an intervention group in which students worked in groups to develop questions. All students in one first-year class participated in the intervention group (n=104); all students in the first-year class two years earlier served as the control group (n=106). Among students in the intervention group, the response rate to a post-intervention survey measuring students' attitudes about the experience was 70% (N=73). The results showed that the students generating their own assessments developed higher cognitive-level exam questions than the instructor-generated assessments. The intervention group (with student-generated assessments) also performed as well or better on tests compared to the control group (with instructor-generated assessments). In the intervention group survey, the vast majority of students agreed that the exercise was helpful for their overall learning experience, but working in teams was said to be the least valuable component of the activity for their learning. This study suggests that student-driven, collaborative assessments can be an important tool for building critical thinking skills in dental classrooms and that it may be worthwhile to expand this type of exercise into other courses.

Dr. Gonzalez-Cabezas is Associate Professor, School of Dentistry, University of Michigan; Dr. Anderson is Clinical Assistant Professor, School of Public Health, University of Michigan; Dr. Wright is Director of Assessment and Associate Research Scientist, Center for Research on Learning and Teaching, University of Michigan; and Dr. Fontana is Professor, School of Dentistry, University of Michigan. Direct correspondence to Dr. Carlos Gonzalez-Cabezas, School of Dentistry, University of Michigan, 1011 N. University Ave., Room 2395, Ann Arbor, MI 48109-1078; 734-763-3391; carlosgc@umich.edu.

Keywords: dental education, educational assessment, test taking skills, testing, critical thinking

Submitted for publication 10/7/14; accepted 4/29/15

Testing is one of the most effective ways to promote student learning.[1-3] The Commission on Dental Accreditation (CODA) suggests in its predoctoral dental accreditation standards that a key way to build critical thinking is through "use of questions by instructors that require students to analyze problem etiology, compare and evaluate alternative approaches, provide rationales for plans of action, and predict outcomes" (p. 22).[4] Critical thinking was defined by Hendricson et al. as "the reflective process in which individuals assess a situation or evaluate data by using mental capacities characterized by [verbs] such as compare, analyze, distinguish, reflect, and judge."[5] Therefore, to effectively assess if dental students can demonstrate critical thinking skills, instructors must make an effort to develop assessments that allow students to develop analysis and judgment, the skills frequently referred to as "higher-level thinking," a reference to Bloom's taxonomy.[2] Assessments based on low cognitive-level questions—such as recall—do not help students develop the kind of learning they will need to demonstrate as practitioners.[6]

Multiple-choice questions are one of the most frequently used assessment methods in dental schools.[7] However, many of those assessments are still based on low cognitive-level questions (i.e., representing students' ability to understand and remember) versus higher cognitive-level questions (i.e., representing students' ability to evaluate and synthe-

size).[2,7] This problem is not limited to dentistry since studies show similar low cognitive-level exam questions in other fields.[8,9] A key challenge in developing complex assessments is the difficulty instructors report in developing higher cognitive-level questions.[9] Another significant obstacle is that multiple-choice question formats pose more acute challenges for measuring higher-level thinking skills.[10,11]

This study examined an innovative student-centered method to develop exams with questions at higher cognitive levels to enhance learning. Exam questions and scores were compared across two years. In one group (control), instructors wrote the questions. In the other group (intervention), students contributed to question development. Our research questions were as follows: 1) Are student-generated exam questions associated more with higher cognitive-level assessments than instructor-generated questions? 2) Are student-generated exam questions at a higher cognitive level associated with similar or improved performance than achieved by students tested with lower cognitive-level questions? 3) Do students perceive that collaborative production of exam questions is helpful to their learning?

In a dental education setting, there have been no prior published studies about student item generation. Two key studies outside of dentistry documented positive outcomes from the pedagogical practice, but they had some methodological limitations. Fellenz described a similar activity, the Multiple Choice Item Development Assessment (MCIDA), in a small business school course.[12] Those students participating in the MCIDA reported that the exercise improved their understanding of key course concepts and served as a valuable learning experience. However, that study did not examine direct measures of learning outcomes, such as student performance on key course assessments, and it did not include a comparison group. Others examined a student-generated question exercise in medical school and found a decrease in students' self-reported testing anxiety but mixed results about its impact on their exam performance.[13] However, that study did not control for cognitive level of the assessment questions across cohorts.

The aim of our study was to use a quasi-experimental design to assess student learning outcomes through direct measures of exam performance, accounting for student background and cognitive complexity of the assessments. The study also sought to understand the students' perceived value of the exercise and to isolate components of the activity said to be more or less useful to the students.

Materials and Methods

Permission was granted by the University of Michigan's Institutional Review Board to conduct this study. The course used in the study was the Cariology II course. The Cariology I (fall term) and II (winter term) courses are taught in the first year of the DDS curriculum to establish didactic foundational knowledge (through both traditional lecture formats and online content delivery), caries detection skills development (through hands-on laboratory and clinical exercises), patient case discussions, and additional clinical experiences to enhance and facilitate active student learning, critical thinking, problem-solving, and use of evidence-based information for dental caries detection, diagnosis, risk assessment, prevention, and management. The goal is to prepare students to be able to perform these skills (dental caries detection, diagnosis, risk assessment, prevention, and management) during clinical care throughout their dental education (years 1-4 of dental school), after graduation, and as lifelong learners. During the two courses, these skills are assessed with a combination of examinations, practical exercises, short clinical decision making papers, and development of evidence-based treatment plans in patient scenarios (first individually and then challenging students in groups to reach team-based consensus). Assessment is geared towards having students apply knowledge and critically think through possible solutions, using best available evidence, to solve clinical problems and justify their answers.

All students who participated in the student-generated exam exercise represented the intervention group (n=104) for the study. Exams for this group consisted primarily of questions that were student-generated, with very few faculty-generated questions. The control group was comprised of students enrolled two years earlier, who did not complete the exercise and instead completed instructor-generated exams (n=106).

Student-Generated Exam Question Assignment

In the intervention year (2013), students were assigned to groups of three. Each student group was randomly given a lecture from which they were to generate three multiple-choice exam questions based on a clinical scenario they had to generate. They could use images used in class or downloaded

legally from the Internet. They developed an additional multiple-choice question from a clinical/lab exercise and another from previous knowledge (gained in the Cariology I course), resulting in a total of five student-generated exam questions from each group. The instructors (CGC and MF) awarded extra credit to groups generating questions at high cognitive levels based on Bloom's taxonomy, and they also gave feedback to teams (e.g., about erroneous answers) for the students to enhance the items.[2] The student-generated exam questions were posted to a Google document that was accessible to the whole class and the faculty to read, make comments on, and edit. Student groups performed this exercise for both the midterm and final exams. (The instructions given to students in the course syllabus, which were reinforced verbally in class, are available from the corresponding author.)

For the course exams, the instructors made minor modifications to the student-generated questions, which frequently resulted in a different answer and encouraged students to move beyond a surface-level understanding of the item to be able to answer it correctly. (For example, a student-generated question about treatment planning for a pediatric patient might be altered to focus on a geriatric patient, thereby shifting the correct answer choice.) The students were warned about this possibility before the exams so they could prepare adequately for it and to discourage them from memorizing questions and answers. In the intervention group, nearly all questions (>90%) were student-generated (with slight instructor modifications), but because the student questions did not cover all course objectives, the instructors added one or two items on each midterm and final to address the gaps.

Rating the Exam Questions

All of the exam questions (total N=160) from the intervention (student-generated questions slightly modified by instructors) and control (instructor-generated questions) groups were rated blindly and independently by three expert scorers. The analysis included all exam questions used in each course: 83 questions for the intervention group (45 on the midterm and 38 on the final) and 77 questions for the control group (45 of the midterm and 32 on the final). Two scorers (OA and MW) were assessment professionals at the university's teaching center but had no educational training in dentistry. The third scorer (MF) was an instructor for the course.

The exam questions were given a cognitive score based on a modified Bloom's taxonomy. In this process, it was necessary to concurrently evaluate the test item in the context of the pedagogy used. For example, a test question on evidence about xylitol may measure a concept at a low cognitive level if students needed to recall facts given directly by the instructor in class, but it may be scored at a higher level if students needed to make more inferences about the evidence. Therefore, it was necessary to determine a consensus score collectively and have all three scorers come to agreement about a score for each question. Therefore, overall interrater reliability statistics were not computed; however, there was 88% agreement between the two assessment professionals' initial ratings.

The scorers rated the questions as follows: Level 1=low cognitive level (measuring recall, knowledge, and comprehension); Level 2=medium cognitive level (measuring clinical application and analysis); Level 3=high cognitive level (measuring evidence-based decision making, synthesis, and evaluation). Table 1 shows examples of student-generated questions rated at each level by the scorers. Following the blinded scoring, the experts met to reach a forced agreement for questions with different ratings. Within a month, 10% of the questions were rescored to assess the level of repeatability between the first and second ratings, and a high degree of consistency was found (weighted kappa=0.88). A kappa with linear weights was used in which a weight of 1.0 was used for total agreement, 0.5 was given to adjacent levels, and 0 was assigned to categories spaced two levels apart (or w=1- [i/(k-1)], where k=number of categories [3] and i=difference in raters' categories). Finally, we designed a post-intervention survey to capture the intervention group's perceptions of their learning gains and experience with the exam writing exercise. (The survey is available from the corresponding author.)

Statistical Analysis

Information collected about the students in the intervention and control groups included entering Dental Admission Test (DAT) scores, grade point averages (GPAs), and winter semester and D1 year GPAs. These data were compared between the two groups using a Mann-Whitney Rank Sum Test. The cognitive level of the midterm exam questions was compared to the cognitive level of the final exam questions for both the student-generated and instructor-

Table 1. Sample student-generated test questions and their ratings

A) Example of a test question rated 3 (high cognitive level) because of need for evidence-based decision making:

An upset father marches into your dental office with his 12-year-old daughter complaining that she has white stains on her front teeth. He was watching a segment on ABC News about water fluoridation and decided that the fluoride in the water is causing this discoloration. He is concerned now that her teeth are defective and the white spots will keep getting worse every time she drinks water. Upon examination, you determine that the child has mild fluorosis on #8 and 9 that is limited to the corner of the incisal edge of each tooth. What would you tell the father in response to his concern?

I. Fluorosis can only occur during the development of the permanent teeth. Thus, now that her teeth have erupted, her fluorosis will not get more severe by drinking water.

II. The function and strength of her teeth are not compromised because of the fluorosis.

III. His daughter should limit her tap water intake because with more fluoride exposure she could develop more severe fluorosis on those teeth.

a. II only
b. I and II
c. I and III
d. I, II, and III

B) Example of a test question rated 2 (medium cognitive level) because of need for analysis:

After measuring salivary flow using the Schirmer strip, the result is <15mm in 3 min. You proceed to measure flow rate using the drainage method. Please assess the statements below about this method.

Statement 1: The Unstimulated Salivary Flow Rate Test is generally performed for 5 minutes.

Statement 2: You would expect the saliva sample collected during this test to have a water consistency.

a. Both statements are true.
b. Both statements are false.
c. The first statement is true. The second statement is false.
d. The first statement is false. The second statement is true.

C) Example of a test question rated 1 (low cognitive level) because information was based on recall from lecture presentation:

What is the most common side effect of using xylitol when used in high amounts?

a. Excessive sweating
b. Bitter aftertaste
c. Increased caloric intake
d. Intestinal discomfort

generated sets of questions using a Mann-Whitney Rank Sum Test. Students' performance on exams in the intervention group were compared to the performance of students in the control group also using a Mann-Whitney Rank Sum test. The perceptions of the helpfulness of the exercise for students in the intervention group were analyzed using the Friedman test. Following the Friedman test, each item was compared using a Wilcoxon Paired Rank test with non-parametric Bonferroni post-hoc tests and additional descriptive statistics. Results were considered statistically significant for a p-value less than 0.05.

Results

All students in the two first-year classes participated in activities of the intervention group (n=104) or the control group (n=106). Among students in the intervention group, the response rate to the post-intervention survey was 70% (N=73).

Students in the intervention group had significantly higher mean DAT scores than the control group, although the difference was small (0.7) (Table 2). There was also a significantly higher mean score on the perceptual ability section of the DAT for the intervention group than the control group, but again with a small difference (0.7). There were no significant differences (all $p>0.05$) in entering GPA, first-year dental GPA, and first-year winter semester GPA between students in the intervention and control groups.

Cognitive Level of Student-Generated Questions

Of the student-generated exams, over two-fifths (42.2%) of the exam items were found to assess high-level cognitive skills (Table 3). In comparison, a very small proportion (15.6%) of the instructor-authored items measured skills such as evidence-based decision making, synthesis, and evaluation. However,

Table 2. Grade point averages (GPAs) and Dental Admission Test (DAT) scores of intervention and control groups

GPAs/Scores	Control Group Mean (SD) (n=106)	Intervention Group Mean (SD) (n=104)
Entering GPA	3.47 (0.32)	3.56 (0.26)
DAT	19.6 (1.53)	20.3 (1.66)*
DAT: perceptual ability	19.9 (2.29)	20.6 (2.00)*
1st-year dental GPA	3.47 (0.31)	3.50 (0.36)
1st-year winter semester GPA	3.41 (0.37)	3.47 (0.39)

*Statistically significant at $p<0.05$

Table 3. Percentage (number) of exam items at each cognitive level

	Total		Final Exam		Midterm Exam	
Cognitive Level	Student-Generated	Instructor-Generated	Student-Generated	Instructor-Generated	Student-Generated	Instructor-Generated
High	42.2% (35)	15.6% (12)	55.3% (21)	6.3% (2)	31.1% (14)	22.2% (10)
Medium	14.5% (12)	22.1% (17)	7.9% (3)	15.6% (5)	20.0% (9)	26.7% (12)
Low	43.4% (36)	62.3% (48)	36.8% (14)	78.1% (25)	48.9% (22)	51.1% (23)

there was some variation by type of exam. Students in the intervention group developed final exam questions at a significantly higher mean cognitive level (M=2.18 out of 3.00, SD=0.96) than the instructor-generated final exam questions (M=1.28, SD=0.58) ($p<0.001$) (Figure 1). Although the student-generated midterm had a slightly higher average cognitive level (M=1.84. SD=0.89) than the instructor-generated test (M=1.71, SD=0.82), the difference was not statistically significant.

Interestingly, a second trend was observed in the change in level of questions over time. For the student-generated exams, the cognitive-level rating increased from midterm (M=1.84) to final exam (M=2.18), but this change was not statistically significant. In contrast, the mean level of instructor questions significantly decreased over time (from 1.71 to 1.28) ($p=0.01$).

Student Exam Performance

The students in the intervention group performed at a significantly higher level on the midterm exam (average score 86.6%) compared to the students in the control group taking the instructor-generated exam (average score 82.4%) ($p<0.001$) (Figure 2). Additionally, the students in the intervention group performed higher on the final exam (average score 89.0%) than did the control group (average score 86.3%), but this difference was not statistically significant.

Students' Perceptions of the Exercise

To assess how the students perceived the process of student-centered exam creation, students in the intervention group were surveyed. A majority (79%) of the responding students agreed that the exercise was helpful for their overall learning experience, over three-quarters (77%) that it helped them on exams, and most (73%) that the assignment enhanced their critical thinking skills. The students also praised the integrative functions of the pedagogical activity: 80% agreed that it helped them apply ideas from lecture to clinic, and 81% reported that it assisted them with making connections between ideas learned in Cariology I and other courses.

Positive student feedback was also reflected by comments on the post-course survey. One student stated, "I think that this exercise was the best learning tool I've seen here. I really learned well having to

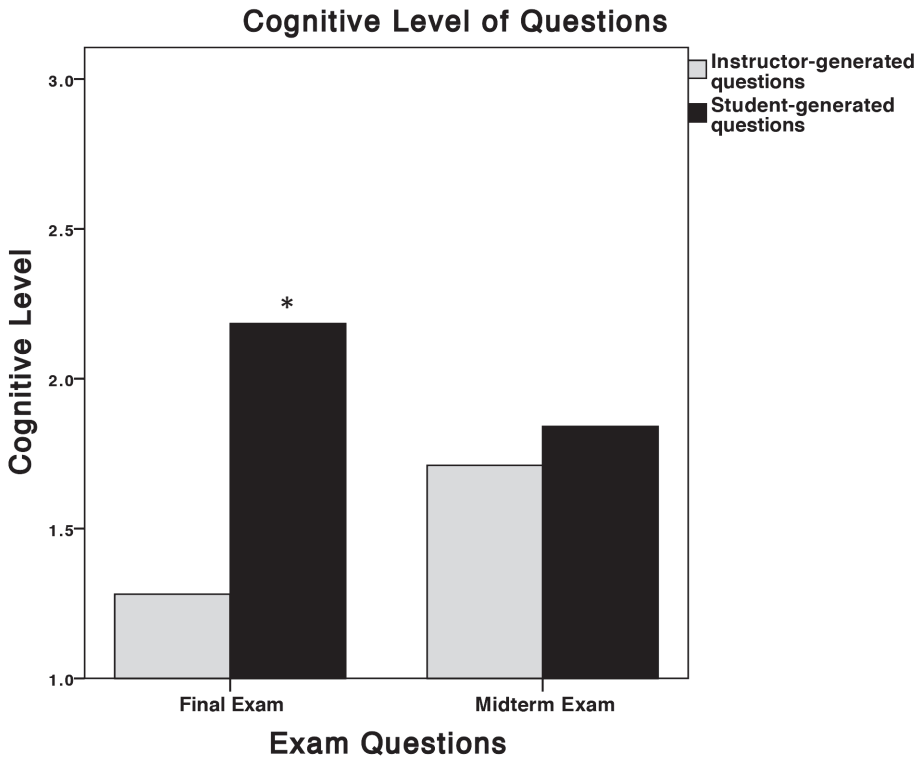

Figure 1. Mean cognitive-level scores of exam questions generated by instructors and students

*Statistically significant at $p<0.05$

critically analyze the material myself in order to make questions." Another student affirmed, "This exercise forced me to evaluate questions and review why they were right and wrong, instead of just taking a test and never getting to see the test and both understand why the answer choices are right or wrong based on my study of the material and the teacher feedback."

To understand the relative perceived usefulness of various components of the exercise, students were asked to rate the value for their learning of key aspects of the exercise: working in teams, use of a collaborative web-based tool (Google Docs), the opportunity for instructor feedback, the chance to practice exam questions, and the opportunity for extra credit. Of these components, working in teams was the only one rated significantly differently ($p=0.009$) in comparison to use of Google Docs and the opportunity for instructor feedback. Additionally, the students rated working in teams as the least valuable aspect for their learning (M=2.5 on a scale from 1=least valuable to 5=most valuable) (Table 4).

The students were also asked for their suggestions to enhance the exercise for future classes. Key suggestions were to improve the student group dynamics by assigning fewer questions per group and to allow each group to generate questions from multiple lectures instead of only one. Another common suggestion was to have the instructors provide students with continuous feedback throughout the semester. One student explained, "I really found the feedback from the instructors extremely helpful . . . especially if it is one of my questions. I would recommend more feedback from instructors." Another common suggestion was for the instructors to generate clearer guidelines for generating exam questions and to provide more examples for the students.

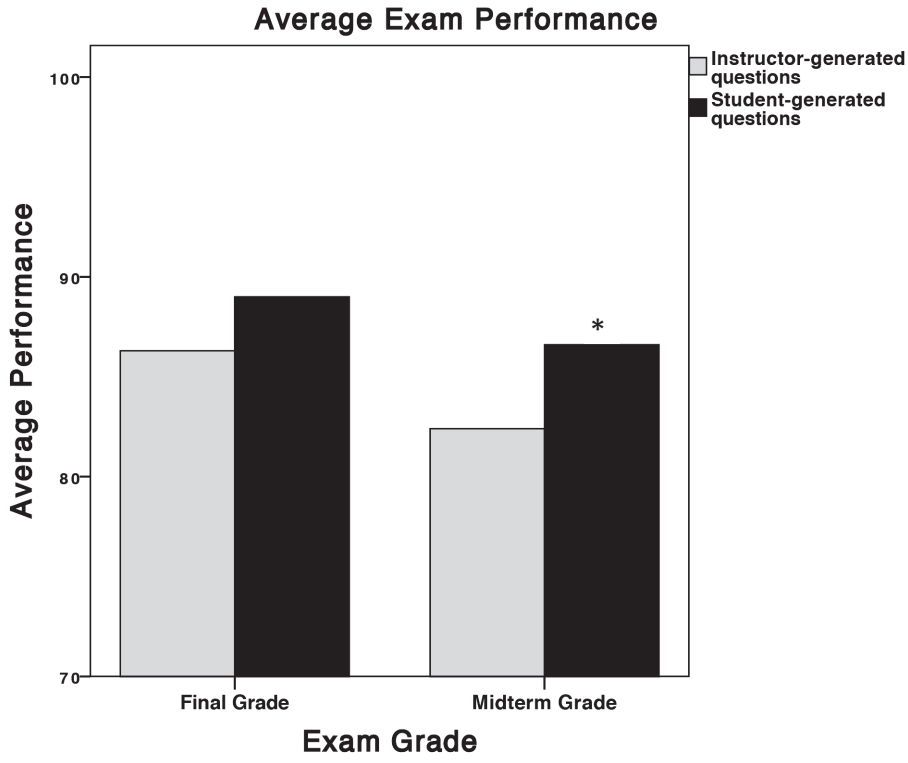

Figure 2. Students' mean performance on final and midterm exams

*Significant difference between the intervention group (student-generated questions) and control group (instructor-generated questions) ($p<0.05$)

Discussion

In this study, the cognitive levels of student-generated final exam questions were found to be significantly higher compared to instructor-created assessments (Figure 1). Despite this increased level of cognitive complexity, students performed at an equivalent level on the exams (Figure 2). Although cognitive complexity is not equivalent to difficulty,[9] the higher cognitive-level questions did not diminish student performance, suggesting that the exercise aided in increased learning.

There was no statistically significant difference between the cognitive levels of the two midterms. Because the final student-generated exams were found to be more cognitively complex than the instructor-generated ones, it could be that students benefited from instructor feedback and learning from previous groups' attempts. These results suggest that the exercise may have cultivated students' critical thinking skills as a result of peer discussion and communication through group work. Instructors are therefore advised to continue the exercise throughout a term or year, rather than as a one-time assignment. However, in spite of the equivalent levels of complexity, students in the intervention group scored significantly better on the midterm than those in the control group, again suggesting enhanced learning.

One limitation of the study is the possibility that these differences were due to groups' significantly different entering DAT scores although their entering, first-year, and winter semester GPA scores were not different (Table 2). However, this DAT score difference was very small and considering that college GPA has generally been found to be the best

Table 4. Students' perceived value for their learning of exercise components, by percentages of total and mean

	Least Valuable 1	2	3	4	Most Valuable 5	Mean
Working in teams	31.5%	23.3%	19.2%	15.1%	11.0%	2.51*
Using Google Docs	19.2%	21.9%	28.8%	20.5%	9.6%	2.79
Getting extra credit	20.5%	15.1%	15.1%	23.3%	26.0%	3.16
Getting instructor feedback	17.8%	15.1%	21.9%	23.3%	21.9%	3.19
Seeing exam questions before taking exam	11.0%	24.7%	15.1%	17.8%	31.5%	3.34

*Statistically significant at $p<0.05$ in comparison to items using Google Docs and getting instructor feedback

predictor of dental school academic performance,[14] it is very unlikely that the differences observed were due to the groups' not being comparable. Another possible limitation of the study is that the students were not randomly assigned to the intervention and comparison groups (who took different tests); therefore, there may have been unobserved differences between the two populations that would account for differential performance.

The students' perceptions of the exercise were positive overall. However, the students found team-based learning to be the least valuable aspect of the exercise. Although the assignment could be completed individually, many instructors wish to foster collaborative skills as a key pedagogical goal. Key recommendations made by students to improve the group process were assigning fewer questions per group and allowing groups to generate questions from multiple lectures instead of one. Instructors may also find it helpful to consult resources that offer recommendations about how to help teams function well, such as deliberate group assignment and peer evaluation.[15] Future research could test if there are differential rates of learning, comparing individually written questions with team-authored assessments. Additionally, because one limitation of this study is that we used a non-validated instrument for the post-intervention survey, future researchers may wish to explore this specific question with a validated instrument on team-based functions.[16]

A testing effect (having students practice questions first before answering them again on the exam) could potentially explain the increased performance on the midterm exam questions.[17,18] This effect would not necessarily be a negative explanatory factor because practice testing has been identified as one of the most powerful study techniques for student learning.[13] It could be a possible drawback if students memorize items seen in advance, a practice that would reflect lower-level Bloom's taxonomy learning outcomes (e.g., recall). Indeed, others have found that a possible drawback of student-centered test generation is that students will memorize items seen in advance, limiting possible gains in cognitive development.[13]

However, as shown in Table 4, many students did not perceive this to be the case, with only about a third (31.5%) rating this aspect of the exercise as the most valuable component. As one student stated, "This is a great way to learn. I really enjoyed it and it helped me focus on learning the material, not on what they will ask questions about." Furthermore, the instructors made minor modifications to the student-generated questions for the actual exams, and the students were warned about this possibility beforehand. Both student perceptions and instructor practice suggest that any testing effect may not necessarily be a drawback to the collaborative assessment approach (i.e., performance benefits would be attributed to learning processes other than memorization). Nonetheless, future research might address the question of a possible testing effect more conclusively by using a research design that compares accuracy rates for student-generated questions that have not been distributed to an entire class (e.g., accuracy for authors of questions as compared to correctness for non-authors).

Why were these students more successful at creating higher level exam questions than the expert instructors? We can only speculate, but it could be that the conditions under which the students wrote the questions (collaborative generation, with guidance and feedback from instructors) were very different from conditions in which many faculty typically write exam questions (individually, with little feedback from others). Indeed, in a faculty development

workshop for dental faculty with peer feedback and collaborative learning, instructors were better able to design learning activities and assessments that deepened students' use of critical thinking skills than they were alone.[19] Another possible explanation is that experts' difficulty with articulating tacit knowledge is well documented,[20] so it may be that novices are better able to identify new applications and evaluative perspectives for concepts.

Overall, the students generated higher cognitive-level exam questions than the instructors and performed well on assessments, suggesting that student-driven, collaborative assessments are an important tool for building critical thinking skills in dental classrooms. Given the documented challenges of instructors' development of high-level assessment items, this approach offers another vehicle for generating multiple-choice questions with improved levels of cognitive complexity.[21] Other ideas for fostering dental students' critical thinking skills in the classroom include using interactive pedagogies that ask students to make predictions and synthesize data,[22,23] teaching formal decision analysis,[21,24] giving writing assignments (e.g., written critique of an advertisement),[25] and asking students to think aloud as they solve problems.[26]

Conclusion

This study sought to determine if an exercise that allows students to collaboratively write exam questions would help cultivate higher cognitive levels of learning. The results showed that students who generated their own exam questions developed higher cognitive-level questions than the instructor-generated ones. The intervention group (with student-generated assessments) performed as well or better on the exams than the control group (with instructor-generated assessments). The data presented here support the expansion of this exercise into other predoctoral dental classroom experiences and offer another assessment approach that is viable for larger classrooms.

Acknowledgments

We would like to gratefully acknowledge the Roy H. Roberts Dental Education Innovation Awards for financial support of this research. Thanks to Dr. Ronit Greenberg for literature review assistance and Ms. Susan Flannagan for organizing and managing the data.

REFERENCES

1. Henzi D, Davis E, Jasinevicius R, Hendricson W. In the students' own words: what are the strengths and weaknesses of the dental school curriculum? J Dent Educ 2007;71(5):632-45.
2. Bloom BS. Taxonomy of educational objectives: the classification of educational goals. Handbook 1, cognitive domain. New York: David McKay Co., 1956.
3. Dunlosky J, Rawson KA, Marsh EJ, et al. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. Psychol Sci Public Interest 2013;14(1):4-58.
4. Commission on Dental Accreditation. Accreditation standards for dental education programs. Chicago: American Dental Association, 2013.
5. Hendricson WD, Andrieu SC, Chadwick DG, et al. Educational strategies associated with development of problem-solving, critical thinking, and self-directed learning. J Dent Educ 2006;70(9):925-36.
6. Oliver R. Curriculum structure: principles and strategy. Eur J Dent Educ 2008;12:74-84.
7. Albino JE, Young SK, Neumann LM, et al. Assessing dental students' competence: best practice recommendations in the performance assessment literature and investigation of current practices in predoctoral dental education. J Dent Educ 2008;72(12):1405-35.
8. Momsen JL, Long TM, Wyse SA, Ebert-May D. Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. CBE Life Sci Educ 2010;9(4):435-40.
9. Momsen J, Offerdahl E, Kryjevskaia M, et al. Using assessments to investigate and compare the nature of learning in undergraduate science courses. CBE Life Sci Educ 2013;12(2):239-49.
10. Piontek M. Best practices for designing and grading exams. CRLT Occasional Paper no. 24. Ann Arbor, MI: Center for Research on Learning and Teaching, 2008.
11. Haladyna TM. Developing and validating multiple choice items. 2nd ed. Mahwah, NJ: L. Erlbaum Associates, 1999.
12. Fellenz MR. Using assessment to support higher-level learning: the multiple choice item development assignment. Assess Eval Higher Educ 2004;29(6):703-19.
13. Papinczak T, Peterson R, Babri AS, et al. Using student-generated questions for student-centered assessment. Assess Eval Higher Educ 2012;37(4):439-52.
14. Ranney RR, Wilson MB, Bennett RB. Evaluation of applicants to predoctoral dental education programs: review of the literature. J Dent Educ 2005;69(10):1095-106.
15. Finelli C, Bergom I, Mesa V. Student teams in the engineering classroom and beyond: setting up students for success. CRLT Occasional Paper no. 29. Ann Arbor, MI: Center for Research on Learning and Teaching, 2011.
16. Ohland MW, Loughry ML, Woehr DJ, et al. The comprehensive assessment of team member effectiveness: development of a behaviorally anchored rating scale for self and peer evaluation. Acad Manag Learn Educ 2012;11(4):609-30.
17. Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. Psychol Sci 2006;17(3):249-55.

18. Nguyen K, McDaniel MA. Using quizzing to assist student learning in the classroom: the good, the bad, the ugly. Teach Psychol 2015;42(1):87-92.
19. Behar-Horenstein LS, Schneider-Mitchell G, Graff R. Promoting the teaching of critical thinking skills through faculty development. J Dent Educ 2009;73(6):665-75.
20. Chi MTH. Two approaches to the study of experts' characteristics. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. The Cambridge handbook of expertise and expert performance. Cambridge, UK: Cambridge University Press, 2006:21-30.
21. Johnsen DC, Finkelstein MW, Marshall TA, Chalkley YM. A model for critical thinking measurement of dental student performance. J Dent Educ 2009;73(2):177-83.
22. Abrams RG. Questioning in preclinical and clinical instruction. J Dent Educ 1983;47(9):599-603.
23. Behar-Hornstein LS, Dolan TA, Courts FJ, Mitchell GS. Cultivating critical thinking in the clinical learning environment. J Dent Educ 2000;64(8):610-5.
24. Braunette DM. Critical thinking: understanding and evaluating dental research. Carol Stream, IL: Quintessence, 1996.
25. Chambers DW. Lessons from students in a critical thinking course: a case for the third pedagogy. J Dent Educ 2009;73(1):65-82.
26. Greenfield LB. Teaching thinking through problem-solving. In: Stice JE, ed. New directions for teaching and learning: developing critical thinking and problem-solving abilities. San Francisco: Jossey-Bass, 1987:5-22.