

Practical Approaches for Enhancing Student Performance in Summative Assessment

Se-Lim Oh^{1*}, Yoon Jeong Kim², Eun-Hwi Cho³ and Ji-Young Han⁴

¹Clinical Associate Professor, Department of Advanced Oral Sciences and Therapeutics, University of Maryland, School of Dentistry

²Associate Professor, Department of Periodontics, School of Dentistry, Loma Linda University

³Assistant Dean, Educational Quality and Outcomes, School of Dentistry, Loma Linda University

⁴Associate Professor, Department of Periodontology, Divisions of Dentistry, College of Medicine, Hanyang University

*Corresponding author: Se-Lim Oh, Email: soh@umaryland.edu

Received: 09 April 2019; Accepted: 28 April 2020; Published: 05 May 2020

Abstract

Introduction: An average of 15 out of 131 students have achieved below 70% on midterm examinations in the second-year periodontics course during the past 2 years in 2015 and 2016. The purpose of this study was to improve student performance in the midterm examination by implementing a formative assessment and improving the quality of multiple-choice questions (MCQs).

Methods: In 2017, second-year dental students were asked to submit student-generated test questions as a formative assessment before the midterm examination. Two periodontists from other institutions reviewed the MCQs in the midterm examination and provided comments on each question independently before the 2017 midterm examination. The students took the midterm and final examination. The quality of the 2017 midterm examination was evaluated using item analyses of the MCQs. Student scores were compared using a chi-square and a paired t-test.

Results: The 2017 midterm examination exhibited the highest item index, indicating the easiest examination among three midterm examinations ($F = 2.8, p = 0.042$) from 2015 to 2017. One student failed the midterm examination in 2017, showing the least number of students scoring below 70% when compared to ones in 2015 and 2016. The mean midterm score (88.4 ± 8) was significantly higher than the mean final examination score (84 ± 7) in 2017.

Conclusions: The study demonstrated that the implementation of formative assessment and peer review of examination questions improved student performance in summative assessment.

Keyword: Formative assessment; Peer review; Student learning; Summative assessment

Introduction

Assessment is critical to measure student learning. Two types of assessments have been typically applied in education; formative and summative assessment [1]. Formative assessment is a part of the instructional process. Its main goal is to gather feedback that can be used by instructors and students to guide improvements in the ongoing teaching and learning context [1]. Studies have reported that formative assessments may help students improve their performances in summative assessments [2,3]. Summative assessment measures the level of student proficiency attained at the end of an instructional unit or term [4].

Student course grades are generally based on their performances in summative assessments [5]. For the past 2 years, most dental students

have achieved a passing score ($\geq 70\%$) on the midterm examination in the second-year periodontics course, but an average of 15 out of 131 students have achieved lower than 70% (a failing score). This led to queries from the student-initiated examination critique committee (ECC) after the examination. Students criticized an average of 30 questions from a total of 40 questions in the midterm examination and complained about the quality of the multiple-choice questions (MCQs) in the examination.

The multiple-choice question (MCQ) is commonly used to assess student knowledge in health care field. However, studies have shown that faculty generated MCQs exhibited relatively low quality, containing nearly 46% of flawed or poorly written questions [6,7]. Researchers also concur that poorly written questions might affect student performance in the examination [6,8]. To improve the quality of examination questions, committee review, peer review, and faculty training have been suggested [7,9,10]. While the faculty who wrote the questions had reviewed the critiques from the student-initiated ECC, in 2017, peer review from independent periodontists on the MCQs was obtained before the midterm examination.

All lecture materials and actual lectures were posted on an online learning site at the University of Maryland School of Dentistry (UMSOD). Nonetheless, most critiques from the ECC were mainly stating that lecturers did not mention the specific sentences in the examination during lectures. Most arguments made by students were due to incorrect memorization of the examination questions and the lecture materials. This may suggest that students could not process test materials thoroughly. To promote the self-assessment of learning, a formative assessment was implemented in 2017.

Formative assessment had not been used in 2015 and 2016. Only a few quizzes with a true-or-false type question were delivered at the end of some lectures to verify student attendance. The feedback from the lecturers after those quizzes was not provided. Therefore, student-generated test questions in a MCQ format were implemented as a formative assessment in 2017 before the midterm examination.

Even though it is generally accepted that formative assessment has effects on student learning and the quality of MCQs also affects student performance in summative assessment, there is a sparse evidence regarding the effect of formative assessment and the quality of examination questions on student performance in the examination in predoctoral periodontal education. The purpose of this study was to improve student performance in the midterm examination by implementing a formative assessment and independent peer review of MCQs. The study compared the quality of 2017 midterm MCQs to those of the examinations from previous years using item analyses for the MCQ. The study investigated the effectiveness of formative assessment on student learning by comparing student assessment scores among different classes and within the class.

Materials and Methods

Outline of the study

This interventional education action research was conducted under an approved Not Human Subjects Research protocol by the Institutional Review Board at the University of Maryland, Baltimore (HP-00082574). The second-year periodontics course is a year-long course starting August each year. During a fall semester in 2017, the second-year dental students were asked to submit student-generated test questions in the MCQ format as a formative assessment before they took the midterm examination. Peer review on the midterm examination questions was performed by two periodontists from other institutions. The students took the midterm and final examinations.

Implementation of formative assessment

Second-year dental students were asked to write one question that they might expect to see on the midterm examination. Students received feedback regarding their questions and answers from the course director (SO). Table 1 shows examples of student-generated questions and answers and feedback from SO. With this process, students had opportunity to discuss the contents of the course material with the course director individually. Although students did not have an opportunity to discuss their questions with their classmates during the class, the discussion between individual student and the course director took place via the online learning site. Students could download and share the contents from the learning site. Therefore, student-generated test questions were not used in the midterm examination and not subject to the peer review, either.

Peer review of examination questions

The midterm MCQs and short answer questions (SAQs) were written by in-school faculty members at the UMSOD. Two periodontists (YK and JH) from other institutions were asked to review the MCQs. They provided critiques and/or comments on each question independently. Six questions from the midterm examination were revised based reviewers' comments. The examples are as follows: "Terms are confusing. There are controversial results depending on studies. Be more specific in diagrams".

Evaluating the quality of the midterm examination

This study used two item analyses (item difficulty index and item total correlation value) to compare the quality of MCQs in the second-year periodontics from 2015 to 2017. (Table 2) describes the two item analyses used in this study.

Student performances in summative assessments

The total number of questions in the midterm examination from 2015 to 2017 were similar, ranged from 40 to 43. The point values allocated to each MCQ and short-answer question (SAQ) were the same because SAQs were mainly required to provide simple factorial answers.

The midterm (34 MCQs and 6 short answer questions) and the final

(57 MCQs and 6 short answer questions) examination were delivered to the second-year dental students via the assessment technology tool in 2017. The number of students who achieved < 70% in the examinations were compared from 2015 to 2017. The scores from the 2017 midterm and final examination were compared.

Statistical Analysis

Descriptive statistics (mean ± SD) were prepared for all data. A one-way analysis of variance (ANOVA) was used to compare the quality of midterm examinations from 2015 to 2017. The *number of students who achieved < 70% in the midterm examination from 2015 to 2017 was compared using a chi-square test*. Student scores from the midterm and the final examination were compared using a paired t-test. Data analysis was performed with the GraphPad Prism (version 7; GraphPad Software, Inc. CA). A p-value < 0.05 was considered significant.

Results

The quality of the 2017 midterm examination was evaluated using item analyses of the MCQs in the examinations (Table 3). The 2017 midterm examination exhibited the highest mean item difficulty index (0.88 ± 0.13) among the three midterm examinations (F = 2.8, p = 0.042). The item difficulty index of 2017 midterm examination was significantly higher than 2016 midterm examination (mean item index = 0.77 ± 0.16). However, there was no difference in item total correlation values among the three midterm examinations.

The effectiveness of formative assessment on student performance in the summative assessment was evaluated comparing the number of students who achieved < 70% in the midterm examination from 2015 to 2017. The result of the midterm examination in 2017 showed the least number of students scoring below 70% when compared to ones in 2015 and 2016 (Table 4); Chi-square, p < 0.0001).

The mean final scores were significantly higher than the mean midterm scores in 2015 and 2016 when compared within the classes using a paired t-test (Figure 1a). However, the mean midterm score (88.4 ± 7.8) was significantly higher than the mean final score (84 ± 7.1) in 2017 (Figure 1b).

Table 1: Examples of student-generated questions and feedback.

Question generated by a student	Feedback
In periodontal patients, which type of radiograph is the most diagnostically useful? a. periapical b. horizontal bitewing c. vertical bitewing d. panoramic <i>Answer is c.</i>	Both (periapical and vertical bitewing radiographs) should be obtained. While bitewings show crestal bone loss, there is no information regarding the remaining bone support from bitewings.
Which provides information about a single patient: Positive predictive value (PPV) or Sensitivity? <i>Answer is PPV.</i>	The question is very vague. What do you mean "information about a single patient"? The concept of PPV or Sensitivity are applied to a diagnostic test. When you screen a large population, select a diagnostic test with high sensitivity. If you want to diagnose a patient in your private clinic, choose a diagnostic test with high PPV.

Table 2: Item analyses for multiple-choice questions.

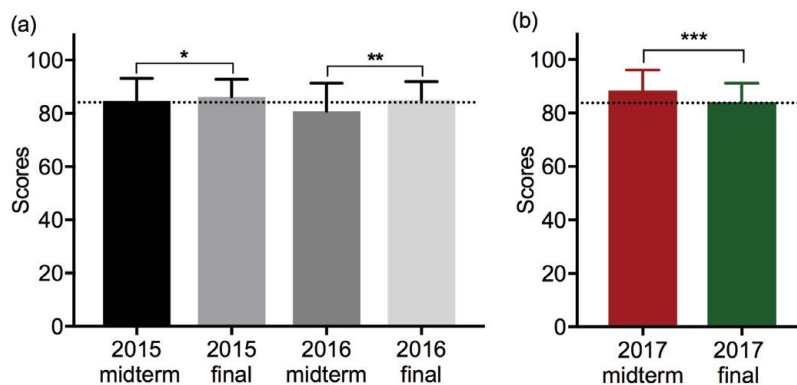
Item difficulty index	Range from 0 to 1
the percentages of students who answer a question correctly.	0 extremely difficult
	1 extremely easy
Item total correlation value	Range from -1 to 1
a correlation relating what students achieve on an item to the assessment scores that they achieve.	1: a student who answered the question correctly also received higher score on the test than those students who answered the item incorrectly
	-1: a student who answered the question correctly received lower score on the test than those students who answered the question incorrectly

Table 3: The quality of multiple-choice questions in midterm examinations (n = the total number of MCQs; ITC = Item Total Correlation).

	2015 (n = 40)	2016 (n = 39)	2017 (n = 34)	one-way ANOVA
Item Index (mean ± SD)	0.81 ± 0.1	0.77 ± 0.2	0.89 ± 0.13	F = 4.33, p = 0.015
ITC (mean ± SD)	0.22 ± 0.1	0.26 ± 0.1	0.25 ± 0.15	F = 0.88, p = 0.42

Table 4: The number of students who achieved < 70% in midterm examinations.

The number of students	2015 (n=130)	2016 (n=132)	2017 (n=130)	Chi-square test
<70%	10 (8%)	19 (14%)	1 (0.8%)	$\chi^2 = 17.2$ p < 0.0002
≥ 70%	120	113	129	

**Figure 1:** Score comparison between the midterm and the final examination. (a) Score comparisons for the 2015 and the 2016. *Paired t-test, p = 0.0124, **Paired t-test, p < 0.0001 (b) Score comparison for the 2017. ***Paired t-test, p < 0.0001.

Discussion

The MCQs are used to assess factual recall, comprehension, and application [11]. While the MCQ format has advantages of being versatile and convenient for grading, the questions need to be developed carefully to ensure validity. The peer review on examination questions might be useful for proofreading and feedback on questions prior to using them in actual examinations. In this study, peer review of MCQ questions focused on correcting confusing statements in the questions, confirming accuracy of the content, and consistency of the formatting. Peer reviews from independent periodontists might reduce student confusion, enabling them to answer correctly.

Item analyses of MCQs use student responses to questions in order to measure the quality of the MCQs. When there is item discrimination power (≥ 0.2 item total correlation value), the test is considered to differentiate low-performing students from high-performing students [12]. Easy questions, e.g. questions with item difficulty index 1, do not possess a discrimination power. As the item difficulty index decreases, the item discrimination power increases. However, there is an optimum level of item difficulty index because too difficult questions, e.g. questions with item difficulty index 0.1, do not distinguish between the high scorers and the low scorers effectively, either. For the past 2 years, the item difficulty indexes of the second-year periodontics course midterm examinations were 0.77 and 0.81, which are considered the ideal difficulty [12,13]. Although the 2017 midterm examination was the easiest in regard to the item difficulty index (0.88), it still exhibited a discrimination power (mean item total correlation value = 0.25).

In higher education, learning is conceptualized as a process where students actively construct their own knowledge and skills, rather than a simple acquisition based on teacher transmission [14]. Students interact with subject contents, e.g. by having discussions with others, in order to internalize meaning [14]. Formative assessment might help the internalizing processes. In this study, all students who submitted test questions received feedback from the course director (SO). Even though student-generated questions were not directly used in the

2017 midterm examination, students were recognizing their questions and could have shared feedback from the course director before the examination. This might lead them to review test materials differently.

No educational interventions were attempted before the final examinations. The mean scores among three final examinations were not different (Fig 1). Interestingly, the mean final scores were significantly higher than the mean midterm scores in 2015 and 2016 (Fig 1a), although new materials were taught between the two examinations and the final examinations contained more questions to cover all materials taught in the entire course. After implementation of the formative assessment and peer review on the 2017 midterm examination questions, the 2017 mean midterm score was significantly higher than the 2017 mean final score (Fig 1b). The second-year dental students in 2017 exhibited the lowest failure rate in the midterm examination compared to the two previous years (Table 4).

Generalization of the study results might be limited by a few factors. The year-to-year comparison may suffer from unnoticeable confounding factors, as there may have been important differences among the students in each year in addition to the intended educational intervention. The written examinations as summative assessment have been slightly different in each year. However, the course director, the lecturers, and the contents in the second-year periodontics course did not change during the study period.

Conclusion

The study demonstrated that the implementation of formative assessment and peer review of examination questions enhanced student performance in summative assessment. Peer reviews from independent periodontists might reduce student confusion, enabling them to answer correctly. Formative assessment accompanied by feedback from the instructor could be a good way to make students engage in learning.

Acknowledgements

The authors report no conflicts of interest related to this study. There is no funding related to this study.

References

1. Taras M. Assessment—summative and formative—some theoretical reflections. *Br J Educ Stud.* 2005; 53:466-478.
2. Dunlosky J, Thiede KW. What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychol (Amst).* 1998; 98:37-56.
3. Vanderlelie JJ, Alexander HG. Learning-oriented assessment increases performance and written skills in a second year metabolic biochemistry course. *Biochem Mol Biol Educ.* 2016; 44:412-420.
4. Harlen W, James M. Assessment and Learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice.* 1997; 4:365-369.
5. Harlen W. Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal.* 2005; 16:207-223.
6. Downing SM. Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? *Acad Med.* 2002; 77:S103-4.
7. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. The quality of in-house medical school examinations. *Acad. Med.* 2002; 77:156-161.
8. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Med Educ.* 2008; 42:198-206.
9. Dellinges MA, Curtis DA. Will a Short Training Session Improve Multiple-Choice Item-Writing Quality by Dental School Faculty? A Pilot Study. *J Dent Educ.* 2017; 81:948-955.
10. Malau-Aduli B, Zimitat C. Peer review improves the quality of MCQ examinations. *Assessment & Evaluation in Higher Education.* 2012; 37:919-931.
11. Albino JE, Young SK, Neumann LM, Kramer GA, Andrieu SC, Henson L, 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; 72:1405-1435.
12. Rao C, Kishan Prasad HL, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *Int J Educ Psychol Res.* 2016; 2:201-204.
13. G. Pope, "Item Analysis Analytics," Questionmark White Paper. 2003; 1–19.
14. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud High Educ.* 2006; 31:199-218.