Confidential Peer-Evaluation as a Method of Learning in Online University Courses

Simon Lemaire, PhD[1], Gladys Brunyninx, MD[2], Miriam Grenon, CL[3], Madisson Kelleher-Radey, BScN[4] , Alexander Yeuchyk, PhD[5]

[1]Department of Cellular & Molecular Medicine, University of Ottawa, 451 Smyth Road, Ottawa, ON, K1H 8M5, slemaire@uottawa.ca (corresponding author).

[2]Faculty of Medicine, University of Ottawa, 451 Smyth Road, Ottawa, ON, K1H 8M5, gbruy042@uottawa.ca

[3]Faculty of Law, University of Ottawa, 57 Louis-Pasteur Private, Ottawa, ON, K1N 6N5, mgren012@uottawa.ca

[4]Faculty of Health Sciences, University of Ottawa, 451 Smyth Road, Ottawa, ON, K1H 8M5, mkell123@uottawa.ca

[5]Teaching and Learning Support Service, University of Ottawa, 136 Jean-Jacques Lussier Road, Ottawa, ON, K1N 6N5, ayeuchyk@uottawa.ca

Abstract

Online teaching of university courses to a large number of students is a major challenge in both its execution and evaluation. The proposed model for solving the logistical issues of these courses is based on using the Wiki tool in the Blackboard learning management system (LMS) to facilitate the submission and evaluation of student essays. The Test tool also permits the collection of confidential peer evaluation based on criteria established by the professor. 220 students studying Human Anatomy and Physiology (A&P) were divided into random groups of 5 by Blackboard to undertake and submit a dissertation on one of 5 different topics and, secondly, to evaluate the work of their teammates. The assessments were collected, calculated, and validated by the professor and teaching assistant in Microsoft Excel® and submitted to the Blackboard Grade Center. A survey among students after the course showed a high degree of satisfaction with the criteria related to the completion of the dissertation and peer review assignment, as a method of learning. https://doi.org/10.21692/haps.2022.011

Key words: online teaching, written essays, peer review, learning, anatomy and physiology

Introduction

One of the biggest challenges in teaching a large number of students online or in university classrooms is finding a method to monitor and appropriately evaluate student performance and degree of knowledge and understanding on the subject matter while ensuring that the method does not permit plagiarism or biased results (Distler 2015). The most popular summative assessment method is based on exams with multiple choice questions (MCQs). MCQs are also used by many professors to provide formative self-testing for students in between the summative exams (Lull et al. 2016). This method has advantages in that it makes it possible to quickly collect the answers and determine the marks of large numbers of students using computer grading. MCQ-based exams can be easily proctored in the classroom; however, they are more challenging to proctor when exams are administered online. An important consideration with MCQ-based exams is the fact that the correct answers exist in the lists of answer choices provided in the questions themselves; the students simply have to identify them. This means that, in order to answer MCQs correctly, students can either study hard and thus know and understand the material of the course (McConnell et al. 2015) or, at times, proceed by deduction or guesswork to arrive at the correct answer for some questions.

Given the difficulties in assessing the performance of the growing number of university students and the limited number of faculty (Burgess et al. 2014) as well as the emergence of online teaching (Distler 2015), there is a need to develop a method that would allow for assessment of student competency and degree of knowledge while encouraging their engagement, motivation, and accurate understanding of course content. Writing assignments, coupled with peer review of these assignments, appear to be a means of providing quality learning opportunities if properly structured and monitored (Pond et al. 1995).

Peer evaluation has previously been studied in undergraduate courses in pharmacy (Dochy et al. 2006; Storjohann et al. 2019) and biology/chemistry (Finkenstaedt-Quinn et al. 2017; Shultz and Gere 2015). These studies assessed the implementation of peer- and self-grading systems and showcased positive results. The methods relied on the development of a great degree of honesty, content knowledge, and self-confidence by the students and they appeared to provide a valuable learning experience for students when coupled with effective instructor guidance.

continued on next page

In this study, students in a large human anatomy and physiology (A&P) class were divided into random groups of five and asked to complete a two-part assignment in which they each wrote an essay on one of five assigned topics and then conducted confidential peer evaluations of the four other written assignments completed by their group of peers.

Methods

General procedure

Within the class of 220 students, random groups of 5 students were formed using the Blackboard LMS. A content area in Blackboard was also created to define the "Rules of the Dissertation" (Table 1) which detailed the subject topic options, the content, the procedure, and deadlines for the dissertation. At the beginning of the course, an oral presentation of 15-20 min was given to explain the "Rules of the Dissertation" and emphasize the various tasks to be completed.

The steps were as follows:

1. Each student in the group first chose one of the five topics related to the different body systems studied in A&P: pernicious anemia (digestive system), hypothyroidism (metabolism), acquired immunodeficiency syndrome (immune system), renal failure (acute or chronic; renal system), or infertility (male or female; reproductive system). Students then published both their subject choice and basic references to be used for their dissertation in the Wiki group. The references published in Wiki allowed the professor or teacher's assistant (TA) to have rapid access to the information. Subject choices were first come first served as students in the same group could not choose the same topic.
2. Approval of student selected topic and references was done by the professor or TA by making a brief comment on the student's Wiki page to endorse the literature or make suggestions.
3. Following subject approval, students composed their essay by adhering to the directives outlined in the content area entitled "Rules of the Dissertation." Briefly, the dissertations were to contain the following elements: a brief summary of the subject (maximum of 200 words), a description of signs and symptoms and possible cause(s) of the disease, a description of the organs and/or systems and/or mechanism involved, a description of treatment and prevention methods of the disease, and references (maximum of 10-12). This work with a limited number of single-spaced pages (2-3 pages) could be supplemented with figures and tables with appropriate legends. Student submitted their dissertation on the Wiki page within their group.
4. Each student read and evaluated the dissertations of their teammates according to the following criteria: a) clarity; b) coverage of the subject and c) overall quality and relevance of the work. They submitted their evaluations in percentages, using a confidential assessment grid (Table 2) through the Test tool in Blackboard. The Test tool allowed the professor (or TA) to export students' grades and comments to Excel, evaluate the variability and reliability of the grades, and relay the feedback to the students.

Dissertation target dates

The dissertation assignment had three deadlines namely for (1) the choice of the subject, (2) the submission of the dissertation, and, (3) the submission of the peer evaluations (Table 1). Evaluation criteria were established and listed in the "Rules of the Dissertation" as well as in an "Evaluation Grid" used to collect the peer evaluation data (Table 2). Throughout the semester, the professor (or TA) used Blackboard to verify that students were completing the tasks on time and to identify those students who needed to complete one or more given tasks. In the latter case, late students were notified.

The marks and comments entered in the "Evaluation Grid" by the students were solely viewed by the professor or the TA in the Blackboard Grade Center from which the data were downloaded in the form of an Excel spreadsheet. Working within Excel, the professor (or TA) calculated the averages, the standard deviations, and variances of the marks to verify and detect any evaluations that fell outside the range. The final grade for each student was calculated in Excel by the professor (or TA) and uploaded in a new column in the Blackboard Grade Center to be viewed by students.

To conserve confidentiality, student comments were randomly number coded from 1 to 220 and displayed in a spreadsheet on the Blackboard announcement page. The codes of the students were uploaded in a new column of the Blackboard Grade Center so that each student could find their individual code number (and feedback from peers on their essay). The professor (and/or TA) also graded some students who either received a nonconforming evaluation (high margin of error between the assessments) or were assigned to a smaller group due to course withdrawal by some students. Students who did not submit their work, or submitted it too late and without justification, or who incorporated plagiarism or provided biased assessments were informed of the possibility of receiving a grade of zero (Table 1).

continued on next page

THEMES	DESCRIPTION
A. TOPICS	1. Pernicious anemia (digestive system) 2. Hypothyroidism (metabolism) 3. Acquired immunodeficiency syndrome (immune system) 4. Renal failure (renal system) 5. Male or female infertility (reproductive system)
B. CONTENT	Brief summary; introduction; problem presentation; discussion; references; tables and figures.
C. EVENTS	1. Formation of random groups of students by Blackboard. 2. Student choice & registration of a dissertation subject plus select references in a Wiki. 3. Approval of the subject and documentation by the professor or TA. 4. Writing and submission of the dissertation by the student within the Wiki.
D. EVALUATION	Values (%) based on criteria (1. Clarity; 2. Coverage of the subject and 3. Overall quality and/or relevance) are set for the essay. Use of the confidential "Evaluation Grid" for the student submission of teammate evaluation grades and comments (strengths and weaknesses).
E. TIME LIMITS	Dates (deadlines) are established for: 1. The choice of the subject. 2. The submission of the dissertation. 3. The submission of the evaluations.
F. PENALITIES	Unjustified late submissions, plagiarism and biased evaluations are prone to be downgraded to zero.

Table 1. *Rules of the dissertation.*

Examples

To facilitate the dissertation process, a content tab was created to show examples of search engines (PubMed, CINAHL, Medline, PasseportSanté, etc.) that the students could use to find appropriate documentation on their subject. An example of a dissertation on a different topic (Crohn disease (digestive system)) was provided to illustrate the format and quality level that was expected from each student. Finally, an evaluation of the provided example was also shared to show how to analyze a dissertation according to the specific criteria established for the essay and how to formulate a sound review to highlight its strengths and weaknesses.

D2L Brightspace learning platform

The method described above with the Blackboard LMS can also be applied to other learning platforms including, for example, D2L Brightspace. The "Group" tool in D2L can be used to perform events 1, 2, 3 and 4 in Table 1, while the collection of confidential student evaluations (event 5) can be accomplished using the D2L "Quiz" tool with the evaluation grid illustrated in Table 2.

continued on next page

Name of student	Group number	Clarity (%)	Coverage of topic (%)	Overall quality & significance (%)	Comments to justify evaluations
#1					
#2					
#3					
#4					

Table 2. *Rubric used by students to evaluate the dissertations of their 4 teammates.*

Results

Validation of student evaluations

The work was completed in accordance with the deadlines by all students. The dates of the deadlines were spread over the entire course session to allow the students time to undertake and enjoy each step, all of which are integral parts of the learning process. The essay counted for 10% of the final grade. The student evaluations were collected and averaged in Excel. The variances (VARP) of the marks (4 per student) were calculated and those that exceeded 7% of the average mark were judged as unreliable and the dissertations were re-evaluated by the professor and TA. Out of 220 evaluations, only 9 were judged to be suspicious. However, after re-evaluation by the professor (and TA), only two of them were slightly modified, indicating that the student evaluations were essentially valid.

Evaluations of the dissertations by the students tended to be higher than those verified by the TA (average of 89% as compared with 84% on a sample of 4 groups of students) or obtained with the final MCQ exam (average of 86% as compared with 76% for the whole class of 220 students). One way to avoid overvaluation (or undervaluation) was to warn the students in advance that they could be penalized if they gave their classmates marks that were too high or too low without proper justification in the comments section of the evaluation rubric. On the other hand, the high grading of students in these essays proved to be valid based on the stronger work of the students in the essays compared with the general closed-book examinations, in part due to having free access to the documentation while generating their dissertations.

Student satisfaction and possible use of D2L Brightspace instead of Blackboard LMS

A survey of the students (Table 3) after this exercise demonstrated a high degree of satisfaction for this method of learning and self-evaluation (Figure 1). Students reported satisfaction levels of 86.9%, 84.4%, 56.4% and 63.5% with

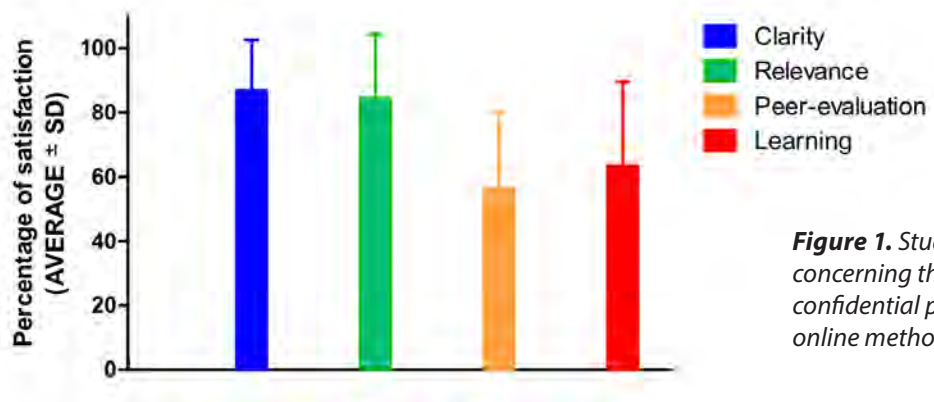

Figure 1. *Student satisfaction rate concerning the clarity, relevance, confidential peer evaluation and online method of learning.*

continued on next page

Give a rating of 0 to 5 (0= strongly disagree, 5= strongly agree)	***0***	***1***	***2***	***3***	***4***	***5***
The instructions given in class and on the website about the dissertation were clear.						
The topic choices were relevant and appropriate to the course objectives						
I consider my essay evaluation by my colleagues to be fair and valid						
Overall, I consider the confidential peer-evaluation of dissertations to be a good learning tool.						
Comments:						

Table 3. *Rubric used by the students to rate the dissertation process for its clarity, relevance, validity of peer evaluation and online method of learning.*

regard to the criteria of: clarity of instructions, relevancy of topics, validity of peer evaluation marks, and value of the online method of learning through peer-reviewed dissertations, respectively. Most of the student comments were positive, and favoured the continuation of this method of learning, especially for A&P courses which demand significant memorization in preparation for summative examinations. This method of teaching and learning a specific A&P subject was novel and refreshing for us compared to the traditional methods of teaching and learning. Some comments indicated the importance of properly instructing the students about the necessity to read and understand their classmate's essays prior to evaluating them according to the established criteria. Following the switch from Blackboard to the D2L Brightspace LMS at the University of Ottawa, the method was applied to other groups of students with the same ease, and peer reviews were collected in the same anonymous manner as with the Blackboard LMS.

Discussion

The main advantages of peer-evaluations are the incentives for students to both produce high-quality works and provide accurate assessments of their classmates' essays which ultimately enhance the quality of the learning process itself (Adachi et al. 2018; Chen 2012). Students who know they will be evaluated by their colleagues are generally more competitive and want to submit quality essays. After taking the time to write and submit their own essay, they are eager to learn what their colleagues have submitted. They can then evaluate the work of their colleagues according to the criteria they themselves have previously followed, and with which they are familiar. We found that students were not prone to misjudging their colleague's work because they knew they could be penalized for giving an unjustified mark that was found to be either too high or too low. Also, in their assessment, they were asked to add short comments to indicate the strengths and weaknesses of the work, thereby justifying their assessments.

In terms of learning, the students achieved this by gathering information on a given subject, integrating that material, and describing the important features, while following the well-defined guidelines for essay format. Furthermore, assessment of the work of four of their colleagues on different topics added to their learning about other aspects of the course. As reported by others, we found that students tended to give their colleague's works higher marks in comparison with the professor's (or TA's) evaluation (Li et al. 2016; Reinholz 2016; Rudy et al. 2001). On the other hand, the higher grades observed with the peer-review essays in comparison with those obtained in the final MCQ exam may have been due, in part, to the fact that the final exam covered the entire course and was timed and closed book, whereas the dissertation essays covered specific aspects of the course and students were given much more time to research and write their articles. Nonetheless, due to a tendency of the essays to have higher scores than the MCQ exams and the fact that they were targeting only certain aspects of course content, we suggest that such works should be worth between 5% and 15% of the final grade.

An ability to properly evaluate the work of peers is part of the learning process required in several schools of medicine and implies that the students must learn to recognize their strengths and weaknesses in writing, in topic knowledge, and in time management (Chen 2012). The group of students referred to herein were primarily Faculty of Health Sciences students registered in an A&P course given by the Faculty of Medicine of the University of Ottawa. However, this teaching and learning strategy could be applied to any other university group, regardless of their discipline. As part of their academic

continued on next page

courses, or future career, all students will eventually be evaluated by their peers, and they may also have to assess the performance of their colleagues. We therefore consider that the peer review system described in this study is a form of learning that is not only easy to implement, convenient, and valid with large groups of students, but it is also necessary for the preparation of the students for their future careers.

According to Papinczak et al. (2007), the peer review process may have a negative impact on problem-based learning (PBL) groups among medical students. These authors mentioned that collaboration and collegiality within a PBL group can be decreased in response to negative feedback from colleagues. To prevent this possibility, we formed random groups of students by making use of a Blackboard tool that allowed the formation of groups wherein most students did not know one another. Furthermore, the evaluations and comments submitted using the Blackboard Test tool were confidential and the professor collected them and established averages and variances prior to submitting the anonymous marks and comments to the Blackboard grade center. This ability to process marks and comments confidentially allowed the students to get to know their strengths and weaknesses without developing any resentment towards their teammates.

Students in the Faculty of Health Sciences as well as those in Medicine experience pressure to succeed. Due to previous experiences in education, some students may feel uncomfortable or skeptical about the value and validity of peer evaluations (Balantyne et al. 2002). Typically, their academic abilities have been developed to learn a large array of content-specific information, but they generally do not know yet how to proceed to evaluate essays. In this context, the professor's role is particularly important to determine the evaluation criteria and reassure the students of their capabilities to pick up essential knowledge from other people's work and to form their own judgments about the work's merits. Students can develop these abilities quite rapidly after being trained by first creating their own work and then applying the same criteria when evaluating colleagues' works (Newbold et al. 1995). The assignment described above required the professor and TA to allocate time to verify the different stages of the learning process, to calculate the averages and variances of the collected marks, and, if necessary, to evaluate the essays themselves if scores looked suspicious. Such verification allowed us to detect one case of plagiarism in a group of 60 students when this technique was reapplied in another class (unpublished data). Therefore, it is very important that the professor and/or TA remain vigilant and proactive in order to ensure the success of this peer-review method of evaluation and self-learning.

We did not explore all of the possibilities that this peer-review learning procedure may have for developing reflective and critical thinking as described by Harasym and colleagues (2008). For example, the 3 or 4 best essays on each research topic within the class of 220 students could have been selected with the help of the Blackboard LMS after the compilation of student evaluations and included as reference content testable in the final exam. This way, the students would get to know what can be considered as the required knowledge on each topic. On the other hand, another way to strengthen the student integration and comprehension of the research topics could also be to form student subgroups among the existing groups wherein each student would be put in contact with a determined number of students of other groups having chosen the same subject. These student subgroups could be requested to hold online meetings to discuss the research subject before writing their own essay. This way, their work would result not only from their own research and way of thinking, but also from the discussion with other students preparing their dissertation on the same topic. A brief report of these online meetings could be put on the subgroup Wiki so that the professor or TA could verify the dates and contents of the meetings and monitor or advise the students whenever necessary. Finally, another way to strengthen this peer-review method of learning could also be to ask the students to submit a revised version of their work after receiving and considering the comments of their teammates for improvement of the essays.

Previous studies have reported that students' unfamiliarity with concepts or peer review activities of essays can be an obstacle to the learning process (Halim et al. 2018). In this regard, we spent much time establishing and explaining the rules of the dissertation and the time limits of the various tasks to be carried out by the students (Tables 1 and 2). The degree of clarity of the instructions as indicated in the student survey was relatively high. However, some students still indicated in their comments that the rules of the essay became clear to them only after the TA and/or professor provided a PowerPoint presentation to the class outlining the dissertation procedures, suggesting the importance of person-to-person communication for the disclosure of instructions to students.

The rationale for the project described above of groups of 5 students writing an essay on a disease related to a system in A&P followed by peer evaluation by colleagues who have written an essay on a different topic linked to a different body system stands on the need to develop education tools that promote the ability of students to be self-directed learners, to be creative, and to become more fully immersed in their education. In this assignment, students had to conduct research and form their own opinions on issues that were related to systems studied in the A&P course. This type of assignment encourages students to make links between concepts and physiological and/or disease conditions. The research on these new education avenues is justified at this time as many of our courses may remain hybrid or even fully online as we gradually transition our teaching during the post-pandemic period.

continued on next page

About the Authors

Simon Lemaire, PhD, is a Professor in the Department of Cellular and Molecular Medicine at the University of Ottawa. He teaches A&P and pharmacology to graduate and undergraduate students in the Faculties of Medicine, Health Sciences and Science. His research focuses on developing online and hybrid courses for students. Gladys Brunyninx, BSc, is an undergraduate medical student at the University of Ottawa. She was a TA for the A&P course. Miriam Grenon, BSc (Hons), is a JD 2022 candidate at the University of Ottawa and was also a TA for the A&P course. Madisson Kelleher-Radey, BScN, was a TA for the A&P course and was particularly involved in collecting the anonymous survey data. Alexander Yeuchyk, PhD, was with the Teaching and Learning Support Service for the development of hybrid courses at the University of Ottawa. He is now a professor in the Languages Department, University of Ottawa.

Literature Cited

Adachi C, Tai JHM, Dawson P. 2018. Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assess Eval High Educ* 43(2):294-306. https://doi.org/10.1080/02602938.2017.1339775

Ballantyne R, Hughes K, Mylonas A. 2002. Developing procedures for implementing peer assessment in large classes using an action research process. *Assess Eval High Educ* 27(5):427-441. https://doi.org/10.1080/0260293022000009302

Burgess A, McGregor D, Mellis C. 2014. Medical students as peer tutors: a systematic review. *BMC Med Educ* 14:115. https://doi.org/10.1186/1472-6920-14-115

Chen JY. 2012. Why peer evaluation by students should be part of the medical school learning environment. *Med Teach* 34:603-606. https://doi.org/10.3109/0142159X.2012.689031

Distler JW. 2015. Online nurse practitioner education: Achieving student competencies. *Nurse Pract* 40(11): 44-49. https://doi.org/10.1097/01.NPR.0000472249.05833.49

Dochy F, Sergers M, Sluijsmans D. 1999. The use of self-, peer and co-assessment in higher education: A review. *Stud High Educ* 24(3):331-350. https://doi.org/10.1080/03075079912331379935

Finkenstaedt-Quinn SA, Halim AS, Chambers TG, Moon A, Goldman RS, Gere AR, Shultz GV. 2017. Investigation of the influence of a writing-to-learn assignment on student understanding of polymer properties. *J Chem Educ* 94(11):1610–1617. https://doi.org/10.1021/acs.jchemed.7b00363

Halim AS, Finkenstaedt-Quinn SA, Olsen LJ, Gere AR, Shultz GV. 2018. Identifying and remediating student misconceptions in introductory biology via writing-to-learn assignments and peer review. *CBE—Life Sci Educ* 17(2):ar28. https://doi.org/10.1187/cbe.17-10-0212

Harasym PH, Tsai TC, Hemmati P. 2008. Current trends in developing medical students' critical thinking abilities. *Kaohsiung J Med Sci.* 24(7):341-355. https://doi.org/10.1016/S1607-551X(08)70131-1

Li H, Xiong Y, Zang X, Kornhaber ML, Lyu Y, Chung KS, Suen HK. 2016. Peer assessment in the digital age: a meta-analysis comparing peer and teacher ratings. *Assess Eval High Educ* 41(2):245-264. https://doi.org/10.1080/02602938.2014.999746

Lull ME, Mathews JL. 2016. Online self-testing resources prepared by peer tutors as a formative assessment tool in pharmacology courses. *Am J Pharm Educ* 80(7):124. https://doi.org/10.5688/ajpe807124

McConnell MM, St-Onge C, Young ME. 2015. The benefits of testing for learning on later performance. *Adv Health Sci Educ Theory Pract* 20(2):305–320. https://doi.org/10.1007/s10459-014-9529-1

Newbold SK, Curran CR, Henson JL, Lee T, McDowell DE, Gassert CA. 1995. A research proposal to study the impact of an automated charting and assessment system. *Medinfo* 8(2):1088. https://pubmed.ncbi.nlm.nih.gov/8591375/

Papinczak T, Young L, Groves M. 2007. Peer assessment in problem-based learning: a qualitative study. *Adv Health Sci Educ Theory Pract* 12(2):169-186. https://doi.org/10.1007/s10459-005-5046-6

Pond K, Ul-Haq R, Wade W. 1995. Peer review: A precursor to peer assessment. *Innova Educ Train Int* 32(4):314–323. https://dx.doi.org/10.1080/1355800950320403

Reinholz D. 2016. The assessment cycle: A model for learning through peer assessment. *Assess Eval High Educ* 41(2):301-315. https://doi.org/10.1080/02602938.2015.1008982

Rudy DW, Fejfar MC, Griffith CH, Wilson JF. 2001. Self- and peer assessment in a first-year communication and interviewing course. *Eval Health Prof* 24(4):436-445. https://doi.org/10.1177/016327870102400405

Shultz GV, Gere AR. 2015. Writing-to-learn the nature of science in the context of the Lewis dot structure model. *J Chem Educ* 92(8):1325–1329. https://doi.org/10.1021/acs.jchemed.5b00064

Storjohann T, Pogge T, Peckham A, Raney E, Barletta JF. 2019. Evaluation of a peer- and self-grading process for clinical writing assignments. *Curr Pharm Teach Learn* 11(10):979-986. https://doi.org/10.1016/j.cptl.2019.06.003

Back to TOC