

Peer Grading Exams with Video Rubrics

Shawn Lupoli

Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County (UMBC)
Baltimore, MD, USA
slupoli@umbc.edu

Abstract -Previous research in peer grading and rubric systems has focused mostly on their use with homework assignments, not tests. Some reasons for this could be that tests are much longer than assignments, are of higher value in grading schemes, and could be highly contested. All of this would take valuable and scarce class time for a class that meets face to face. Managing time while maintaining control of the tests and protecting student privacy adds to that challenge. The author has been able to accomplish this with the use of a rubric, video editing software such as Camtasia, and YouTube. This model can increase grading accuracy, and greatly reduce the time it takes to grade exams while answering student inquiries and concerns, saving the instructor time.

I. INTRODUCTION

Starting with Topping's [10] work on benefits of peer grading on students, there have been many studies substantiating that hypothesis [1, 5, 6, 8]. But consistent issues arise such as the accuracy of the graders, since they themselves are students, and the number of graders it takes to achieve that consistency. Many articles on peer grading are based around online or MOOC classes [3, 9] or online tools such as PRAZE [15], so finding an established procedure for a successful face-to-face peer grading experience for all involved is described here. This in-class version of peer grading uses concise videos to answer, expand on, and grade the an exam question. The in-class setup also allows the student to "skip" grading a question if they are confused by it, allowing the instructor to make the final grading decision. The development cycle of exam, rubric, blind coupon, rubric video and feedback ensures a successful peer grading experience for both instructor and student if continually developed for content and clarity. This cycle for peer grading of exams could be used for any subject matter. There are several features of the exam peer grading cycle that would be familiar to anyone who has tried peer grading, but this model has advantages over others in terms of consistency, time investment, privacy and return on investment (ROI).

II. THE DEVELOPMENT CYCLE

A. Veteran Exam

In the Data Structures course, a vetted and veteran exam consists of such complex questions as problem solving and creating code using the various data structures covered in class. The type of questions (multiple choice to essay) on an exam is not a factor that would favor a positive experience with peer grading, over that of a rubric. After several semesters of teaching the same class and exam, the development of veteran exams with clearly defined instructions and solutions is crucial since grading for both

the student/peer grader and instructor will be easier and the chance of outliers kept to a minimum.

B. Solid Rubric

Having a strong rubric also serves to narrow the gap that can exist in terms of knowledge and skill between students and the instructor [4, 7]. The rubric takes most judgment away from students so they can focus on what and how the solutions are formed. This has another benefit of increasing the accuracy of grading for any random grader. If all else fails and a student is not comfortable with grading a particular answer, they can opt out.

Peer grading is not meant to completely relieve the instructor of grading, as there may be confusion on the peer grader's part, which may take time to resolve. In such cases, the instructor might still break even on their return on investment (ROI) in developing this peer grading cycle.

Peer grading will still work for questions that do not have particular solutions such as essay or open ended questions. A solid rubric does not have to be rigid. It can have several options. Another approach for grading problems that don't have a single solution is that the instructor may grade this portion of the exam. For example, in Computer Science, there are many ways to program a Binary Search Tree (BST) traversal. Giving options for solutions or providing a wide range with assigned points would still meet the goal of timely feedback with graded results.


C. Blind "coupon" cover page

Since this grading is done in class, protecting the privacy of the gradee/grader is essential but also more challenging since both can be in the same classroom. Previously conducted grading sessions have been completed anonymously, such that neither the grader nor the exam-taker knows who the other is. Both student and grader have taken the same exam, so both share the same topics, instructions, success and rubric. While there have been some interesting work in this area [12], the simplest way to protect the participant's identity is done using a perforated 8x11 "coupon" paper. This auto generated coupon sheet is in three sections: grader, instructor and student. While only the instructor and student coupons contain the name of the test taker, all three sections contain a randomized string sequence in order to match the test to a student when the instructor retrieves the peer graded test. When taking the test, the students are given a test with the blind "coupon" cover sheet. (Figure 1)

This coupon system, along with the random distribution of tests for grading, ensures the test-taker’s privacy and eliminates any bias that could happen if a grader knew they were grading a friend’s paper.

Topic:

Lupoli's CMSC 341



Section:

Test Ticket

Exam 1

Grader Initials:


Score:

Test Ticket -> ECDFFAAD

DO NOT REMOVE THIS PORTION

Topic:

Lupoli's CMSC 341



Section:

Test Ticket

Exam 1


Student Name: _____ (please fill name, and section above)

Test Ticket -> ECDFFAAD

Instructor Copy

Topic:

Lupoli's CMSC 341



Section:

Test Ticket

Exam 1

Student Name: _____ (please fill name, and section above)

Test Ticket -> ECDFFAAD

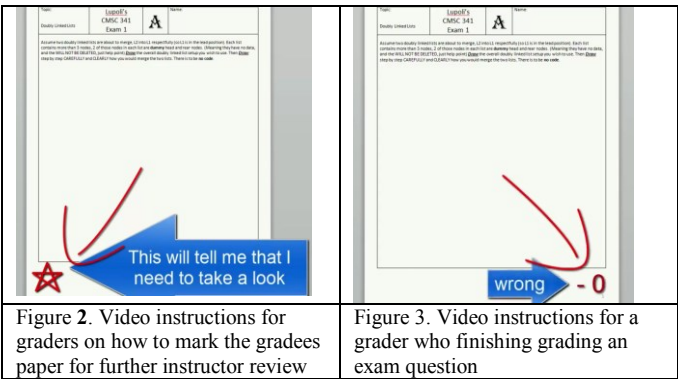
Figure 1. Blind "coupon" cover sheet

D. Video Rubric

A video rubric is a compressed and refined explanation of the solution to an exam question, followed by a rubric point breakdown. The editable video enables the instructor to use graphics, other videos, and whatever means to communicate the solution clearly with limited time in mind. The only time when all stakeholders are together is during class. While trying to save time, grade, and reveal answers to as many questions as possible are the primary reasons for peer grading, this is not the main goal of the video rubric. The student needs to understand what the correct answer is, and why, as the question could be a fundamental concept used throughout the course.

Grading the question for the random unknown student is secondary. In order to accomplish this, the most successful structure for the video is as follows: 1. Show the correct answer first. 2. Explain and show how the correct answer came to be. 3. Show partial credit for each portion building up to the correct answer. Showing the correct answer first gives the grader the possibility of four scenarios: A. Both the grader and gradee got the answer correct. B. Only the grader got the answer correct. C. Only the gradee got the answer correct. D. Neither got the question correct. While the fourth scenario is the most dire, the student can opt out of grading by placing a star at the lower left hand corner which indicates to the instructor that the question needs to be further evaluated. This would enable the grader to focus (Figure 2) on their own learning and be able to understand the correct answer about to be shown. In any of the other scenarios but D., the grader only has to focus on one portion of the video therefore understanding and grading to the correct answer. (Figure 3). Another learning opportunity

emerges for the grader during peer grading: the ability to see how others approached the answer. Whether good or bad, the grader gets to appreciate and analyze the test taker’s original response.



The creation of the videos can be done with any of the video editing software options available. Many come with built in tutorials, customer assistance and support groups for those with little experience. While the initial amount of work to create these videos can be steep, using the same test or version of test for years, including minor edits, minus much of the grading, quickly gives the instructor a solid ROI.

Finally, storage for these videos is as easy as uploading them to YouTube. With unlimited storage capacity and the ability to set the videos to private, the instructor will be able to view and show these videos at any given time during class or in their office with a student.

E. Student Perceptions of Peer Grading

While not as detailed as Mulder’s, et al. study on how peer grading influences perceptions [8], a survey was conducted just after grading the first exam to gather feedback from students about their perceptions and experiences with peer grading using the video rubrics. The online survey has four sections: student concerns, learning from the peer grading, overall peer grading experience, and matching expected grades with the results. Fifty-seven students out of 75 in two sections responded to this particular survey.

Students are seldom exposed to peer grading. In the survey conducted, only 38% of students had previous experience with peer grading. Many are concerned that their peers might make mistakes, but these fears are addressed, as previously mentioned, by having the exam viewed three times: by the grader, then the instructor, and back to the test taker. As the students are handed back their peer graded exams, student concerns are either minimal or quick to resolve. With the use of peer grading and video rubric, the back and forth disagreements are minimal since every student knows the overall breakdown of the exam. If a student was graded incorrectly, the student is informed and can quickly articulate the issue to the instructor for a proper grade change. The students were asked if the grader had

made a mistake on their exam. 87% responded that no mistake was made. Of the other 13%, only 1 respondent considered the mistake not easily fixable, the rest considered it a quick fix. To underscore the point that there is minimal difference on who grades the test, 73.7% believed they would have received the same grade, while almost 20% were unsure (Figure 4).

10. Do you believe you would have received the same or a similar grade from the instructor as you		
Answer Options	Response Percent	Response Count
Yes	73.7%	42
No	7.0%	4
Unsure	19.3%	11

Figure 4. Confidence in peer grading

Using video rubrics, of course, saves time grading. But the true test of a good rubric is to see if the students learned or confirmed their understanding of the particular exam question/content while viewing these videos in class. During grading, students will have questions to clarify either the answer and how to grade a student’s response. Some might require exam question, rubric, or video clarification. 75% of those surveyed found the video rubrics beneficial to their learning. And video rubrics were not the only way to learn from grading. Receiving peer graded comments is another way to learn. At least 59% of students surveyed thought the grader comments were somewhat helpful, while 25% did not receive comment from their grader at all. Grading the exam itself is another obvious route to learning the material since the gradee gets to view how other students attempted a solution while being given an official answer. Nearly 82% percent perceived that interacting with other students’ test solutions helped them understand the subjects better. Finally, 71.4% of the class at least felt somewhat confident that peer grading was beneficial to their learning, while 21.4% were undecided. Taking note of student questions and concerns during the grading process helped further develop and improve the peer grading experience.

Before the grading event happened in class, the aforementioned instructions were given in both auditory and video form and cover learning expectations: how to handle different grading situations, placing comment on the gradee’s exam, and how to handle problems that are not easy to grade. Those instructions include reassuring students that the goal is for them to understand the question and solution, which helps to alleviate students’ trepidation with the process. Using the model developed, 92.9% agreed the instructions were clear. Furthermore, the same percentage of students were clear that expectations of grading were trumped by learning. Using the peer grading model, 98.2% of students received their graded exam within two week of taking the exam. This included one week for Spring Break. If the grader was unable to grade a response, many “did their best” to give partial credit at various portions, but correctly

marked the page with a star on the bottom left of the paper for the instructor to look at personally.

Finally, with students having intimate knowledge of the exam, how well did they predict their own grade to what was actually received? The first exam in Data Structures covers almost eight weeks of material. Because of this, this exam is the hardest out of all three. After grading, 82.6% of all students expected to pass the exam, while in reality, 72% in fact did. (Figure 5)

8. What grade did you expect on the exam prior to the grading session?		
Answer Options	Response Percent	Response Count
90-100%	1.8%	1
80-89%	40.4%	23
70-79%	40.4%	23
60-69%	15.8%	9
Below a 60%	1.8%	1

9. What grade did you receive on the exam?		
Answer Options	Response Percent	Response Count
90-100%	8.8%	5
80-89%	24.6%	14
70-79%	38.6%	22
60-69%	12.3%	7
Below a 60%	15.8%	9

Figure 5. Expected versus received grades on Peer Graded exam

While data collection and analysis is just beginning, results have been favorable in terms of the success of peer grading providing timely feedback (days instead of weeks), and correcting misconceptions in the early stages of the course. With other peer grading practices, there are often concerns that anonymous peer graders will leave inappropriate comments [13]. In the current study this has not been the case, possibly for two reasons: the graders’ initials on the blind “coupon” are stored and only viewed by the instructor, and the rubric leaves little for the grader to independently judge or comment on.

F. Instructor’s ROI

There is always great skepticism in the idea that creating a video rubric is time-consuming. Since the instructor already has much of the prep work done in creating a rubric to do instructor-graded exams, creating the video should be straightforward. Scripting what the speaker will say in the video usually requires a lot of time, but that is also done already since the instructor knows the problem, the solution, how to get to the solution and the grading breakdown. The trick is not to “restart” if an error is made, but to “pause” and start that section again, then cut the error out later. If there

are any minor edits that need to be done to the video or rubric, they can be done within a short amount of time and then reloaded back up to YouTube. There are many common sense ways and simple tricks to get the video created in a timely manner in order to ensure a good RIO for the instructor.

As far as this study was concerned, the author did not have any official training. Camtasia, which the author used, had a two-day training session that covered more than what an instructor would usually use. The online tutorial is only 30 minutes in total [14], and most of what it covers is not necessary for what is required. In addition, the rubric, the test and the solutions are already created, so no extra work was incurred.

The creation of an exam and rubric do not add to this investment since it has to be done in class. But the ability to use this work semester after semester is key for the rest of the work required. The making of videos after practice can be done easily since the established rubric already helps explain and determines the point breakdown. At times, creating the video first for a new question make the hardcopy rubric better.

As for overall grading, the first test in CMSC 341 Data Structures, both classes conducted peer grading. After a few hours to review all tests, the entire time to grade was 4 hours for around 80 students. For the second test, only one class conducted peer grading, while the other class was graded by the instructor. Grading of just around 40 tests took about six hours for the instructor, plus an hour and a half for the peer graded class: already nearly double the time and only 50% were peer graded.

G. Room for Improvement

In this rendition of the study, a few items through data collection or logistics can be improved. And the overall class setup for CMSC 341 Data Structures used in this study is in question.

Since the students had only seen the “blind coupon” once, many students either did not know how to fill it out, or did not complete it in its entirety. This does make it cumbersome for the instructor to correctly identify sections so not to give a test to be peer graded to the wrong section. The ideal setup would be using a projector to visually show instructions where to fill out the coupon while the students are completing the exam.

Graders’ comments needed further clarification or hand holding. Only 40% of students found that their peers’ grading comments were helpful in their overall understanding. And when asked what was helpful, the consistent responses were either the grader did not leave comments or were minimal. To rectify, instructions at the beginning of the grading event were clarified and examples shown so to give a richer feedback to the exam taker. Those instructions asked the graders to write both what exactly was incorrect/correct using the rubrics and to add comments as long as appropriate.

The class structure itself may not be the best fit for this study. Compared to many foundational courses, CMSC 341 Data Structures covers various topics that do not rely on previous knowledge covered only a few weeks ago. After the first test, there is some foundational knowledge reused for the second test. This became apparent when a second small survey taken just after the second test asked if the peer grading in the first exam had helped with the second.

“To be honest, it really didn’t. I see the peer grading for Exam 1 and studying for Exam 2 to be completely different things. I don’t think one has to do with the other. I suppose it helped me know what to sort of expect and how the partial credit works. This may have subconsciously affected my studying for Exam 2.”

The overwhelming majority stated that peer grading helped them understand how the test was to be graded or how points were to be graded.

“Reviewing the exam with other students was constructive criticism toward my own approach in deciphering and attempting the problems on the exam. Peer grading prompts valuable discussion and also provides deeper understanding toward the problems.”

III. RESEARCH DESIGN

A quasi-experimental design has been chosen in which each of three sections of Data Structures will receive different treatment in order to test the hypothesis that exam peer grading is successful. All classes will take the same exams. Class 1 will be the control group and will not experience peer grading. Class 2 will experience peer grading for only the first exam, while Class 3 will peer grade both exams. (The final for all sections is not peer graded.) A comparison of overall final grades will also be recorded. Finally, a portion of a focus group with student graders will be conducted to get empirical data.

ACKNOWLEDGMENT

The author would like to thank the invaluable efforts of Lynn Jensen, Dr. Adam Griffy, Drs. Jennifer Harrison and Kerrie Kephart of UMBC’s Faculty Development Center, and Olivia Wolf of UMBC’s Department of Computer Science and Electrical Engineering.

REFERENCES

- [1] Dunlosky, J, et al. "Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive And Educational Psychology." *Psychological Science In The Public Interest* 14.1 (n.d.): 4-58.
- [2] Gehringer, E.F. "Electronic Peer Review And Peer Grading In Computer-Science Courses." *SIGCSE Bulletin* 33.1 (2001): 139-143.
- [3] Greenberg, Kathleen P. "Rubric Use In Formative Assessment: A Detailed Behavioral Rubric Helps Students Improve Their Scientific Writing Skills." *Teaching Of Psychology* 42.3 (2015): 211-217. Freire, M., A. del Blanco, and B. Fernandez-Manjon. "Serious Games As Edx MOOC Activities." (2014):
- [4] Freeman S, Parks J W. How accurate is peer grading?[J]. *CBE-Life Sciences Education*, 2010, 9(4): 482-488.
- [5] Lu, Jingyan, and Zhidong Zhang. "Understanding The Effectiveness Of Online Peer Assessment: A Path Model." *Journal Of Educational Computing Research* 46.3 (2012): 313-333.
- [6] Phillips, Fred. "The Power Of Giving Feedback: Outcomes From Implementing An Online Peer Assessment System." *Issues In Accounting Education* 31.1 (2016): 1-15.
- [7] March, JL, and MA Erdmann. "Teaching Assistant Training Via A Course Management System: Videos And Rubrics To Increase Consistency." *Abstracts Of Papers Of The American Chemical Society* 247.(n.d.):
- [8] Mulder, Raoul, et al. "How Does Student Peer Review Influence Perceptions, Engagement And Academic Outcomes? A Case Study." *Assessment & Evaluation In Higher Education* 39.6 (2014): 657-677.
- [9] Ran, He, et al. "Peer Grading System Based On Social Relations On MOOC." (2015):
- [10] Topping, Keith. "Peer Assessment between Students in Colleges and Universities." *Review of Educational Research* 1998: 249.
- [11] Xiao, Yun, and Robert Lucking. "The Impact Of Two Types Of Peer Assessment On Students' Performance And Satisfaction Within A Wiki Environment." *The Internet And Higher Education* 11.Special Section of the AERA Education and World Wide Web Special Interest Group (EdWeb/SIG) (2008): 186-193.
- [12] Varshney, L.R., A. Vempaty, and P.K. Varshney. "Assuring Privacy And Reliability In Crowdsourcing With Coding." (2014):
- [13] Yu F Y, Wu C P. Different identity revelation modes in an online peer-assessment learning environment: Effects on perceptions toward assessors, classroom climate and learning activities[J]. *Computers & Education*, 2011, 57(3): 2167-2177.
- [14] Camtasia Studio 8 Tutorials: TechSmith. 2015. <https://www.techsmith.com/tutorial-camtasia-8.html>
- [15] PRAZE:Learning Management System: The University of Melbourne. [Lms.unimelb.edu.au](https://lms.unimelb.edu.au), 2016. <https://lms.unimelb.edu.au/teaching/assessment/praze/>.