Work in Progress: Can In-Class Peer Reviews of Written Assignments Improve Problem Solving and Scientific Writing in a Standard-Based, Sophomore Laboratory Course?

Dr. Casey Jane Ankeny, Northwestern University

Casey J. Ankeny, PhD is an Associate Professor of Instruction at Northwestern University. Casey received her bachelor's degree in Biomedical Engineering from the University of Virginia and her doctorate degree in Biomedical Engineering from Georgia Institute of Technology and Emory University where she studied the role of shear stress in aortic valve disease. Currently, she is investigating equitable student engagement strategies in flipped and traditional biomedical engineering courses. She aspires to understand and improve student attitude, achievement, and persistence in student-centered courses.

Dr. Ken Gentry, Northwestern University

Ken Gentry is a Professor of Instruction and Adviser working mainly with first-year students. He teaches cornerstone design and courses in the biomedical engineering department.

Prof. David P. O'Neill, Northwestern University

David O'Neill is an Associate Professor of Instruction and the Michael Jaharis Director of Experiential Learning for the Biomedical Engineering Department at Northwestern University. David read Engineering Science at University College, Oxford, receiving his M.Eng. and D.Phil. before undertaking a post-doc in the Department of Physiology, Anatomy and Genetics. During these years, he taught undergraduate tutorials for Keble, New, University, and Harris Manchester Colleges, was College Lecturer for New College and a Senior College Lecturer in Engineering Science for Keble College. He taught approximately 75% of the core degree topics, as well as human physiological measurement laboratory classes for medical students.

Philippa Eshun

WIP: Can in-class peer reviews of written assignments improve problem solving and scientific writing in a standard-based, sophomore laboratory course?

Introduction

This work-in-progress study assesses the impact of reflective practices, including peer review using a co-created rubric, on written assignments in a sophomore-level, biomedical engineering laboratory course. As an introduction to experimentation, the course covers the statistical design of experiments and the quantification of measurement data quality. Topics include problem-solving skills, scientific writing, and hypothesis generation amongst other research-related topics. Evidence-based pedagogy used in the course includes standards-based grading and reflection.

This study is motivated by work demonstrating the inclusive and effective nature of peer review, co-created rubrics, and standards-based grading. An earlier study showed a strong positive correlation between instructor review and peer review in a biomedical engineering laboratory, suggesting peer review could be an effective form of feedback [1]. Peer review also resulted in the perceived improvement of students' ability to critique. Additionally, the use of co-created rubrics is an inclusive teaching practice that can improve confidence and self-efficacy. It speeds up future detailed feedback, as the students and instructors have a similar understanding about the elements of the rubric and may enhance self-regulated learning [2]. Finally, standards-based grading shifts the primary objective to individual learning and achievement, removes distraction from low-importance errors and reduces the penalties for those with insufficient background [3].

Previously, we identified weaknesses in the "problem identification" and "interpretation" components of problem solving [4]. As a result, we implemented reflective practices [5] and noticed a trend suggesting that increased reflective practices, namely the addition of reflective engineering notebooks, may have improved student perception of standards-based grading and increased student engagement with fulfilling course standards [6]. To address limitations of our previous study and build upon encouraging results, we are implementing the equitable, reflective strategy of peer review on individually written assignments. We hypothesize that the implementation of peer review for individually written two-page abstracts will result in increased overall class performance in course standards and result in favorable student attitude with regards to the peer review process.

Methods

In-class

At the start of the term, the instructional team and students work together to co-create the "problem-solving" rubric which will be used to evaluate three individually written abstracts. Instructors provide the current rubric [7] comprised of course standards definitions and evidence associated with proficiency for each standard. Students work in groups to provide feedback on both the standard definitions and expected evidence of proficiency. The instructors then compile and review student feedback outside of class and construct the finalized rubric considering the students' feedback. Please see Appendix A for the original and finalized co-created rubrics.

After each lab experiment, students individually draft their abstracts in class for ninety minutes. Then, using Canvas's Peer Review function, each student provides detailed feedback on one randomly assigned draft using the co-created rubric as a guide for thirty minutes. Students can see who they are reviewing and who reviewed their draft. Peer reviewers provide written feedback but no score, or letter-based grading, to follow the evidenced-based process of ungrading [8]. Students can view their received peer-reviews and are thus able to incorporate the feedback before submitting a final version to be graded by the instructional team. The instructional team provides both additional qualitative feedback and a score for each standard (on a scale of 1: does not meet standards to 4: meets standards). The instructional and study teams (two instructors, two teaching assistants, and one research assistant) underwent grader calibration.

Study Analysis

Methodologies used in the study include the analysis of quantitative and qualitative survey data and the analysis of written lab assignments. In addition to the final abstract, the study team grades the draft abstracts on the aforementioned proficiency scale. Draft and final submissions are not blinded to the study team. Scores are converted to Hake gain values to assess the impact of the peer review process on standards proficiency. Hake gain refers to the change in student draft and final submission scores "divided by the maximum possible gain" [9]. The Wilcoxon signed rank test was conducted to look for statistical differences in overall draft and final submission scores.

The study team has developed two instruments to evaluate peer review: 1) an assessment of peer review quality and 2) an attitudinal survey regarding the peer review process. The peer review quality instrument (based on [10] and summarized in Figure 1) will assess the appropriateness and specificity of the criticism as well as the justification for why it is included. Lastly, each review will be assessed for the presence and usefulness of any suggestions. The attitudinal survey (based on [11-13]) contains two sections, one for the critic and one for the critiqued, and covers areas such as utility, assessment of training/grader calibration, impact on future work, and emotion (Appendix B). We investigate correlations with proficiency levels of both the critics and critiqued using Spearman's Rho. This study (IRB #STU00214218) was deemed exempt from continuing oversight by the institutional IRB.

Appropriateness – Is the comment related to rubric criteria?			
2 – Yes, directly related to rubric criteria	1 – Partially related to rubric criteria	0 – No, not related to rubric criteria	
Specificity – Is the comment related to something specific in the draft?			
2 – Yes, directly related to a specific part	1 – Somewhat vaguely related to part of the draft	0 – No, comments are vague	
Justification – Does the comment include justification for why it is included?			
2 – Yes, clear justification based on understanding of rubric	1 – Partial justification given	0 – No, no justification is given	
Suggestion – Does the comment offer suggestion for improvement?			
2 – Yes, helpful suggestion is offered	1 – Unhelpful or partial suggestion is offered	0 – No, no suggestions for improvement offered	
		Total	__/8

Figure 1. Assessment of Peer Review Quality. The study team evaluated the quality of each review on a three-point scale for each of the following elements: appropriateness, specificity, justification, and suggestion. This rubric was used for six of the seven standards (all except teamwork which was evaluated separately).

Results

As shown in Figure 2, we found a statistically significant improvement in problem-solving mastery when comparing the Abstract #1's draft to the final submission (* refers to p=0.000031). Only completed abstracts were included in the analysis (n=19 out of 26). The mean Hake gain for the first abstract is 48 +/– 7 %, suggesting that students earned almost half of the points lost on the draft on the final abstract submission. Additionally, the quality of peer reviews scored 35 +/– 9 out of a maximum score of 48 (maximum of eight points per standard and six standards evaluated). Initial investigation of the correlation between mastery improvement and quality of peer review were not significant for either the critic or the critiqued. Thematic analysis of peer review quality revealed that most of the comments scored "2" on the *appropriate* criterion in Figure 1. There is room for improvement in terms of *specificity*, *justification*, and *suggestions* as many students responded with a simple "yes" in response to rubric criteria. Many of the comments that received lower scores for *suggestions* also received lower scores for *justification*.

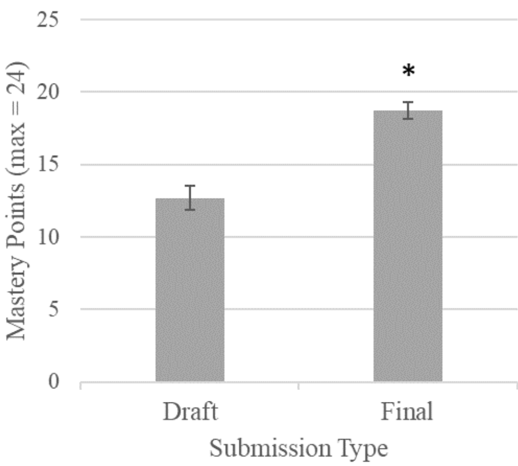

Figure 2. Mastery Points for Draft and Final Submissions. There is a statistically significant improvement in average mastery from the draft to the final (n=19, error bars represent standard error).

Discussion and Conclusion

In summary, analysis on the first abstract shows a significant improvement in mastery between the draft and final that could be partially due to the implementation of co-created rubrics and peer evaluation. A potentially confounding factor could be "time". Some students struggled to complete the abstract within the allotted class time. As students spend more time on the abstract after the peer review, one could hypothesize that this would inherently improve the abstract score. However, this time would be used to implement peer evaluation feedback. Moreover, two of 26 students reached out to the instructor after the peer review session for additional feedback. This could also confound results and will be taken into consideration with future statistical analysis. The quality of the peer review was relatively high with about 35 out of the total 48 points, and analysis yielded areas for improvement. Students will be given the feedback that their comments were *appropriate* but lacked in the *specificity*, *justification*, and *suggestion* categories. They will receive direction on how to provide specific and justified comments with associated suggestions. Lastly, we did not see a correlation between peer review quality and proficiency of standards. Future analysis will 1) investigate correlations between proficiency and each of the four subcategories of peer review quality, 2) include analysis from abstracts #2 and #3, and 3) investigate whether or not there are statistical differences in peer review quality and class performance over time.

In conclusion, this work-in-progress study assesses the implementation of additional reflective practices (peer review and co-created rubrics) with the goal of improving individual proficiency of standards in problem solving and written communication while retaining high student favor.

References

[1] Saterbak, A., & Volz, T. "Implementing Calibrated Peer Review To Enhance Technical Critiquing Skills In A Bioengineering Laboratory Paper", *2008 ASEE Annual Conference & Exposition*, Pittsburgh, PA, USA, 2008, pp. 13.708.1-13.708.13. doi: 10.18260/1-2—3128

[2] Feldman, J. *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Thousand Oaks, CA, USA: Corwin, a SAGE Company, 2019.

[3] Smeding, Darnon, C., Souchal, C., Toczek-Capelle, M.-C., & Butera, F. Reducing the Socio-Economic Status Achievement Gap at University by Promoting Mastery-Oriented Assessment," *PLOS* One, vol. 8, no. 8, pp. e71678–e71678, Aug. 2013, doi: 10.1371/journal.pone.0071678.

[4] Ankeny, C. J., & O'Neill, D. P., & Beckmann, L. (2020, June), WIP: Comparison of a Standards-based Assessment to a Traditional, Summative Rubric in a Biomedical Engineering Laboratory Paper presented at 2020 ASEE Virtual Annual Conference Content Access, Virtual On line . 10.18260/1-2--35529

[5] Diefes-Dux, H. A. and A. R. Carberry, "Cases of Student Reflection within a Course Using Standards-Based Grading," in 2019 IEEE Frontiers in Education Conference (FIE), 2019, pp. 1–9. doi: 10.1109/FIE43999.2019.9028501.

[6] Adkins, A. N., & O'Neill, D. P., & Ankeny, C. J. (2021, July), WIP: Effectiveness of Different Reflection Approaches for Improving Mastery in an Engineering Laboratory Course Paper presented at 2021 ASEE Virtual Annual Conference Content Access, Virtual Conference. https://peer.asee.org/38079

[7] Ankeny, C. J., & O'Neill, D. P. (2019, June), Work in Progress: Aligning and Assessing Learning Objectives for a Biomedical Engineering Course Sequence Using Standards-based Grading within a Learning Management System Paper presented at 2019 ASEE Annual Conference & Exposition, Tampa, Florida. 10.18260/1-2--33591

[8] Coletta, V.P., and J.J. Steinhart, "Why normalized gain should continue to be used in analyzing preinstruction and postinstruction scores on concept inventories," *American Physical Society*, vol. 16, no. 1, Feb. 2020, doi: 10.1103/PhysRevPhysEducRes.16.010108

[9] Kohn, A., & Blum, S.D. *Ungrading*: *Why Rating Students Undermines Learning (and What to Do Instead)*. (First edition. ed.). Morgantown: West Virginia University Press, 2020.

[10] Gielen, S., E. Peeters, F. Dochy, P. Onghena, and K. Struyven, "Improving the effectiveness of peer feedback for learning," *Learning and Instruction*, vol. 20, no. 4, pp. 304–315, Aug. 2010, doi: 10.1016/j.learninstruc.2009.08.007.

[11] Andersson, Magnus, and Maria Weurlander, "Peer Review of Laboratory Reports for Engineering Students," *European Journal of Engineering Education*, vol. 44, no. 3, pp. 417–428, 2019. doi: 10.1080/03043797.2018.1538322

[12] Conde, M.Á., L. Sánchez-González, V. Matellán-Olivera, and F. J. Rodríguez-Lera, "Application of Peer Review Techniques in Engineering Education." *International Journal of Engineering Education*, vol. 33, no. 2, pp. 918–926, 2017.

[13] Larsen, K. and J. Gärdebo, "Retooling Engineering for Social Justice: The Use of Explicit Models for Analytical Thinking, Critical Reflection, and Peer-Review in Swedish Engineering Education," *International Journal of Engineering, Social Justice and Peace*, vol. 5, no.1-2, pp. 13-29, 2017.

Appendices

Appendix A: Co-created Rubric

Problem Identification:

- Definition:
  - Original: Identify the problem and construct a hypothesis that has a strong connection to biomedical engineering/health.
  - Revised: Identify a significant health-related problem and construct an associated hypothesis.
- Original Evidence:
  - Introduction: Relevant Title
  - Introduction: Motivation (meaningful to BME?)
  - Introduction: Statement of objective/gap addressed
  - Introduction: Specific Hypothesis relevant to objective
- Revised Evidence:
  - Relevant title that efficiently describes focus of abstract
  - Motivation: Broader problem context (could include socioeconomic factors if relevant)
  - Motivation: Impact (could be number of people affected or severity of condition)
  - Statement of objective/gap addressed.
  - Specific Hypothesis relevant to objective

Knowledge Processing:

- Definition:
  - Original: Locates, evaluates, integrates, and applies knowledge to support the hypothesis. Assesses the accuracy of conclusions in literature.
  - Revised: Locates, evaluates, integrates, and applies knowledge to support and form the hypothesis. All evidence in support of the hypothesis is delivered succinctly and accurately.
- Original Evidence:
  - Previous work: Integrated summary (correct interpretation of prior work)
  - Previous work: Logical and strong connection to hypothesis
- Revised Evidence:
  - Previous Work: Accurate, relevant, integrated, and paraphrased summary of prior work from multiple sources
  - Previous work: Logical and strong support of hypothesis (in other words, provides strong rationale behind experimental work)

Approach/Experimental Design:

- Definition:
  - Original: Formulate the approach and appropriate experimental design.
  - Revised: Formulate the approach and appropriate experimental design.
- Original Evidence:

- Methods: physical set-up (equipment, settings, images, etc. - as appropriate)
- Methods: experimental set-up (power analysis, randomization, controls, replication, etc.)
- Methods: Considers measurement techniques, variables, blocking, factors/levels, etc.

- Revised Evidence:
  - Methods: Brief overview of approach
  - Methods: physical set-up (equipment, settings, images, etc. - as appropriate)
  - Methods: experimental set-up (power analysis, randomization, controls, replication to ensure high-level of precision/reproducibility, etc.)
  - Methods: Considers measurement techniques, variables, blocking, factors/levels, etc.

Data Analysis

- Definition:
  - Original: Analyzes and graphs appropriately data needed to test the hypothesis.
  - Revised: Analyzes data gathered and includes graph/images/tables to support hypothesis testing.
- Original Evidence:
  - Results: description of analyzed results
    - What are they evaluating the description for (concision, formatting, etc.)?
  - Results graphic(s): Appropriateness
- Revised Evidence:
  - Results: description of analyzed results is accurate and relevant to the hypothesis
  - Results graphic(s): Supports hypothesis testing.

Interpretation

- Definition:
  - Original: Interprets analysis to draw conclusions about hypothesis and ties to greater significance. Includes error considerations, suggested future experiments, etc.
  - Revised: Relates data analysis to the hypothesis (support/does not support) and explains significance of the results.
- Original Evidence:
  - Results logically tied back to the hypothesis.
  - Strong argument for relevance/greater significance
  - Errors are considered.
  - Future experiments and clear why these are meaningful.
  - The main take-home message or conclusion
- Revised Evidence:
  - Discusses data/results supporting or failing to support hypothesis.

- Gives context to findings.
- Includes the strengths and weaknesses (including errors) of experiment.
- Ties results to experimental significance
- Elucidates and supports next steps.
- Delivers the main take-home message or conclusion.

Communication

- Definition:
  - Original: Demonstrates appropriate written and visual communication.
  - Revised: Demonstrates organized and clear written and visual communication appropriate for the targeted audience.
- Original Evidence:
  - Written: Demonstrates clarity
  - Written: Organization
  - Written: Appropriate format
  - Written: Correct scope (appropriate for audience)
  - Written: Presents credible information accurately
  - Written: Uses citations appropriately
  - Graphic(s): Completeness [Axes, Units, legend or key, trendline, etc]
  - Graphic(s): Caption (complete, correct, logical order of info)
- Revised Evidence:
  - Written: writes with clarity and concision
  - Written: writing has a logical order/flow
  - Written: The scope of the abstract is appropriate for audience (necessary information is included, details are not over-explained)
  - Written: Appropriate format.
  - Written: Uses citations appropriately - appropriate format (IEEE, other types)
  - Written: Completeness - no important information missing
  - Graphic(s): placement of graphics
  - Graphic(s): Completeness [Axes, Units, legend, or key, trendline, etc].
  - Graphic(s): Caption (complete, correct, logical order of info)

Teamwork:

- Definition:
  - Original: Demonstrates commitment to high team function
  - Revised: Demonstrates commitment to productively complete tasks as a group
- Original Evidence:
  - Contributed to establishing goals
  - Assisted in planning tasks
  - Accepted Individual Responsibility (completed assigned tasks well by deadline, etc.)
  - Effectively communicated with other team members
- Revised Evidence:

    - TEAMWORK CONCEPTS (self-assessed) DO NOT COMPLETE FOR PEER REVIEW
    - Part 1 (Self): Contributed equally to establishing goals.
    - Part 1 (Self): Assisted in planning tasks.
    - Part 1 (Self): Accepted Individual Responsibility (completed assigned tasks well by deadline, etc.)
    - Part 1 (Self): Effectively communicated and collaborated with other team members.
    - Part 2: Team Member Chart (see assignment for more information)
- Please note that Teamwork is not being evaluated in this work-in-progress. For the course, students write a short paragraph self-evaluating their teamwork. They also evaluate peers' team performance in terms of effort and quality.

Appendix B: Student Attitude Survey about Peer Review Process

Peer Review Assessments:

Likert Scale: Totally Disagree (1) to Totally Agree (4)

For Critic:

1. I understand how to **assess** abstracts in this course.
2. The **training** helped me understand how to **assess** abstracts in this course.
    a. The exemplar of the peer review abstracts in this course…
    b. Classification scheme….
    c. Discussion of fictious abstracts in this course….
    d. Lecture on peer review…
3. It was **useful** for me to see and assess another student's abstract.
4. I feel confident to **assess** abstracts in this course.
5. I **enjoy** giving peer feedback.
6. I feel **reluctant** to give **negative feedback** to my classmates without it being anonymous.
7. I am satisfied with the **overall quality** of the feedback I have given.
8. I believe that it is **important** for me to learn how to give feedback.
9. My **technical writing skills** have **improved** because of reviewing my peers' technical writing.
10. Giving feedback is an effective approach to **improve my critical thinking skills**.
11. Peer review in this course made me feel more **confident to review** technical writing **outside** this course.

For Critiqued:

1. I feel **confident** in writing a good abstract.
2. The **training** in the course helped me to understand how to **write** a good abstract in this course.
    a. The exemplar of the peer review abstracts in this course…

    b. Classification scheme….
    c. Discussion of fictious abstracts in this course….
    d. Lecture on peer review…
3. The review of abstracts in this course that I received was **constructive and fair**.
4. It was tough to have my abstracts in this course **evaluated** by another student **without being anonymous**.
5. It was tough to **write** an abstract in this course knowing it would be evaluated by another student **without being anonymous**.
6. I **enjoyed** receiving peer feedback.
7. I believe that it is **important** for me to learn how to use feedback provided by my peers.
8. Taking feedback is an effective approach to **improve my critical thinking skills**.
9. My **technical writing skills** have **improved** because of having peer review of my technical writing.
10. Overall, my **abstracts have improved** because of the peer review process.
11. I am satisfied with the **overall quality** of the feedback I received.
12. Peers have **adequate knowledge** to evaluate my work.
13. I think I have **learned more from the peers' feedback** than from the instructional teams' feedback.
14. In the **future**, I will have my writing **peer-reviewed** even if I am not required to do so.