

Leveraging Formative Assessments to Improve Student Outcomes

Lessons Learned

Joel R. Hillison¹ and Philip M. Reeves^{2,3}

¹ U.S. Army War College

² Johns Hopkins University

³ Oak Ridge Associated Universities

Abstract

Formative assessments are an effective, if underutilized, way to improve student learning outcomes. Nevertheless, not all formative assessments are the same. This study examines peer review as a formative assessment and the impact on student learning outcomes. The study also looks at the effectiveness of formative assessments in improving student writing skills in a graduate program.

In the past, education at the United States Army War College (USAWC) relied on an instructor-centered model in which teachers transmit information to students (Spooner, 2015). Recently, researchers and practitioners have recommended a shift to student-centered learning (Blumberg, 2009). A key component of student-centered learning is frequent feedback opportunities via formative assessments (Gikandi et al., 2011). Formative assessments occur throughout a course and provide both students and faculty information about learning as it happens (Kelley et al., 2019; Spooner, 2015). This article utilizes a mixed-method approach to examine the impact of using peer review as a formative assessment of student learning outcomes in a graduate-level distance education course at the USAWC.

Background and Literature Review

The USAWC's mission is to promote senior leader development. The War College consists of a resident and distance program accredited by the chairman of the Joint Chiefs of Staff (through the process for accreditation of joint education) and the Middle

States Commission on Higher Education. In the distance program, seminars include up to thirty students with one or two faculty members serving as mentors and evaluators. In the first year of studies, a pool of faculty members evaluates student papers. Therefore, a student will have a different faculty member grade their papers in each course, which allows students to get feedback from multiple faculty members. In the second year of the program, faculty members only evaluate students in their seminar.¹

The college has recently increased its emphasis on formative assessments as a method for applying evidence-based teaching methods and utilizing data to make decisions. According to the college's memorandum on student assessment, "formative events allow students to gauge their understanding of new material as well as learn and practice new skills" (Breckenridge, 2020, p. 3). The memorandum encourages faculty members to provide feedback to students on drafts and outlines before submission for grading.

While the assessment policy acknowledges the role of students "in assessment, critique, and feedback," the memo does not explicitly address peer feedback (Breckenridge, 2020, p. 3). Given the level of experience and expertise of the adult learners in the program, a peer-review process seems like a valuable method for increasing the amount of feedback that students receive without increasing the burden on students or faculty with respect to time. The following sections describe the findings from the literature on peer feedback and its impact on learning outcomes and student writing.

Peer Feedback Has the Potential to Improve Learning Outcomes

The new assessment policy is consistent with research findings, which have found that formative assessments contribute to improved student outcomes and higher teaching quality (Bakula, 2010; Cauley & McMillan, 2010; Cizek, 2010; Petrović et al., 2017). A recent meta-analysis of 54 studies examining learning outcomes found that peer reviews improve academic performance (Double et al., 2020). Student and teacher perceptions also support the use of peer feedback (Brown et al., 2009; Kaufman & Schunn, 2010; Young & Jackman, 2014).

Peer Feedback Has the Potential to Improve Students' Writing

In addition to improving achievement of specific learning outcomes, research indicates that peer reviews can improve student writing (Ibarra-Sáiz et al., 2020; Mirick, 2020; Samarasekara et al., 2020; Wood, 2022). Multi-peer feedback has a greater effect on writing than receiving feedback from an expert or just a single peer (Cho &

¹ Note: This process has changed since the study was first conducted. Now faculty members only evaluate students in their seminar in the first and second years of the program.



LEVERAGING FORMATIVE ASSESSMENTS

Schunn, 2007, p. 15). Additionally, requiring students to make evaluative comments on other student products improves writing more than providing a rating alone (Wooley et al., 2008, p. 2377). Furthermore, by “analyzing other students’ strengths and weaknesses, readers/reviewers become better at recognizing and addressing their own weaknesses” (Ambrose et al., 2010, p. 258). Surprisingly, the value of peer reviews does not appear to depend solely on reviewing superior student products. Peer reviews are more effective when they expose students to both good and bad examples (Zong et al., 2020).

Based on the benefits described in the literature, the USAWC began to incorporate peer feedback as part of its formative assessments. Since the institutional policy does not specify a particular approach, the faculty could experiment with peer reviews in the curriculum. The following sections describe faculty formative assessments and the peer-review process in one course at the USAWC.

Faculty Formative Assessments (Mentoring) at the USAWC

Since at least 2007, the department’s mentoring program provided formative feedback to students who struggled during the first few courses in the distance program. In this program, faculty members provided additional assistance to students who failed the diagnostic essays during orientation or failed to achieve a satisfactory outcome (e.g., scored below a B) on a graded course. These students could voluntarily submit an outline for each writing requirement. Faculty would review the outline and provide narrative comments to the students to help them improve the content and organization of their ideas. The students would then have at least one week to review and incorporate the feedback into the final submission. They could also talk to the faculty who provided the comments for clarification. The course director developed a guide for faculty that included details on answering the questions and a

Dr. Joel R. Hillison serves as a professor of national security studies at the U.S. Army War College. Hillison earned a PhD in international relations from Temple University, an MA in economics from the University of Oklahoma, and an MSS from the U.S. Army War College. He has published numerous articles, podcasts, and book chapters and is the author of *Stepping Up: Burden Sharing by NATO’s Newest Members* (U.S. Army War College Press, 2014) and the lead editor and contributor to *Sustaining America’s Strategic Advantage* (Praeger, 2023).

Dr. Philip M. Reeves serves as an instructor at Johns Hopkins University and a senior project manager at Oak Ridge Associated Universities. Reeves earned a PhD in educational psychology from the Pennsylvania State University. His personal research focuses on topics that intersect social, cognitive, and educational psychology. He has collaborated on research projects that have examined teacher training, collaborative teaching, student help-seeking behavior, cognitive load, metacognition, and school climate.



standard rubric for evaluating student submissions. The course director also conducted a training session to calibrate faculty on the standards to achieve the desired learning outcomes. This program has always been purely voluntary, and there has never been a requirement for students to participate.

Peer Formative Assessments (Peer Feedback) at the USAWC

In the past two years, faculty have experimented with using peer feedback and faculty feedback to improve student learning outcomes and student writing. Since most students struggle with answering the question prompts and organizing their thoughts, the program decided to focus formative assessments on student outlines rather than draft papers in accordance with the USAWC assessment policy (Breckenridge, 2020). At first, the department used a blog or discussion board features of Blackboard for peer reviews. Subsequently, the department employed an online tool designed to facilitate peer feedback, Peerceptiv. Peerceptiv reduced the administrative burden of implementing a peer-review process. The distance program decided to utilize a double-anonymized system, where both the original author and the reviewers remain anonymous. This system reduced the potential for bias and encouraged more honest critical feedback.

It was clear from the initial planning stages that “in order for students to be able to engage in this process effectively, the reviewers need a structure to guide their reading and feedback, the writers need reviews from several readers, and the writers need sufficient time to implement feedback and revise their work” (Ambrose et al., 2010, p. 257). Therefore, the program developed a peer-review process that followed the approach outlined in Ambrose et al.

Structure to Guide Peer Reviews

The course director instituted three strategies to guide peer reviews. First, he developed an instructor guide and a rubric, which included numerical ratings using a Likert scale and narrative comments on each assignment element. He later conducted a training session to teach students how to navigate the peer feedback process and provide useful feedback. Finally, the course director recorded the session for students unable to participate synchronously. Table 1 provides an example of the peer-review prompt.

The Three-Stage Process and Adequate Time

The program developed a three-stage peer-review process to be consistent with a previous study on peer review (Ambrose et al., 2010). In the first stage of the peer-review process, students uploaded outlines in Peerceptiv. In the second stage, students



Table 1

Sample Peer Review Prompt

SCORE	DESCRIPTION
5	The thesis and essay map are complete and clearly outlined and implemented.
4	Between 3 and 5.
3	The thesis and essay map are mostly clear.
2	Between a 1 and 3.
1	The position is vague. The organization of the argument is missing, unclear, or incomplete.

Note. Rate the thesis and essay map in the introduction of this essay. Briefly expand upon your rating of the introduction.

had three days to review other student outlines. Each student evaluated two or three outlines during this period. After completing their final review, students could access their feedback. Again, these were double-anonymized reviews, so students did not know whose outlines they were evaluating, nor did they know who provided comments to them. After seeing the feedback, students gave feedback on the clarity and usefulness of the comments they received. Reviewers could then access this feedback to help them improve their peer-review skills.

In the third and final stage of the peer-review process, students had about a week to incorporate the feedback they received into their papers before submitting the completed assignment. Because the use of peer feedback was new in the program, this study represents the first formal assessment of the effectiveness of the peer-review process at the USAWC.

Based on the literature above, this mixed-method study tested the following hypotheses to determine the effectiveness of the implementation of the peer-review process as a formative assessment:

- H1: Students are satisfied with the peer-feedback process.
- H2: Participation in peer feedback improves student outcomes (grades).
- H3: Participation in formative assessments has an enduring impact on writing skills in subsequent courses, as evidenced by fewer failures.

Methods

This study used a mixed-methods approach to evaluate these hypotheses. The USAWC Institutional Review Board approved the research design. The data included quantitative data on student grades and qualitative reflections from the students via surveys.



To test the first hypothesis, the authors analyzed the results from student surveys to determine students' views of formative assessments. To test the second hypothesis, the authors compared course results to previous years (percentage of failures and grades). The authors used grades from the first two core courses (Strategic Leadership–DE2301 and National Security Policy and Strategy–DE2302) for comparison. Every student completed the summative assessments in DE2301; formative assessments (both mentoring and peer review) were voluntary in DE2302. Unlike the US-AWC resident program, a different faculty member conducted the summative and the formative assessments in these courses.

To test the final hypothesis, the authors compared the course results (percentage of failures and grades) of participants and nonparticipants in the peer-review process in subsequent courses (War and Military Strategy–DE2303 and Global and Regional Issues–DE2304; see Appendix A for course descriptions).

Because the survey examines data across multiple years, it is important to note that the student cohort composition in the distance education program was similar from academic year (AY) 19 through AY24. Most students are military colonels or lieutenant colonels and in either the Army National Guard or Army Reserve. The gender ratios also remained constant. Thus, there were no significant demographic changes before and after the course director introduced peer reviews.

Results

H1: Students are satisfied with the peer feedback.

Results from the student surveys suggested that students were satisfied with the peer-review process; 96% of students were satisfied or very satisfied with faculty feedback and 73% with peer feedback. Students also had the chance to rate the feedback they received in Peerceptiv, on a scale from 1 to 5, based on how helpful that feedback was. Most students (93%) rated their feedback as helpful at a three or higher level, with five being the highest grade. These results are consistent with other previously mentioned studies on peer feedback (Brown et al., 2009; Kaufman & Schunn, 2010; Young & Jackman, 2014).

Descriptive comments suggested that the peer-review process helped students identify gaps in their approaches to the questions and sharpen the focus of their arguments. The following end-of-course survey comments were reflective of sentiments supporting peer reviews:

I (appreciated) the opportunity to review other essays and work to provide productive feedback to my peers. Critically reviewing other essays assisted me to be more critical in reviewing my work.



LEVERAGING FORMATIVE ASSESSMENTS

Table 2

Change in Grades Over the Past Six Years

Academic Year (AY)	Total High Pass/Course Participants	High Pass (%)
AY19	88/398	22.1%
AY20	81/417	19.4%
AY21	107/402	26.6%
AY22	85/401	21.2%

It was a useful tool and receiving peer feedback helped to see different perspectives from the other students. I found it very valuable.

Comments that were skeptical of the peer-review process centered around concerns about the “blind leading the blind” and the varying quality of peer feedback. In some cases, the feedback was neither specific nor constructive. There were also instances where one set of feedback contradicted another.

H2: Participation in faculty and peer feedback improves student outcomes (grades).

Results suggested that faculty and peer formative assessments positively impacted student learning outcomes as measured by course grades. In the first two years of incorporating faculty and peer formative assessments in DE2302 (AY23 and AY24), excellent grades (A- and above) increased from 21% to 43% of student summative assessments and failure rates dropped from a four-year average of 12.4% (from AY19 to AY22) to an average of 7.4% (from AY23 to AY24). About 58% of students participated in the voluntary formative assessments in AY23 and AY24.

Rates of high passing (grades of A and A+) increased significantly in DE2302 from the period AY19 to AY22 (before peer feedback) to the period AY23 to AY24 (after peer feedback) (see Table 2).

The authors used a logistic regression equation to evaluate individual student outcomes in DE2302. The independent variables were grades in DE2301 (high pass, pass, or fail) to account for student potential before formative assessment; participation in Peerceptiv (yes or no) to account for the impact of peer feedback; and participation in mentoring (yes or no) to account for the effect of faculty feedback. The dependent variable was the student grade in DE2302 (high pass versus low pass and fail).

The result using these variables was not statistically significant (McFadden $R^2 = .01$, $X^2 [3] = 6.7$, $p = .08$). Participating in mentoring was the only statistically signif-



icant predictor $\beta = 1.5, p = .006$. When participating in mentoring (i.e., going from 0 to 1), the odds of achieving a high pass grade versus the combination of the other two grade categories are 4.4 times greater when holding the other variables in the model constant. When participating in Peerceptiv (i.e., going from 0 to 1), the odds of achieving a high pass grade decreased slightly, to 0.7 (See Table 3). However, we cannot rule out random error when interpreting the results because participating in Peerceptiv was not statistically significant. These results were also robust using other statistical analyses (see Appendix B).

H3: Participation in formative assessments has an enduring impact on writing skills in subsequent courses, as evidenced by fewer failures.

Finally, the authors compared course grades (DE2301 to DE2304) before and after the course director introduced peer reviews in AY23 to see if they led to any changes in the number of failures. Table 4 contains the results from the past six years.

Average failure rates dropped significantly in DE2302 in AY23 and AY24 (after the introduction of peer reviews) from the average in the period AY19 to AY22. (The spike in failure rates in AY22 appears to have been the result of a poorly worded essay prompt.) The proportion of participants who failed DE2302 differed across years, $X^2(5, \mathcal{N} = 2424) = 26.9, p < .0001$. The number of failures was fewer than expected based on historical trends in AY23 (expected 41.8, actual 25) and AY24 (expected 44.6, actual 35).

While failure rates dropped in DE2302 after the introduction of peer reviews, the failure rates increased significantly in DE2303. Most of that increase occurred between AY22 and AY24. This suggests that any positive impact of peer feedback did not persist in the following course. There was also not a significant relationship between academic year and failure rates in DE2304 $X^2(5, \mathcal{N} = 2333) = 3.36, p = .645$.

Discussion

H1: Students are satisfied with the peer feedback.

The analysis provided support for the first hypothesis and indicated that students found that receiving peer feedback was valuable. This result was consistent with the previously cited literature. The three-stage peer-review process employed by the course director was consistent with other studies on peer review (Ambrose et al., 2010) and took advantage of an innovative software application (Peerceptiv) to smoothly administer and complete reviews. Given the time limitations of faculty and students, the positive reaction to peer reviews from most students was an important



LEVERAGING FORMATIVE ASSESSMENTS

Table 3

Odds Ratio from Logistic Regression Model to Predict Grade in DE2302

	Odds Ratio	Lower	Upper
Peerceptiv participation	0.7	0.5	1.1
Mentoring participation	4.4	1.5	12.7
Fail vs. High Pass in DE2301	1.1	0.5	2.3
Pass vs. High Pass in DE2301	0.7	0.4	1.0

finding. Faculty also expressed satisfaction with the peer-review process in course after action reviews. That said, course directors could make improvements to the structure of the process to address concerns expressed in the student surveys.

First, the course directors can increase emphasis on how to conduct peer reviews (Sanchez, 2019) and how to integrate feedback into their work. Providing examples of feedback on both strong and poor submissions could increase both the quality of the peer reviews and subsequent student performance (Verleger et al., 2016). Second, a rehearsal using the actual rubrics and sample products might also improve student feedback. This could include practice evaluations and faculty feedback to students during the practice sessions. Third, course directors could monitor the results more closely to ensure that every student receives multiple reviews. Many negative survey comments pointed to only receiving two of the three promised peer reviews. Changing the settings or due dates in Peerceptiv could remedy this issue. As a fail-safe, faculty members could provide the third review where necessary. Finally, faculty can present research findings to students that explain the benefits of the peer-review process. Doing this at the beginning of the program could preempt the “blind-leading-the-blind” comments discussed earlier. For example, Wu & Schunn (2020) analyzed the quality of feedback provided in peer reviews and found that low-quality feedback was infrequent. Providing students with this study and other research on the benefits of the peer-review process (e.g., Double et al., 2020; Ibarra-Sáiz et al., 2020; Mirick, 2020; Samarasekara et al., 2020; Wood, 2022) should “allay concerns about the blind-leading-the-blind in peer feedback” (Wu & Schunn, 2020, p. 1).

H2: Participation in peer feedback improves student outcomes (grades).

The implementation of peer reviews increased the number of high pass grades and reduced the number of failures within DE2302 from an average of 12.4% (AY19 to AY22) to 7.4% (from AY23 to AY24). This result suggests that there were benefits to student performance in DE2302. However, when controlling for prior grades and



Table 4
Course Failures and Resubmission Rates

AY	DE2302		DE2303		DE2304	
	Total failures/ Total students	RESUB RATE (%)	Total failures/ Total students	RESUB RATE (%)	Total failures/ Total students	RESUB RATE (%)
AY19	47/398	11.8%	42/390	10.8%	68/386	17.6%
AY20	45/417	10.8%	51/407	12.5%	81/404	20.0%
AY21	40/402	10.0%	36/393	9.2%	81/390	20.8%
AY22	68/401	17.0%	59/389	15.2%	78/383	20.4%
AY23*	25/390	6.4%	64/375	17.1%	79/367	21.5%
AY24*	35/416	8.4%	68/404	16.8%	71/403	17.6%

Note. * Indicates when peer reviews started in DE2302

participation in faculty mentoring, the results did not indicate that participating in the peer-review process could predict a final grade. This result was surprising, especially considering the literature on the positive impact of peer reviews on student outcomes (Bakula, 2010; Cauley & McMillan, 2010; Cizek, 2010; Deiglmayr, 2018; Double et al., 2020; Petrović et al., 2017).

There are plausible explanations for the result. The limited impact on student performance could be related to a lack of trust in their peers' feedback. Some of this sentiment came out in student surveys. Successful peer review requires students to trust their judgment and the judgment of their peers (Ibarra-Sáiz et al., 2020, p. 140). The recommendations in H1 to improve student satisfaction with peer feedback should also improve student outcomes.

H3: Participation in formative assessments has an enduring impact on writing skills in subsequent courses, as evidenced by fewer failures.

While the average failure rates in DE2302 decreased in the two years with peer reviews (AY23 and AY24), the positive impact did not persist in subsequent courses. Failure rates were worse in DE2303 in AY23 and AY24 and did not change significantly in DE2304 during this period. This result was surprising given the previously cited literature linking peer feedback and improved learning outcomes (Ibarra-Sáiz et al., 2020; Mirick, 2020; Samarasekara et al., 2020; Wood, 2022).



LEVERAGING FORMATIVE ASSESSMENTS

There are several possible explanations for this finding. Multiple institutional policies might have interacted to impact this result. First, the college increased its overall emphasis on writing in AY23 and AY24 and thereby increased the standards expected of students. This certainly resulted in more critical evaluations and may have improved student writing. Second, the course director reduced the number of written requirements from three short essays from AY19 to AY22 (600 words each) to two short essays in AY23 and AY24 (750 words each). This gave students fewer chances to improve their grades over multiple assignments. Finally, the course director added a dedicated writing preparation week to DE2302 before the summative assessment in AY23 and AY24 to enable students to review and incorporate both faculty and peer feedback. Other courses (DE2301, DE2303, or DE2304) did not include dedicated writing weeks. Unfortunately, these changes coincided with the implementation of the peer-review process adding several confounding variables to the study.

Limitations and Peer Feedback Training

There are two main limitations of this study. First, some programs require a standardized test (GRE or GMAT) or a graduate skills diagnostic test for student admissions. This program does not. Each service component holds a board that selects students for enrollment based on a review of past performance in the student's field and the student's potential for future service, but there is no requirement for a graduate skills test. However, students may participate in an orientation program that includes a diagnostic essay. Based on the diagnostic essay results, students may enroll in an additional writing assistance program before the first credit-granting course begins. Since not every student participates in the diagnostic essay program, it is not practical to use this as a control variable for student writing ability.

Second, as described in the previous section, the study occurred in conjunction with several other policy changes at the USAWC. The other policy changes could have impacted the metrics of student learning and skill development. Similarly, parts of the study occurred during the height of the COVID-19 pandemic. While the pandemic did not directly impact educational practices in the distance program, it could have influenced students' ability to manage coursework in relation to other changes that occurred to their personal and professional responsibilities during this time.

While not necessarily a limitation, it is important to note that students conducted peer reviews only in DE2302, not subsequent courses in the first year. The previously cited 2020 study on peer feedback found a positive correlation between feedback frequency and satisfaction with implementation, though not necessarily with improved learning outcomes (Wu & Schunn, 2020, p. 11). Thus, satisfaction with peer review would improve with more opportunities and practice. Additionally, it might be valuable to explore the impact of peer reviews on the full papers instead of outlines alone.



Conclusion

This study describes the outcomes of formative assessments (both faculty and peer) on students of the USAWC. As expected, both faculty and students saw merit in the peer-review process. The process implementation also corresponded with a decrease in failure rates in the course that included peer reviews though there could be other explanations for that finding. The effect that peer reviews had on writing quality was less clear. The faculty will continue to look at ways to improve student outcomes without adding unnecessary burdens on the faculty and students.

Regardless of what future analyses find, USAWC students will be in leadership positions and will provide feedback to subordinates and peers through developmental assessments. This requires critical thinking skills. Peer-review practice in academic environments can enhance those skills. ☞

References

- Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. Jossey-Bass.
- Bakula, N. (2010). The benefits of formative assessments for teaching and learning. *Science Scope*, 34(1), 37–43.
- Blumberg, P. (2009). *Developing learner-centered teaching: A practical guide for faculty*. Jossey-Bass.
- Breckenridge, J. G. (2020). *Student assessment* (Carlisle Barracks Memorandum 623-1). U.S. Army War College.
- Brown, G. T., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. (2009). Use of interactive–informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction*, 19(2), 97–111. <https://doi.org/10.1016/j.learninstruc.2008.02.003>
- Cauley, K. M. & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas*, 83(1), 1–6. <https://doi.org/10.1080/00098650903267784>
- Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers and Education*, 48(3), 409–426. <https://doi.org/10.1016/j.compedu.2005.02.004>
- Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. Andrade & G. J. Cizek (eds.), *Handbook of formative assessment* (pp. 15–29). Routledge.
- Deiglmayr, A. (2018). Instructional scaffolds for learning from formative peer assessment: Effects of core task, peer feedback, and dialogue. *European Journal of Psychology of Education*, 33(1), 185–198. <https://doi.org/10.1007/s10212-017-0355-8>
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32, 481–509. <https://doi.org/10.1007/s10648-019-09510-3>



LEVERAGING FORMATIVE ASSESSMENTS

- Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), 2333–2351. <https://doi.org/10.1016/j.compedu.2011.06.004>
- Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., & Boud, D. (2020). Developing student competence through peer assessment: The role of feedback, self-regulation, and evaluative judgement. *Higher Education*, 80(1), 137–156. <https://doi.org/10.1007/s10734-019-00469-2>
- Kaufman, J. H., & Schunn, C. D. (2010). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science*, 39, 387–406. <https://doi.org/10.1007/s11251-010-9133-6>
- Kelley, K. W., Fowlin, J. M., Tawfik, A. A., & Anderson, M. C. (2019). The role of using formative assessments in problem-based learning: A health sciences education perspective. *Interdisciplinary Journal of Problem-Based Learning*, 13(2). <https://doi.org/10.7771/1541-5015.1814>
- Mirick, R. G. (2020). Teaching note—online peer review: Students' experiences in a writing intensive BSW course. *Journal of Social Work Education*, 56(2), 394–400. <https://doi.org/10.1080/08841233.2020.1813235>
- Petrović, J., Pale, P., & Jeren, B. (2017). Online formative assessments in a digital signal processing course: Effects of feedback type and content difficulty on students learning achievements. *Education and Information Technologies*, 22, 3047–3061. <https://doi.org/10.1007/s10639-016-9571-0>
- Samarasekara, D., Mlsna, T., & Mlsna, D. (2020). Peer review and response: Supporting improved writing skills in environmental chemistry. *Journal of College Science Teaching*, 50(2), 69–77. <https://www.jstor.org/stable/27119243>
- Spooner, E. (2015). *Interactive student-centered learning: A cooperative approach to learning*. Rowman & Littlefield.
- Verleger, M. A., Rodgers, K. J., & Diefes-Dux, H. A. (2016). Selecting effective examples to train students for peer review of open-ended problem solutions. *Journal of Engineering Education*, 105(4), 585–604. <https://doi.org/10.1002/jee.20148>
- Wood, J. (2022). Making peer feedback work: The contribution of technology-mediated dialogic peer feedback to feedback uptake and literacy. *Assessment & Evaluation in Higher Education*, 47(3), 327–346. <https://doi.org/10.1080/02602938.2021.1914544>
- Wooley, R. S., Was, Christopher A., Schunn, C. D., Dalton, D. W., (2008, July 23-26). *The effects of feedback elaboration on the giver of feedback* [Conference presentation]. Cognitive Science Society, Washington, DC, United States.
- Wu, Y. & Schunn, C. (2020). When peers agree, do students listen? The central role of feedback quality and feedback frequency in determining uptake of feedback. *Contemporary Educational Psychology*, 62, Article 101897. <https://doi.org/10.1016/j.cedpsych.2020.101897>
- Young, J. E. J., & Jackman, M. G.-A. (2014). Formative assessment in the Grenadian lower secondary school: Teachers' perceptions, attitudes, and practices. *Assessment in Education: Principles, Policy & Practice*, 21(4), 398–411. <https://doi.org/10.1080/0969594X.2014.919248>
- Zong, Z., Schunn, C., & Wang, Y. (2022). What makes students contribute more peer feedback? The role of within-course experience with peer feedback. *Assessment & Evaluation in Higher Education*, 47(6), 972–983. <https://doi.org/10.1080/02602938.2021.1968792>



Appendix A

Description of Courses

DE2301—Strategic Leadership

The strategic leadership course introduces the students to foundational concepts and analytical frameworks they will use throughout the two-year program. There are two summative assessments for this course. The first is online discussion board participation, and the second is a set of three short essays (600 words each). Faculty members offered eligible students mentoring for all three essays but no peer feedback. The thought process was to have faculty demonstrate what valuable feedback looks like before starting peer feedback in the second course. The grades on this essay represent one variable.

DE2302—National Security Policy and Strategy

The national security policy and strategy course introduces new analytical frameworks (international relations theory and decision-making models). It covers the actors, institutions, and processes in the global and domestic environment and introduces students to the strategy formulation framework. There are also two summative assessments for this course. The first is an online discussion board simulating an interagency policy committee, and the second is a set of two short essays (750 words each). Students could participate in peer feedback using Peerceptiv for the second essay. Faculty members also provided eligible students mentoring for both essays. The grades on this essay represent another variable.

DE2303—War and Military Strategy

The war and military strategy course introduce students to classical theories of war and strategy. It has a case study on World War II and a block devoted to contemporary security challenges. As in the previous courses, there are two summative assessments for this course. The first is an online discussion board, and the second is a set of four short essays (ranging from three hundred words to 750 words). Students do not participate in a peer feedback assessment in this course. However, faculty members offer mentoring to eligible students for both essays. The statistical analysis below includes those paper grades.



DE2304—Global and Regional Issues

This is the last online course before students come to the first two-week resident course. In the first block of this course, students study new analytical frameworks and both conventional (e.g., major power) and nonconventional threats. Once again, there are summative assessments for this course. The first is a timed online exam consisting of three short essays (600 words each), and the second is an online discussion board developing a regional strategy. There is no peer feedback or mentoring provided. The statistical analysis includes the student grades on the exam.



Appendix B

Statistical Analysis

Table 1: The proportion of participants that who achieved excellent grades in DE2302 differed across years, $X^2(5, \mathcal{N} = 2424) = 107.8, p < .0001$. The number of high passes was greater than expected in AY23 (expected 112.6, actual 162) and AY24 (expected 120.1, actual 177).

Table 2: The authors also conducted Kendall Tau and Chi-square analyses to test the relationship between variables. There is a significant correlation between the DE2301 grade and the DE2302 grade ($n = 401, \tau_b = .12, p = .013$) and between participating in Peerceptiv and participating in mentoring ($X^2 [1] = 9.5, p < .001$). There was not a statistically significant relationship between participating in Peerceptiv and grades in DE2302 ($X^2 [2] = 1.4, p = .49$) or between Peerceptiv and grades in DE2301 ($X^2 [2] = 1.0, p = .59$). Given the low correlation coefficients (less than .3), it was not surprising that the ordinal regression equation was not statistically significant.

Table 3: The proportion of participants who failed DE2303 differed across years, $X^2(5, \mathcal{N} = 2358) = 17.9, p = .003$. The number of failures was more than expected in AY23 (expected 50.9, actual 64) and in AY24 (expected 54.8, actual 68). The authors thought that writing might also improve in subsequent courses even without peer reviews. Again, the composition of the cohorts remains consistent across academic years.

