Systematic Analysis of Formative Feedback, Focus on Electrical Engineering Assessments

Miss Bahar Memarian, University of Toronto

Bahar Memarian is a PhD candidate in Industrial Engineering and the Collaborative Specialization in Engineering Education at University of Toronto, Canada. Her primary research interests are in Human Factors and Systems Engineering, specifically their application in Education (e.g. learning outcomes assessment, engineering problem solving). Before that, she completed her MASc. (2015) and BASc. (2012) in Electrical Engineering from University of Toronto.

Dr. Susan McCahan, University of Toronto

Susan McCahan is a Professor in the Department of Mechanical and Industrial Engineering at the University of Toronto. She currently holds the positions of Vice-Provost, Innovations in Undergraduate Education and Vice-Provost, Academic Programs. She received her B.S. (Mechanical Engineering) from Cornell University, and M.S. and Ph.D. (Mechanical Engineering) from Rensselaer Polytechnic Institute. She is a Fellow of the American Association for the Advancement of Science in recognition of contributions to engineering education has been the recipient of several major teaching and teaching leadership awards including the 3M National Teaching Fellowship and the Medal of Distinction in Engineering Education from Engineers Canada.

Systematic Analysis of Formative Feedback, Focus on Electrical Engineering Assessments

Bahar Memarian, Susan McCahan
Department of Mechanical and Industrial Engineering, University of Toronto, Canada
bahar.memarian@utoronto.ca, susan.mccahan@utoronto.ca

Abstract

Inspired by a system thinking approach, in this research paper we re-conceptualize assessment as a set of inputs and outputs and examine the relationships between them. Specifically, we examine if granularity (grading scheme increments) and total mark variations on engineering problems that are set up by the course instructors lead to significantly different formative feedback delivery on student solutions by the assessors across a large collection of graded student work. The data draws on actual graded test samples from three electrical engineering courses; two first year circuits course, and a second-year electromagnetics course from a large public university. The summative and formative quantity of feedback for the graded problem solutions (440 in total) are counted and coded. The finding of pair-wise correlational analysis confirms that assessment inputs (i.e., total mark and level of grading breakdown or granularity in a problem) which are set by the course instructor, significantly influence the formative notations of assessors' feedback but not the summative. Interestingly, the total mark of problem (r_τ=0.43, $p<.05$) has a higher effect on assessors' formative feedback than the granularity of problem (r_τ=0.38, $p<.05$) across the set of naturalistic material collected in electrical engineering. The findings of this study can be further strengthened by applying it across a larger collection of electrical engineering courses. If similar results are seen, this may suggest that research in the field should not presume that granularity is the only factor that influences assessors' feedback. Instead, both total mark and granularity should be taken into consideration and in fact, more attention should be paid to the total mark of the problem. If an electrical engineering problem is more fundamental, for instance, it should not be assigned a relatively low total mark necessarily because assessors seem to regard formative feedback on problems having a low total mark as unworthy of feedback.

Introduction

Assessment in electrical engineering (EE) utilizes a conventional practice. Feedback delivery is mainly achieved by assessors reviewing and evaluating an assessment instance against a problem's ideal solution and marking schema that are set up by the course instructor. The product of this practice can be broadly categorized into two groups; namely summative and formative. Summative relates to the solution grade while formative relates to any feedback communicated on student solution besides the grade.

The influence of assessment setup on the feedback provided by assessors in engineering education, and particularly in electrical engineering, remains to be explored. To date, research in the field has primarily investigated assessment through two methods: 1) exploratory and 2) intervention studies [1]. A limitation of most studies carried out are that they are highly contextualized to a one-time instance and lack a reproducible and generalizable methodology. Further adding to this gap, the findings and conclusions are primarily drawn from indirect means of data collection. Examples are post evaluation surveys of student perceptions and attitudes [2]–[6]. As such little investigation has been done on naturalistic data directly, particularly on tests and assignments that are considered to be the backbone of assessments in engineering.

Engineering tests and assignments primarily utilize closed-ended problems. Closed-ended problems are known to be less complex and of smaller scope than real world (open-ended) engineering problems that are more common in industry. Feedback on closed-ended problems may tend to be more mechanistic and sequential, since the problem has been intentionally structured to have one correct result and an optimal solution strategy. In addition, teaching assistants' (TAs) time is generally constrained by the large number of test papers assigned for evaluation, which may also lead to less feedback. Nonetheless, conventional assessment of closed-ended problem solving is endemic in electrical engineering, and a great deal of assessor time and institutional cost is spent on these types of assessments. It would be therefore useful to examine the trends and patterns of feedback in this area more fully. Understanding complexity of assessment systematically can aid in reporting its specific trouble spots. Future interventions can, in turn, focus on addressing more explicit gaps.

Methods

The goal of this paper is to identify factors set by the course instructor that impact the feedback assessors provide on closed-ended problem solving work. We propose a system thinking approach to carryout coding and analysis of naturalistic feedback provided on assessment instances (i.e., tests). A system thinking approach offers tools and principles that are conveniently aligned with values of pedagogical practices [9],[10]. Broadly speaking, we regard assessment as a set of phases, namely:

1. Problem set up; instructor expertise and program requirements impact problem quality
2. Marking scheme; instructor expertise impacts solution manual and grading scheme quality
3. Test taking; student performance and situational factors impact solution quality
4. Assessment and feedback; assessor expertise and training impact feedback quality

Completion of each phase and moving to the next one is dependent on the work progress of user groups involved in each phase. Examples of user groups are instructors, students, and assessors. At each phase the inputs and outputs are influenced by decisions and actions of one or multiple user groups, the environment, and possibly technologies. The environment may include grading policies, and norms within an institution that may influence decisions and hence assessment outputs. For example, one instructor may strictly enforce its course average to be around a certain percentage (i.e. 75%), while another instructor may not have this rigid preference. The teaching assistants, knowing their course's evaluation practice preference, may consequently adopt a different approach to what they consider a high and low achieving solution strategy. We therefore propose applying a system thinking approach on a collected set of naturalistic assessment data, as shown in Figure 1. These inputs and outputs afford metrics that can be used to characterize the system independent of assessment instrument, material, or assessors involved.

There are two inputs under analysis for this study, namely the total mark of the problem and the number of mark breakdowns set by the course instructor in the grading scheme (hereafter referred to as grain or granularity). There are various other characteristics that can be coded as inputs and included in the analysis. Examples include, but are not limited to, the: a) visual degree of problem (i.e. visual model or framework presented or not), b) flow of student solution (legible or not), c) dependency of problem (part a-b-c question or an independent question) and so on. For this study we limit our analysis to the total mark of the problem and granularity, as we find these two metrics to be almost always readily available. Our particular interest is in examining how changes in assessment inputs relate to assessment outputs.

The two outputs under investigation are formative feedback, and summative grade. Formative feedback is derived by counting the freeform feedback pieces assessors provided per problem solution irrespective of its nature (i.e. checkmarks, cross marks, textual pieces etc.). Summative feedback of each problem solution is obtained by taking the grade and classifying it into a five-point scale: Fails (0/4), Below (1/4), Average (2/4), Meets (3/4) and Exceeds (4/4) expectations for performance level.

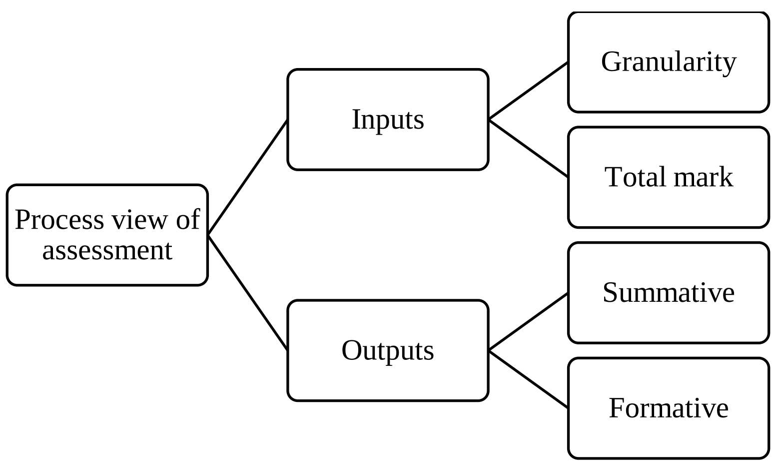

Figure 1 Decomposition adopted for assessment inputs and outputs

The system thinking approach in our work is therefore intended to serve two methodological purposes:

- To examine the functional characteristics of assessment processes
- To collect and analyze data based on the underlying goals of "Assessment for Learning"

Assessment for learning is a well established concept in literature, with formative feedback taking precedence over the use of assessment solely for summative grading [9]. To accomplish these purposes, data is collected from the actual environment in which the process is happening and objectively analyzed to understand the characteristics of the process. A systems approach has been shown to be useful in assessing system reliability, redesigning processes, estimating workload and training requirements in engineering workplaces [10]. We suggest that it may be beneficial for engineering education research as well.

The overview of data collection and analysis methodology is illustrated in Figure 2. The findings of this study are specific to actual midterm and final exam samples from 3 EE courses collected from a large public university. The details of this approach are provided in a concurrent publication [11]. All of the courses represent foundational, core elements of an EE program. A total of 21 problems were selected from the assessments across the courses. The courses are a first-year circuits course (EE1, problems 1-7), a similar circuits course taught in a different program (EE2, problems 8-15), and a second-year electromagnetics course (EE3, problems 16-21). Overall between 16 and 27 graded student solutions were randomly picked from each assessment, leading to a total sample size of 440 graded problem solutions. Before conducting the study, a research protocol approved by the Research Ethics Board (ID: 37507). Both descriptive and inferential tests were carried out on the collected data.

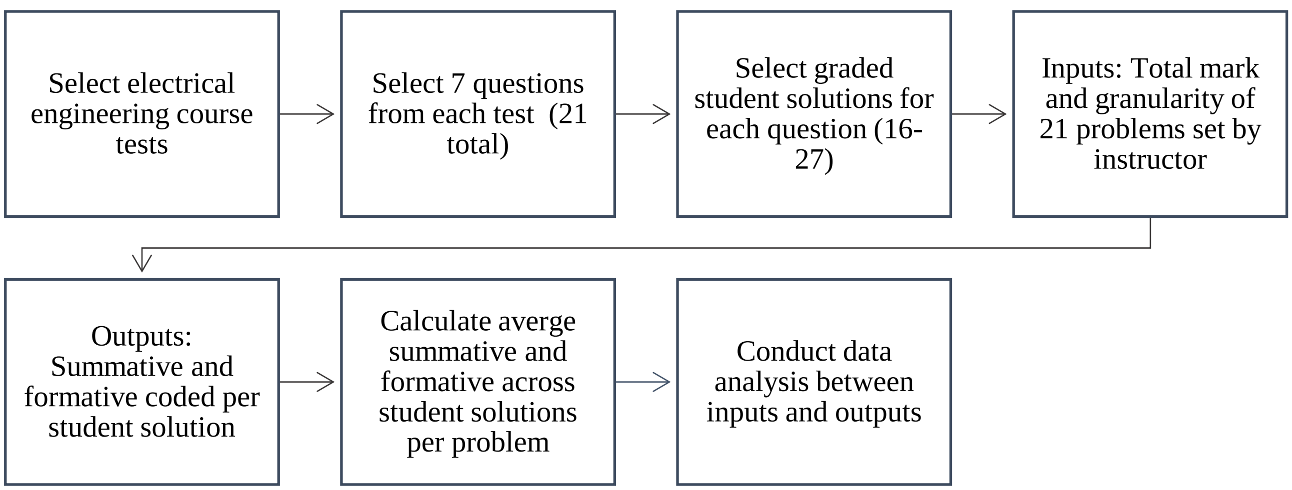

Figure 2 Overview of methodology for coding and examining conventional practice

Results and Discussion

The histogram summaries of assessment inputs for the 21 EE problems are illustrated in Figure 3 (Granularity top left and Total Mark top right). The Total Marks of the problems varied between 2 to 10, with higher proportion (5) having a total mark of 5, followed by a total mark of 2, 4, and 10 (3 each). The granularity set by the course instructor for the 21 EE problems varied from 1 to 13. A granularity of 1 implies that only the total mark of problem was given. For a short answer or multiple-choice question, this may not pose an issue, but for complex problem-solving tasks that were examined in this study, this may signal poor assessment set up. Three out of the 21 problems had the highest granularity (9 or more). However, this is less than the number of problems that had the highest total marks of 8 or 10 (5), suggesting that some large size problems had less detailed granularities. Overall it could be surmised that EE test problems mostly tend to have a granularity between 1 to 7 when the total mark varies between 2 to 6.

In addition to summarizing data descriptively, the values of input and output variables for each of the 21 EE problems were visualized as shown in Figure 4. As mentioned earlier, problems 1-7 pertained to a first-year circuits courses (midterm from EE1), problems 8-15 to another first-year circuits course (final exam from EE2) and problem 16-21 pertained to a second-year electromagnetics course (midterm from EE3). The average feedback and summative mark given to each student solution are data reproduced from a concurrent publication to demonstrate the relationship between these output variables and the input variables being examined here [11].

Further to descriptive analysis, we wanted to see if changes in Granularity and/or Total Mark of problem have a direct relationship with the Formative and Summative feedback outputs. This motivated us to carryout correlational analysis for each of the problems. Since each problem contained student solutions (n between 16 to 27) with diverse performance qualities, any significant correlations between assessment inputs and outputs were generalizable to the EE problems under examination, making the analysis systematic. The results of Kendal Tau correlation analysis on non-parametric data conducted on the 21 EE problems instances are presented in Figure 5. The correlation plot reveals that the Granularity of an engineering problem highly correlates with the Total Mark of that problem. This is reasonable as the higher the Total Mark of a problem, the more space is made available to make the grading scheme finer grained. What are the most interesting findings are the other two significantly correlating results: Granularity (GRAN) vs. Mean Formative (FORM) and Total Mark (TOT) vs. Mean Formative (FORM). The findings suggest that the formative feedback of assessors or put another way the freeform commentary they provide on student solutions during grading is significantly influenced by the problem's set level of grain as well as the total mark or put another way size of the problem. Total mark of the problem ($r_\tau=0.43$, $p<.05$) had a slightly higher effect on formative feedback provided by the assessors in comparison to the granularity of problem ($r_\tau=0.38$, $p<.05$) across the naturalistic pool of data collected in electrical engineering. What is worth mentioning is that in literature, the benefits of making grading schemes finer grained has been explored [12]. These works suggest that increasing either the total mark using an adequate degree of granularity for the grading schemes may lead to more consistent evaluations between assessors. For the purpose of providing formative feedback, we see that the total mark of problem in fact has a higher impact on the formative validity of assessments, in the context of three courses examined at least. It is important to note however, that the findings do not suggest that reliability of conventional assessment may be improved by increasing the grain or total mark per se, but rather the degree of formative feedback provided by the assessors. In addition, the results are relatable to collective performances of assessors, in broad terms, and the specific individual differences of assessors are not examined in this study. These are considerations for future studies.

When looking at one assessment instance with just one granularity and total mark, the formative work of assessors is prone to be influenced by situational solution factors. This is an important factor to take into consideration. In technical domains such as in EE, it is reasoned that a closed ended problem's solution has a set number of steps deterministically, leaving little room for assessors to face any difficulty with evaluating student work. An example is the necessary and specific procedure to carryout mesh analysis for a circuits problem. It may be posited that there are prescribed steps to a mesh analysis, and a student solution would receive or lose points based on the completeness of their work. Broadly speaking this perspective may seem reasonable. However, what also needs to be considered is that the way students present and detail their solution strategy is unique and variable. The work of assessors as a result gets more complicated because they often need to be decrypting and interpreting student's solution thought process and solution strategy, particularly on tests and exams which are time constrained for the students taking them.

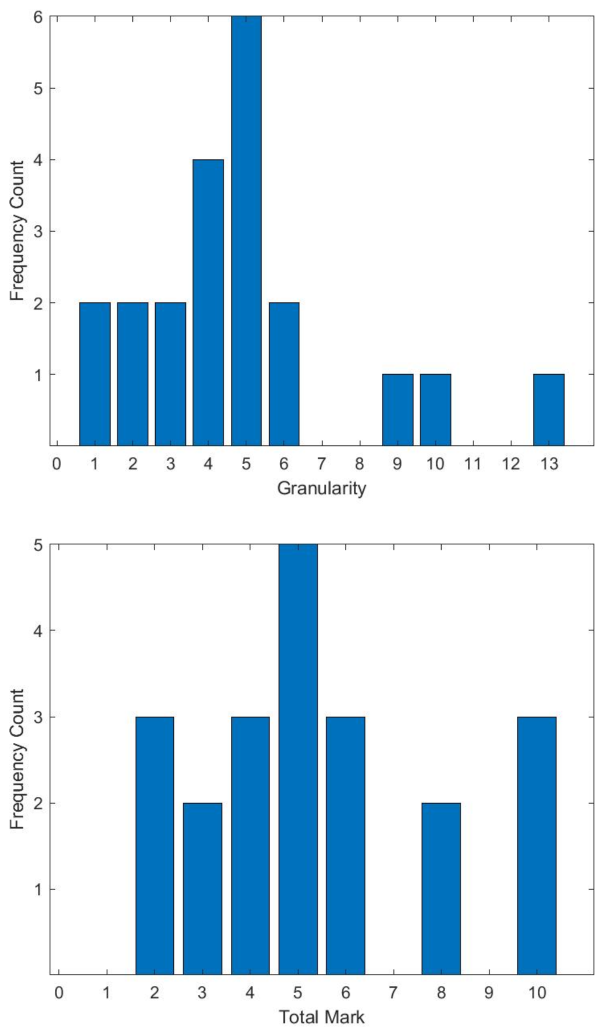

Figure 3 Histograms of assessment inputs; Granularity (top) and Total Mark (bottom)

Granularity	Total Mark	Problem
13	8	21
5	5	20
5	5	19
6	6	18
10	9	17
5	5	16
4	5	15
4	5	14
4	4	13
2	5	12
5	8	11
9	10	10
5	8	9
2	2	8
5	6	7
4	4	6
3	3	5
3	3	4
1	2	3
1	2	2
6	6	1

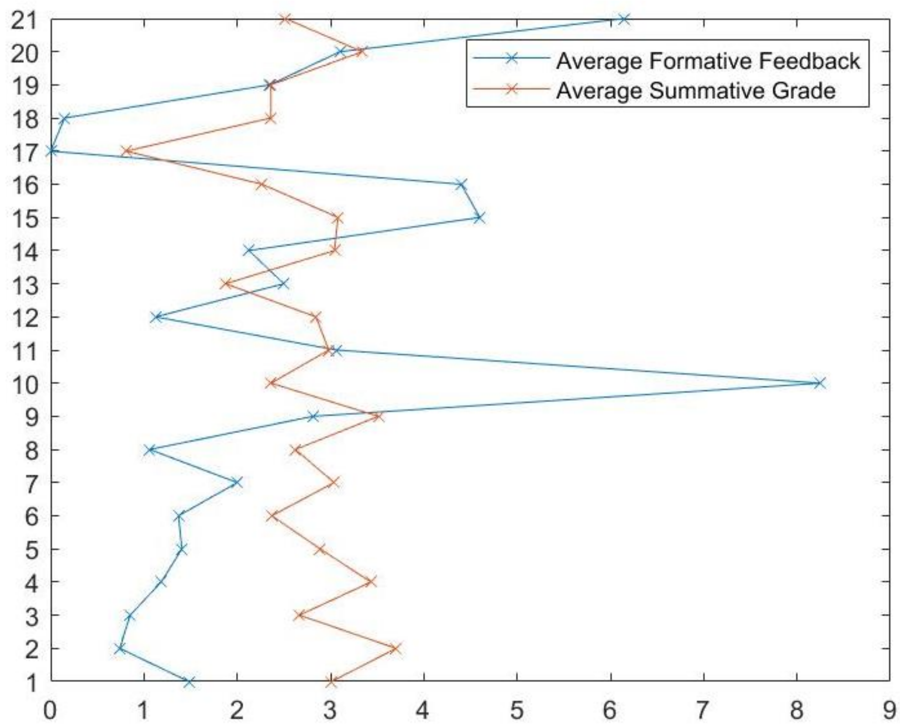

Figure 4 Demographic of Engineering Problem and Assessment Guide. Average formative feedback and average summative grade are taken from Memarian and McCahan, 2019 [11].

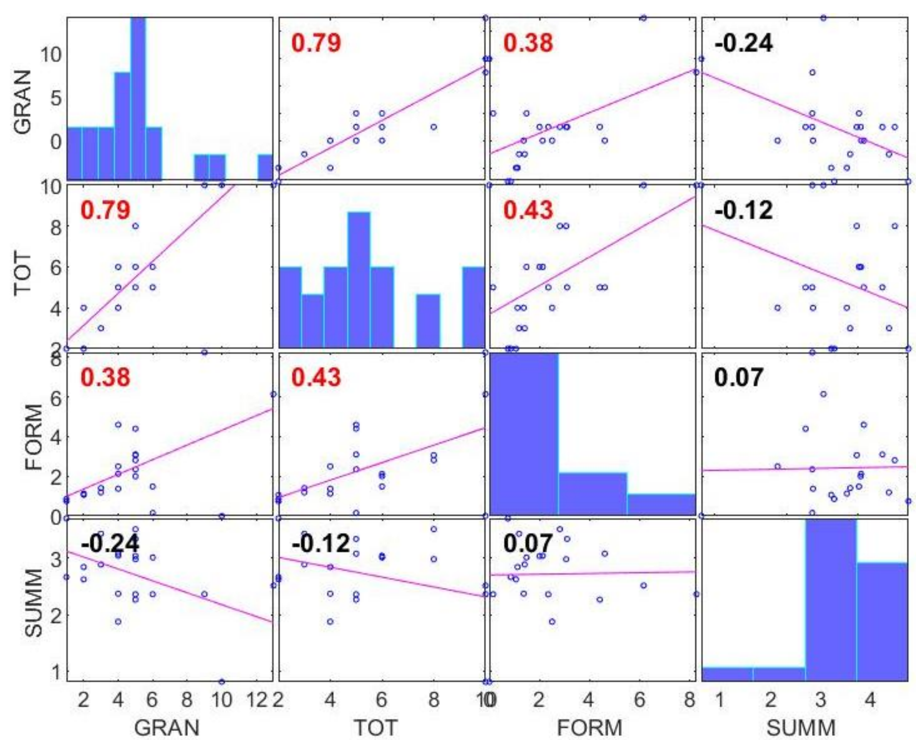

Figure 5 Correlation analysis on Problem Complexity, Granularity, and Feedback Quality

Conclusion

The goal of this study was to explore if changes in the level of detail in evaluation guidance that are set by the course instructors (Granularity or GRAN) along with the total mark (Total Mark or TOT) of a problem influence the formative (Formative) feedback provided by assessors across student performance levels (Summative). In this study we specifically focused on conventional grading practice which is commonly used by assessors in electrical engineering (EE) courses. The statistical analysis results suggested that both total mark and level of granularity in a problem, which are set by the course instructor, influence the formative feedback delivery of assessors. Meaning that when the total mark of problem or level of grading breakdown detail of a problem's grading scheme increases, formative feedback is provided by the assessors at a significantly higher rate. The findings of this study can offer a robust and analytical approach for research in the field. If data is collected and analyzed longitudinally, the findings can shed light on types of shortcomings conventional practice may impose on formative feedback and student learning.

Adopting a system thinking research methodology can help in identifying areas where conventional assessment makes it difficult for assessors to achieve their formative feedback objectives. This can in turn support redesigning assessments to improve factors [10] such as: a) Assessment validity: Such that freeform notations of higher formative value are regarded as feedback, b) Assessment reliability: Such that the degree of process delivery is maintained across product categories and courses appropriately, c) Evaluating staffing requirements or training guides: Such that time and resources are optimally used, d) Estimating workload: Such that problems and Summative performance levels requiring more formative feedback can be taken into account when dealing with assessors' workload. The benefits of systematizing assessment do not only apply to the assessors, but more importantly to the students. What is apparent from literature is that students find some types of comment considerably more usable [13]. A systematized collection of assessment data, when carried out over a course of time period, can aid in tracking the types and amount of feedback that lead to optimal student learning in engineering courses.

References

[1] R. L. Olds, B. M., Moskal, B. M., & Miller, "Assessment in engineering education: Evolution, approaches and future collaborations," *J. Eng. Educ.*, vol. 94, no. 1, pp. 13–25, 2005.

[2] K. Chanock, "Comments on essays: do students understand what tutors write?," *Teach. High. Educ.*, vol. 5, no. 1, pp. 95–105, 2000.

[3] P. Orsmond, S. Merry, and K. Reiling, "A study in self-assessment: tutor and students' perceptions of performance criteria," *Assess. Eval. High. Educ.*, vol. 22, no. 4, pp. 357–368, 1997.

[4] S. E. Pitts, "'Testing, testing...' How do students use written feedback?," *Act. Learn. High. Educ.*, vol. 6, no. 3, pp. 218–229, 2005.

[5] K. J. Rodgers, A. K. Horvath, H. Jung, A. S. Fry, H. Diefes-Dux, and M. E. Cardella, “Students’ perceptions of and responses to teaching assistant and peer feedback,” *Interdiscip. J. Probl. Learn.*, vol. 9, no. 2, p. 2, 2015.
[6] M. Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, “'It’d be useful, but I wouldn’t use it’: barriers to university students’ feedback seeking and recipience,” *Stud. High. Educ.*, vol. 42, no. 11, pp. 2026–2041, 2017.
[7] J. Biggs, “Enhancing teaching through constructive alignment,” *High. Educ.*, vol. 32, no. 3, pp. 347–364, 1996.
[8] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, DC: National Academy Press, 2000.
[9] V. J. Shute, “Focus on formative feedback,” *Rev. Educ. Res.*, vol. 78, no. 1, pp. 153–189, 2008.
[10] J. D. Lee, C. D. Wickens, Y. Liu, and L. N. Boyle, *Designing for people: an introduction to human factors engineering*. CreateSpace, 2017.
[11] B. Memarian and S. McCahan, “Analysis of Feedback Quality on Engineering Problem Solving Tasks,” in *Proceedings of the American Society of Engineering Education (ASEE) Annual Meeting*, 2019.
[12] N. Falkner, R. Vivian, D. Piper, and K. Falkner, “Increasing the effectiveness of automated assessment by increasing marking granularity and feedback units,” in *In Proceedings of the 45th ACM technical symposium on Computer science education*, 2014.
[13] M. Walker, “An investigation into written comments on assignments: do students find them usable?,” *Assess. Eval. High. Educ.*, vol. 34, no. 1, pp. 67–78, 2009.