# Analysis of Feedback Quality on Engineering Problem-solving Tasks

**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.

# Analysis of Feedback Quality on Engineering Problem Solving Tasks

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

## Abstract

In this research paper, we examine the types of feedback provided to students on engineering problem solving tasks. The *assessment for learning* conceptual framework is adopted, which suggests that assessment tasks should further learning rather than being only summative. In this work the feedback observed on marked midterm tests and final exam papers is coded to investigate whether the feedback aligns with an assessment for learning approach. The types of feedback on the papers is characterized using a hierarchical schema with check marks (basic validating feedback) being the least effective, and textual comments (elaborating feedback) being the most effective. The proposed classification is then used to code graded student test papers (naturalistic material) from three electrical engineering courses. The material includes 7 problems from each course, leading to 21 engineering problems in total. Between 16 and 27 graded student solutions are randomly selected for analysis for each problem. The results demonstrate that poor quality student solutions receive less, and less valuable feedback than high quality student work. The results also exhibit a high degree of variability between types of feedback provided on student work. The findings of this study are useful in informing instructional design and changes to assessment practices.

## Introduction

It has been established that not all forms of feedback are equally effective in promoting learning [1]–[6]. Using only grades to improve learning, for example, is not an effective approach [7], [8]. Black and Wiliam's review of 250 studies pertaining to formative assessment found that descriptive feedback targeted to specific solution errors can improve student learning and retention, particularly for underperforming students, as compared to grades alone [9].

There is growing interest in *assessment for learning* as a conceptual framework for the design of assessment activities and the integration of assessment in courses. Assessment for learning supports the adoption of evaluation and feedback practices that improve student learning going forward, and assessment that is, in and of itself, a learning activity [10]. As one aspect of this "feed forward" approach [11], assessment for learning aims to not only produce a grade as a product of a summative evaluation, but also support improved performance in the future. Assessment becomes not just an add-on to the course design, but a central component of the design. This is aligned with a backward design method and Bigg's constructive alignment framework [12], [13].

Conventionally, assessment is seen as a product of performance or summative activity, while the assessment for learning conceptual framework suggests viewing assessment as formative and as part of the process of learning. Assessment feedback, which includes all information exchanges [14], can be explicit, e.g. use clear specific textual phrases, or be more implicit, such as the use

of a grade deduction to indicate a point of error. When adopting this perspective, assessment becomes a cyclic communication between assessors and students instead of a collection of discrete grading instances [15], [16]. An ideal version of this cycle has been described by Hewson and Little and is part of an aligned system vis-à-vis Bigg's theory of Constructive Alignment [11], [12], [17], [18].

Despite growing awareness of assessment for learning, evidence of improved feedback practices continues to be missing [19]. For instance, a lack of alignment between formative and summative assessment has been noted [20]. Performance based assessment has been extensively studied, in particular, the use of formative assessment tools such as rubrics to provide feedback on student work [21]. However, rubrics are more commonly used on writing assignments, or open-ended projects, such as design reports. Courses that emphasize content (i.e. facts, and the application of physical and mathematical concepts) are common in engineering curricula. In many engineering courses students are asked to solve closed-ended problems to demonstrate their mastery of the material in these types of "fact and principle" courses. Given that the nature of the feedback that engineering students receive has not be investigated methodically, we undertook an analysis of the formative feedback provided to students on major engineering assessments involving problem solving. We wish to examine the alignment between summative and formative marking in engineering assessment practice.

In our research, we are considering the way test questions are used as learning activities for engineering students in "fact and principle" courses. The process of marking tests is often guided by a grading scheme based on an ideal solution which is set by the course instructor. Assessors' notations on a student's paper in conventional practice are a combination of a summative mark, and formative information. The summative mark is the total points a student solution receives. The formative information is any additional messages communicated on the student solution (i.e. descriptive text) other than the final grade. The formative information is often provided as unstructured, freeform notations (cross marks, checks, or short phrases). Literature in this field has classified formative feedback of this kind as corrective, which is more symbolic and carries implicit meaning, or descriptive, which is more textual and expresses meaning explicitly [22]. In our work a large sample of graded student work in electrical engineering course assessments (i.e. midterm tests, and final exams) are coded using a frequency count of qualitative notations on the student papers. The objective of this study is to deepen understanding of current practices for the purpose of supporting changes in the way these assessments are used as learning activities.


## Methods

A method was developed to address the following questions: How does the quantity of feedback of different forms (i.e. cross-marks, text phrase, etc.) vary:
1. in relation to solution quality which is measured using the summative mark?
2. in relation to the problems being graded?

A research protocol was submitted and accepted by the Research Ethics Board (ID: 37507). We examined graded student papers from three electrical engineering courses at a large public university. The study used 7 problems from each course, resulting in 21 unique engineering

problems in total. Between 16 and 27 graded student solutions were selected at random for each of the problems yielding 440 graded student solutions in total. The three engineering courses examined included two first year circuits courses, and one second year electromagnetics course. We identify these courses as EE1, EE2, and EE3, respectively, in this paper. These are fundamental (core) courses in the curriculum. The problems examined were closed-ended problems having one correct end result. For example, the type of question that requires the analysis of a circuit to find the current flowing through a given point. The problems selected necessitated a combination of conceptual, factual, and procedural cognitive processing, and to some degree metacognitive processing. These skills are integral components of knowledge construction for problem solving [23]. The feedback provided on these papers is representative of the feedback given by teaching assistants who have received no special instruction related to providing feedback. This is consistent with the preparation of teaching assistants in these courses. There is a teaching assistant training program available at our campus, but it is optional. The teaching assistants are graduate students or upper year undergraduate students who are specializing in the field.

After obtaining samples of naturalistic material, we devised an approach for data coding. The summative mark given to each of the 440 student solutions and types of formative feedback provided by the assessors were recorded. The coding schema that was applied to the freeform feedback notations on each paper is shown in Figure 1. From the literature in assessment, and particularly assessment for learning, the quality of feedback improves as assessors change their feedback strategy from a symbolic and verification form to one that is more formative and elaborative [22]. That is, the feedback quality is improved as the meaning of the information goes from being implied and implicit to being explicit. Based on this framework, and observations of frequent forms of feedback seen on the naturalistic samples available, a hierarchy was developed that characterized the formative value of feedback from least valuable to most valuable. A check mark alone, for example, is the least valuable form of feedback because the meaning is least explicit and specific. It does not explicitly provide information to the student about the number or nature of the errors in their solution and neither does it provide clear information about what they did well. On the other end of the spectrum, feedback can be communicated most explicitly through use of textual phrases that explain the issue specifically, (i.e. elaborate).
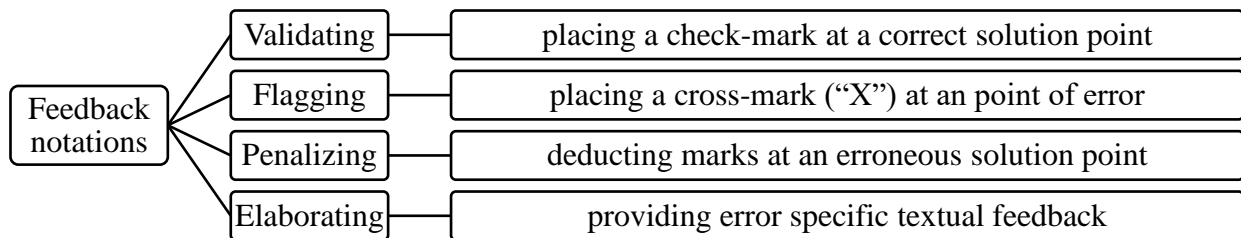


Figure 1. Coding schema used to classify formative feedback notations.

Because there is generally little guidance given to the teaching assistants about the amount of feedback that is expected, it can vary widely. In an unstructured feedback system, at any point in the student's work an assessor may choose to provide one or multiple types of feedback. To collect the data for this study, we considered each notation on the student's work as a datum and

recorded the type of feedback it represented using the schema in Figure 1. We did not record the relationships between the notations (e.g. a check mark with an elaborating text phrase was recorded as two unique data). This methodology could be considered a type of task analysis. In this case we are analyzing the product of a task (i.e. marks on a paper) to understand the task itself (i.e. the process of grading the paper).

In assessment research it is common to divide the learners into quartiles by performance. However, in this case dividing into quartiles does not work because the total points available, and the distribution of marks, was insufficient to allow us to divide the learners in this way. Instead, in order to compare the data across the 21 different test problems, the student scores were normalized to a 0 to 4-point scale. This was done by dividing the student's score by the ideal score for the particular problem, which produced a percentage, and then multiplying by 4. This created 5 performance categories:

1. 0 to 0.49 – fails, which is often no solution at all, or mostly blank
2. 0.50 to 1.49 – unsatisfactory
3. 1.50 to 2.49 – marginal
4. 2.50 to 3.49 – good, but not fully correct
5. 3.50 to 4. – which generally represents a perfect or near perfect correct solution

It is the three categories in the middle of this range that represent the data of most interest because in this range the student has provided some work to mark, but clearly has a need for formative feedback to improve. The data is not evenly distributed across these categories, nor is there a normal distribution that would be easy to characterize using a mean and standard deviation. However, by dividing up the data in this way, we can parse out the ceiling effect at the top end of the scale and focus on the feedback to students in the unsatisfactory, marginal and good categories.

**Results**

Essentially, we were interested in obtaining a count of formative feedback notations to see if they are significantly different in quantity across student performance levels, and courses. We first analyzed the feedback across the five performance levels for all courses combined. We then examined the feedback across the problems and courses represented in the exam material. This was to see if students receive a consistent degree and form of feedback on different problems. The average grade student solutions scored per problem are also reported. Generally, it was assumed that problems with a lower average summative grade indicated a more difficult problem, while a higher grade indicated an easier problem.

On the 440 problem solutions examined, a total of 988 pieces of feedback were coded. Much of the feedback is validating and little of it is elaborating. This is clearly shown in Table 1 which presents the frequency count of feedback notations by performance level. The total number of solutions in each performance level (n-values) and percentage of the total feedback received in each category is also shown. If the feedback was evenly distributed across the performance categories then the percentage in the right-hand column would match the percentage in the left most column, which shows the distribution of the data across the performance levels. Overall, the number of solutions receiving validating feedback is higher for better quality solutions (good

and excellent, n= 342) as compared to lower quality solutions (fails, unsatisfactory, and marginal, n =120). This makes sense because poor solutions probably have little work that shows mastery. However, use of flagging feedback exclusively in an otherwise poor solution (fails, unsatisfactory, and marginal) may be discouraging for students who are struggling. Elaborating feedback can also be called descriptive, and symbolic feedback, which includes Validating, Flagging, and Penalizing notations, are corrective [22]. The total corrective feedback is also shown in Table 1. Throughout the levels there is relatively little descriptive elaborating feedback being given. What we find most interesting is the low number of student solutions in the "fails" category receiving elaborating feedback (4 of 38). Many of the solutions in this category received flagging feedback (17 of 38) instead, which may have little formative value and a negative connotation that may demotivate students.

Table 1. Frequency count of feedback notations across performance levels.

| | Validating (V) | Flagging (F) | Penalizing (P) | Elaborating (E) (Descriptive) | Corrective (V+F+P) | Percent of total feedback |
|---|---|---|---|---|---|---|
| 1. Fails (n=38) 8.6% | 4 | 17 | 8 | 4 | 29 | 3.3% |
| 2. Unsatisfactory (n=41) 9.3% | 23 | 21 | 9 | 6 | 53 | 6.0% |
| 3. Marginal (n=96) 21.8% | 93 | 125 | 96 | 22 | 314 | 34.0% |
| 4. Good (n=66) 15.0% | 92 | 41 | 71 | 36 | 204 | 24.3% |
| 5. Excellent (n=199) 45.2% | 250 | 20 | 32 | 18 | 302 | 32.4% |
| Total (n=440) | 462 | 224 | 216 | 86 | 902 | 100% |

Overall, students in the "marginal" and "good" categories received more feedback on average than those in the "fails" and "unsatisfactory" categories. The uneven nature of the feedback overall is evident in figures 2 and 3 which show the total frequency distribution of formative feedback. A chi-square test was conducted comparing the degree of formative feedback pieces across the five performance levels for the three courses combined (CI=95%). The results show that the degree of corrective feedback provided is significantly higher than the descriptive feedback for all performance levels other than the unsatisfactory group.

The total corrective and descriptive notations for each of the 21 problems analyzed are illustrated in Figure 4. Passing solutions (summative grade >=50%) on average received a higher degree of feedback per problem and particularly more corrective feedback in comparison to failing solutions (summative grade <50%). As mentioned earlier, problems 1 to 7 pertained to a first-year circuits courses (midterm from EE1), problems 8 to15 to another first-year circuits course (final exam from EE2) and problems 16 to 21 pertained to a second-year electromagnetics course (midterm from EE3). For EE1, some degree of elaborating feedback is given for passing solutions across different problems. However, for the same course, failing solutions received a lower and inconsistent degree of elaborating feedback (i.e. problems 5 and 6 high, Problems 1 to

3 none). A similar pattern of higher elaborating feedback for passing solutions in comparison to failing ones can be seen for EE2 and EE3 as well.
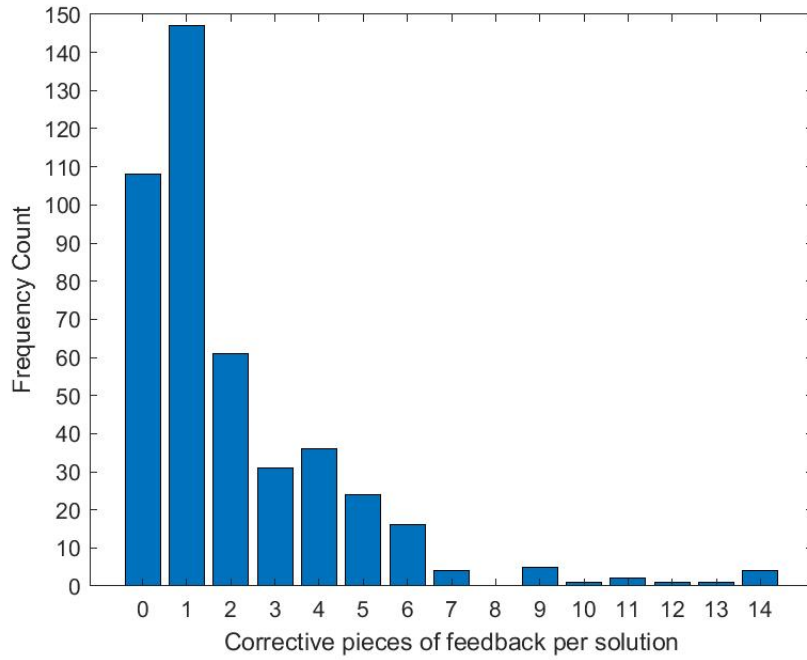


Figure 2. Total number of corrective feedback notations on each student solution. The frequency count is the number of student solutions.
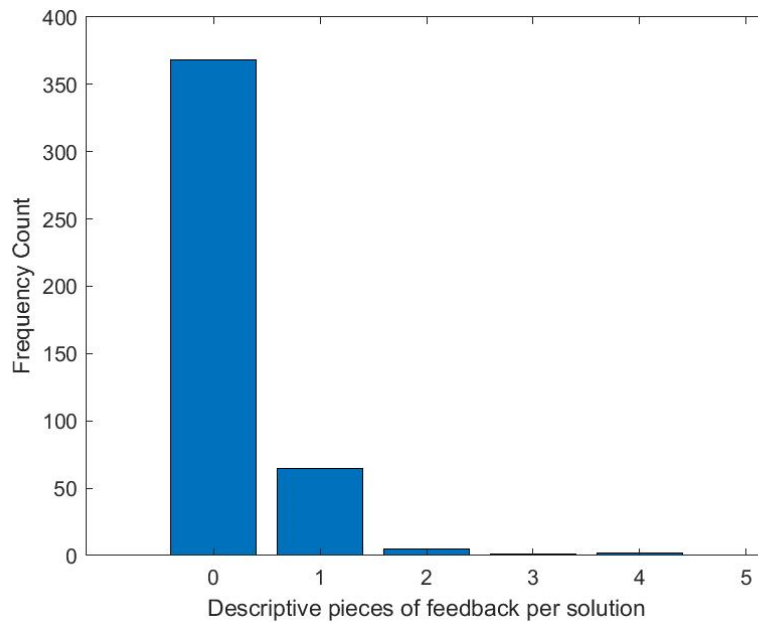


Figure 3. Total number of descriptive feedback notations on each student solution. The frequency count is the number of student solutions.
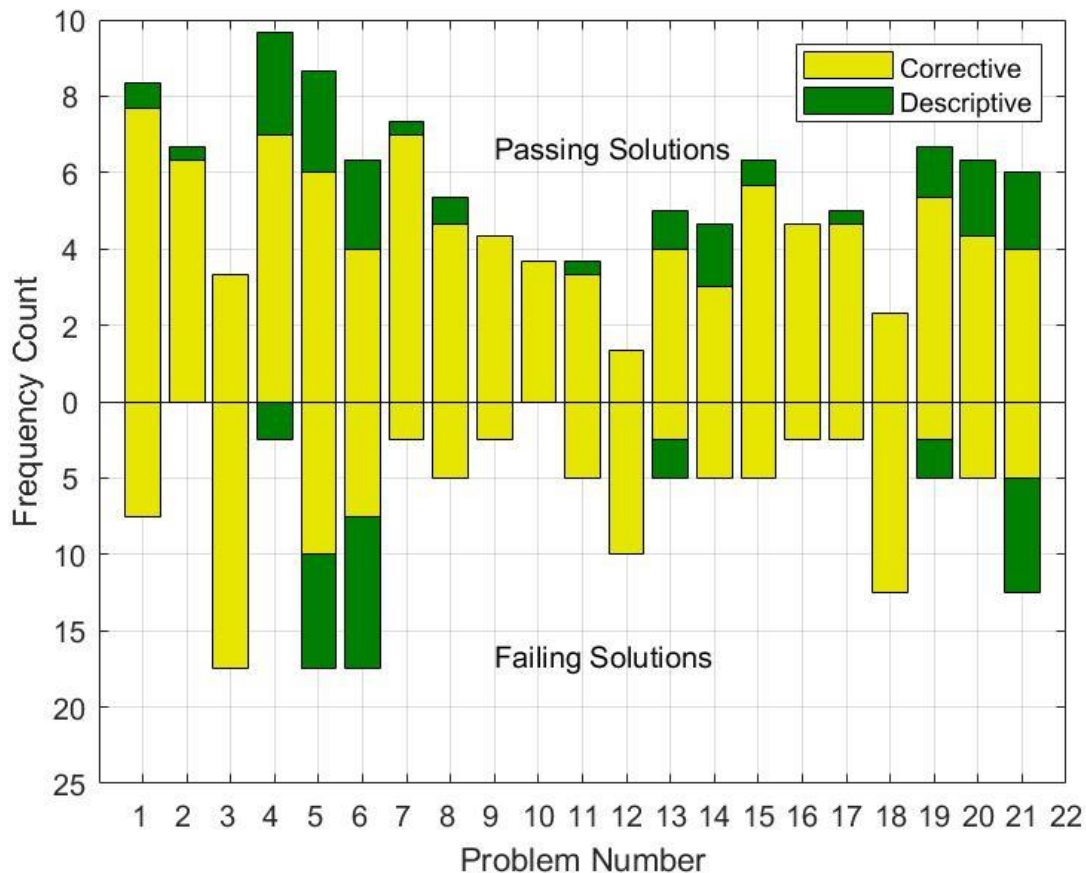
Figure 4. Frequency count of feedback categories across the 21 EE problems examined.

The average formative notations for each of the 21 EE problems are also shown in Figure 5. It can be seen that minimal feedback was provided for problems 17 and 18 in the second-year electromagnetics course (EE3). What makes the lack of formative feedback surprising is that the average summative grade for the two problems was 1/4 (25%) and 2/4 (50%) respectively, suggesting that students were not performing well and in fact would have benefited from some formative feedback. In contrast, a high degree of formative feedback was seen on problem 10 in the first-year circuits course (EE2). Such inconsistency suggests that assessment training is needed to level the assessors' feedback practices. Overall it can be seen that the mean formative feedback across the problems of the second-year electromagnetics course was consistently higher than for the two circuits courses. We can only speculate that this may be due to a smaller class size or more experienced teaching assistants. A chi-square test (CI=95%) was conducted comparing the degree of formative feedback pieces for the 21 EE problems. The results showed that the degree of corrective feedback provided is significantly higher than the descriptive feedback for all problems other than problems 17 and 18.

The frequency distribution of feedback types across the three courses is illustrated in Figure 6. The degree of elaborating feedback was noticeably lower for EE2 course which seems concerning. The degree of validating feedback was considerably higher for EE1 and EE3. The amount of validating feedback is largely attributable to the substantial quantity of this type of

feedback given to student solutions in the "excellent" category. The data in Table 1 confirms this by basically showing that the "excellent" performance category dominates the data set in that it represents 45.2% of all of the data. Chi square (CI=95%) analysis shown in Tables 2, 3, and 4 illustrates that the number of validating notations is significantly lower than the other feedback categories in the unsatisfactory performance level. In contrast, the number of validating notations is significantly higher than the other feedback categories in the marginal and good performance levels. Throughout the levels it is clear that there is relatively little elaborating feedback being given.
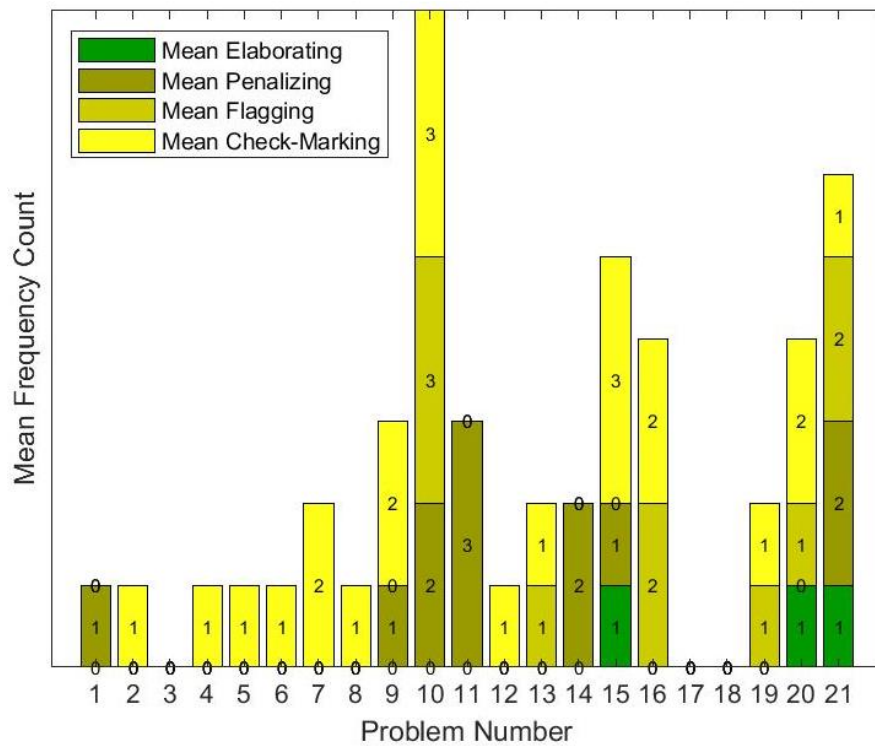


Figure 5. Average number of corrective (sum shades of yellow) and descriptive (green) feedback notations across the 21 EE problems.

| Table 2. shows which types of feedback were most significant in the "unsatisfactory" performance category. | |
|---|---|
| EE1 | Penalizing > Validating |
| EE2 | Flagging > Validating |
| EE3 | None |
| ALL | Flagging > Validating |

| Table 3. shows which types of feedback were most significant in the "marginal" performance category. | |
|---|---|
| EE1 | None |
| EE2 | Validating > Elaborating |
| EE3 | Validating > Penalizing and Elaborating |
| ALL | Validating > Elaborating |

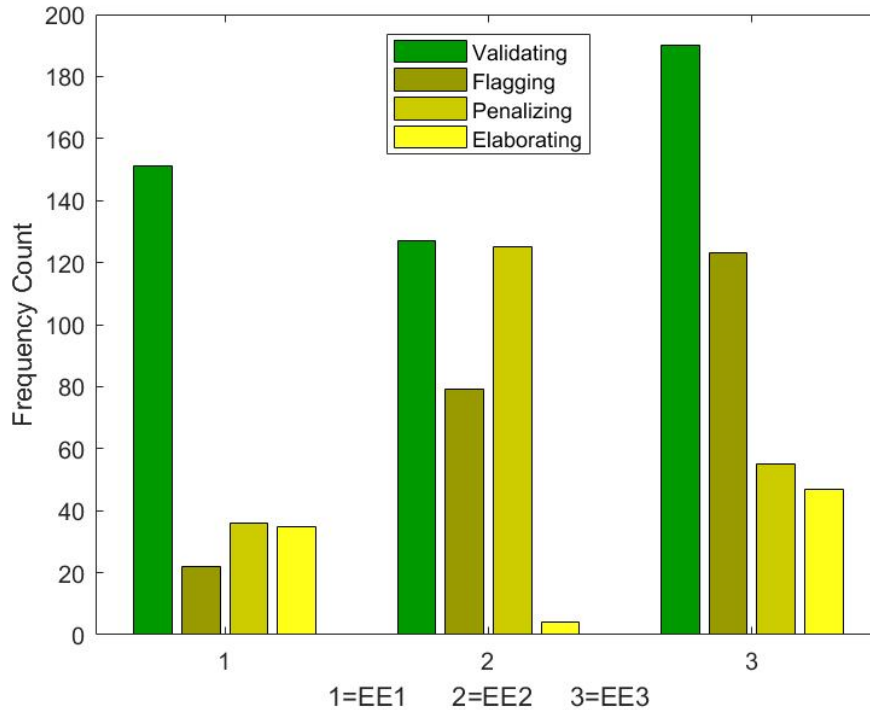| Table 4. shows which types of feedback were most significant in the "good" performance category. | |
|---|---|
| EE1 | Validating > 3 Other Categories |
| EE2 | None |
| EE3 | None |
| ALL | Validating > Flagging and Elaborating |

Figure 6. Frequency count of feedback categories by course.

## Discussion

Our findings reveal that underperforming students across three fundamental electrical engineering courses are receiving less feedback, and particularly less feedback with high formative value, as compared to their better performing peers. As the results demonstrate, conventional assessment practices are failing to provide high quantities of formative feedback especially of the flagging and elaborating type. Further it can be seen that student solutions at the lowest performance levels, on average, receive little or no feedback.

Assessment has an important role in learning, and feedback can play a key part in teaching engineering problem solving [24]. This is one of the prime principles of assessment for learning. One strategy that is used by instructors is to provide a solution key (i.e. an ideal solution), to students when returning a test. Solution keys, while helpful if the student has access to them and uses them appropriately, are not a replacement for individualized information on performance. Coming from a novice perspective, the student may not perceive the errors they have made in the same way that an expert in the field does. What appears to be a simple sign error to the student, may appear to the expert assessor as a fundamental misconception regarding an underlying physics concept.

As such, feedback can play an important role in guiding students on what concepts to review and the degree of their misunderstanding, (e.g. does it warrant taking additional steps to get assistance). Grade deductions for instance can reveal the severity of student misconceptions, or a textual feedback piece can save the underperforming student considerable time figuring out what exactly they need to work on.

The data presented here suggests that the assessment culture in engineering, at least at one major institution, emphasizes providing a summative grade, and does not actively encourage more in-depth feedback on tests and exams. This may be a result of time constraints, but it also speaks to the relative value placed on formative versus summative appraisal. The results of this study support better training for teaching assistants or other people providing feedback on student work. Encouraging the use of higher quality feedback forms, such as flagging, and suggesting that underperforming student solutions should receive the same amount of feedback would be a good first step. In addition, the results suggest that more work needs to be done to understand the factors that support assessment for learning practices, including the delivery of formative feedback.

It is important to note that equity in assessment may not necessarily mean the quality of feedback needs to be equal across different student performance levels. Adopting such a perspective may suggest that a good solution should receive the same degree and/or type of feedback as a marginal or poor student solution. This is, however, an inefficient approach, as the degree of error and learning adjustment needed are very different. It is worth considering further the development of an approach that provides equitable and efficient feedback to students that is appropriate to their performance level. This issue is related to another important consideration concerning the notion of consistency within and between courses. If all assessors in a course are providing little feedback, this may be consistent within a course. However, there can be instances where different assessors are providing different degrees of feedback within a course, or across a set of courses the student is taking. The issue of consistency between assessors was not explored in this research but is an important aspect of feedback effectiveness.

The work presented here has some significant limitations. It collected data from only 3 courses within one particular department at a large research-intensive institution. Making the results generalizable to "fact and principle" courses broadly would require a larger data set derived from a wider variety of courses and institutions. However, the teaching and assessment practices used at this institution are common across North America. In addition, we intentionally chose large fundamental freshman and sophomore courses where formative feedback is critical. In upper year courses students may be expected to have more fully developed learning skills that would allow them to make use of solution keys and other resources for self-assessment and independent learning. In introductory courses, however, we often emphasize the importance of providing support for learning, and support for developing learning skills, (i.e. provide an assessment for learning environment). The results in this study suggest that the current state of feedback practices is not fulfilling this need.

**Conclusion**

The goal of this study was to take a closer look at the types of feedback assessors provide when marking student solutions in content focused courses where closed-ended problem solving is the most common type of assessment task. What is striking from the statistical analysis is that the underperforming students consistently receive less feedback, and feedback of lower quality, as compared to their peers. The findings of this study suggest that if evaluation in engineering education intends to achieve the purpose of assessment for learning, it needs to adopt assessment

practices that more fully emphasize formative feedback and shift away from conventional practice.

## References

[1]     P. Black and D. Wiliam, "Assessment and classroom learning," *Assess. Educ. Princ. policy Pract.*, vol. 5, no. 1, pp. 7–74, 1998.

[2]     S. M. Brookhart, "Feedback that fits," *Engag. whole child Reflections best Pract. Learn. teaching, Leadersh.*, pp. 166–175, 2008.

[3]     K. E. Dunn and S. W. Mulvenon, "A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education.," *Pract. Assessment, Res. Eval.*, vol. 14, no. 7, 2009.

[4]     H. Hattie, J., & Timperley, "The power of feedback," *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81–112, 2007.

[5]     A. Kluger and A. DeNisi, "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory.," 1996.

[6]     M. Yorke, "Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice," *High. Educ.*, vol. 45, no. 477–501, 2003.

[7]     R. Butler and M. Nisan, "Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance," *J. Educ. Psychol.*, vol. 78, no. 3, p. 210, 1986.

[8]     R. Stiggins, "From formative assessment to assessment FOR learning:," *Phi Delta Kappan*, vol. 87, no. 4, pp. 324–328, 2005.

[9]     P. Black and D. Wiliam, "Inside the black box: Raising standards through classroom assessment," *Phi Delta Kappan*, vol. 80, no. 2, pp. 139–148, 1998.

[10]    D. Hounsell, V. McCune, J. Hounsell, and J. Litjens, "The quality of guidance and feedback to students," *High. Educ. Res. Dev.*, vol. 27, no. 1, pp. 55–67, 2008.

[11]    L. Forster, A. & Hewson, "'Universities learning,'" in *Staff Development in Open and Distance Learning*, K. P. C. Latchem & F. Lockwood, Ed. London, 1998, pp. 221–30.

[12]    J. Biggs, "Enhancing teaching through constructive alignment," *High. Educ.*, vol. 32, no. 3, pp. 347–364, 1996.

[13]    G. Wiggins, G. P. Wiggins, and J. McTighe, *Understanding by design*. Ascd, 2005.

[14]    J. Walther, N. W. Sochacka, and N. N. Kellam, "Quality in interpretive engineering education research: Reflections on an example study," *J. Eng. Educ.*, vol. 102, no. 4, pp. 626–659, 2013.

[15]    J. C. Archer, "State of the science in health professional education: effective feedback," *Med. Educ.*, vol. 44, no. 1, pp. 101–108, 2010.

[16]    A. Cramp, "Developing first-year engagement with written feedback," *Act. Learn. High. Educ.*, vol. 12, no. 2, pp. 113–124, 2011.

[17]    J. Biggs, "Assessment and Classroom Learning: A Role for Summative Assessment?," *Assess. Educ.*, vol. 5, no. 1, pp. 103–110, 1998.

[18]    J. B. Biggs, *Teaching for quality learning at university: What the student does.* McGraw-Hill Education (UK)., 2011.

[19]    J. Orrell, "Feedback on learning achievement: rhetoric and reality," *Teach. High. Educ.*, vol. 11, no. 441–456, 2006.

[20]    C. Evans, "Making sense of assessment feedback in higher education," *Rev. Educ. Res.*, vol. 83, no. 1, pp. 70–120, 2013.

[21]    J. W. Gikandi, D. Morrow, and N. E. Davis, "Online formative assessment in higher

education: A review of the literature," *Comput. Educ.*, vol. 57, no. 4, pp. 2333–2351, 2011.

[22]   V. J. Shute, "Focus on formative feedback," *Rev. Educ. Res.*, vol. 78, no. 1, pp. 153–189, 2008.

[23]   and B. S. B. Anderson, Lorin W., David R. Krathwohl, *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Allyn & Bacon, 2001.

[24]   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.