Mind the Gap: Insights into Student Perceptions During Peer Assessment of Writing

Andrew Tzer-Yeu Chen, Benjamin Tan, Kevin I-Kai Wang
Department of Electrical and Computer Engineering
The University of Auckland
Auckland, New Zealand
{andrew.chen, b.tan, kevin.wang}@auckland.ac.nz

***Abstract*—As educators, we need to be mindful of the gaps between student and teacher: the gaps in conceptual knowledge, the gaps in skills, and the gaps in perceptions of quality. Additionally, engineers need to be equipped with communication skills, including high-quality engineering writing. In pursuit of this goal, we reflect upon our experiences after introducing a writing activity to a large cohort of second-year engineering students. We present our findings from analysis of uncalibrated self and peer assessments of written work, alongside traditional expert-marking, to quantify the gaps in perceptions of quality.**

Keywords—assessment, engineering education, peer review, teaching and learning, writing

I. Introduction

Engineering practice is more than the judicious application of deep technical knowledge for problem-solving. Design work requires engineers to trade-off competing design metrics, and engineers should be able to strive for, and recognise, some measure of *quality*—even when such judgements may be subjective. The ability to appreciate and produce quality extends beyond the purely technical. Engineers need to be comfortable full-duplex communicators, ready to gather and synthesise knowledge for professional development, and then disseminate this to a variety of audiences. This is reflected in the expectation of “effective communication” as a Graduate Outcome for Washington Accord accreditation of engineering degrees (WA10) [1]. Writing is one of the most common forms of communication and is a critical skill for the next generation of engineers. However, despite the known importance of high-quality writing, there remain at least anecdotal complaints of poor undergraduate writing from our academic colleagues, and communication has been identified as a deficiency amongst graduates in industry [2][3]. Furthermore, the question of "when" to address writing, or "how" to add writing into courses earlier in our programmes, or even "who" is actually responsible for teaching writing skills can be of great controversy amongst faculty [4][5].

Peer review and self-reflection have also been identified to be useful for the development of “group-working skills, interpersonal skills, organisational skills, and listening skills” [6], all of which are important for engineers working in industry. It is suggested that peer-review activities help students understand how assessment is conducted, putting them in the shoes of instructors and assessors, thus allowing them to be more strategic in future assessment activities [7]. The effectiveness of these activities comes from having rich experiences to draw from, and as such, Nulty argues that these activities should be introduced as early as possible [6]. Drawing from our experience in integrating peer assessment activities with procedural technical exercises common in engineering courses (where students assess one another against model solutions provided by the instructor), we were interested to know if a similar approach could be introduced in the context of a more subjective, less regimented exercise to our second-year undergraduates. Where existing peer assessment activities in our degree programme involve student responses to technical problems and assessment against clearly defined technical criteria (as in [8]), we wondered if students were sufficiently pre-equipped to evaluate against more subjective quality criteria, despite their "novice" status.

Recognising the importance of writing, we decided to take the opportunity to introduce a research and writing-based assignment to a large cohort of undergraduates in the Department of Electrical and Computer Engineering. We integrated the assignment into a fundamental technical course, which provided a stark contrast given its subjective nature compared with traditional problem-solving assignments. As (emerging) researchers, we were especially curious to investigate if a perception gap existed between self, peer, and "expert" (i.e. instructor) assessment of student work in this softer context. One motivation was to understand better the student "angst" that arises when the final marks they receive do not match their expectations, especially for work relating to softer skills that often have to be assessed subjectively. Insight into student responses to, and engagement with, the writing task would provide us with a clearer indication of the validity and applicability of peer assessment in more subjective elements within our curriculum.

In this paper, we present the results of our study into student perceptions of their own efforts, compared with the efforts of their peers, and "expert" evaluation, within a writing assignment. Furthermore, we present and discuss our motivations and experience arising from embedding this writing activity in a traditionally technical/theoretical course. The rest of the paper is organised as follows: in Section II, we present a literature review of existing work, both in terms of engineering writing and understanding the quality of self/peer-review; in Section III, we present the context in which our writing assignment is designed; in Section IV, we describe the details of the assessment including the marking rubric; in Section V, we report on the results of our study of differences in quality perception quantitatively; in Section VI, we reflect upon our teaching practice qualitatively; and in Section VII we conclude the paper and suggest ideas for future work.

II. Literature Review

Writing has long been advocated as a useful tool for supporting the development of theoretical concept mastery [9], as well as being itself a crucial skill for providing information and persuading others in an engineering/technical context. As the IEEE Professional Communication Society notes, engineers need to be keen communicators, ready to write for several different purposes, and especially for the interface between the profession and clients (lay people) [10]. In [9], the authors advocate for writing to be distributed throughout courses in the engineering curriculum as opposed to concentrated in a single "writing intensive" course. They argue that the act of writing and a level of focus on writing should be properly integrated to support the slow process of improving writing maturity. At the same time, they acknowledge two common concerns from faculty: the administrative/time burden of running writing exercises, and potential self-doubt concerning their ability to assess such work. Considering that the ability to communicate well has been identified as a skills gap, both informally in our discussions with academic peers, and more rigorously identified in [2], we should certainly reflect more on our role as educators in fostering communication and writing skills.

Our study uses writing as a basis to explore the use of peer review for a task with subjective quality, embedded within a technically-focused course. It has been well established in the literature that peer-review activities bring numerous potential benefits for students, be it in the form of better or more timely feedback [8] or the development of student "confidence, understanding, [and] reflection" [6]. Some papers have also shown the use of self and peer assessment, and how these can be used together to encourage reflective behaviour [11]. Several studies in this domain present experiences of incorporating peer review as part of learning at different levels of experience and competency. For example, a study of secondary school students' experience in peer feedback activities is presented in [12], where the authors conclude that peer assessment activities, even when undertaken by students of relatively low levels of experience, can be useful, provided sufficient scaffolding and training. The "validity" of quantitative peer assessment is often measured by comparing the error (or numerical distance) between peer and "expert" marks. This error was found to be quite high in their study, while less so in studies undertaken at the tertiary level [6].

Boud argues in [13] that students' ability to be competent judges of quality can be improved over an extended period of time with regular practice throughout a degree programme. In [14], high correlations between peer assessment in continuous activities and final summative exam marks were observed, supporting the argument that as students develop more experience with peer assessment, their knowledge and abilities to do well in examinations also improve. Hence, it is worth identifying an approximate baseline for student perceptions of marking, particularly in terms of deciding if peer assessments should carry fractional (or even complete) weighting in a summative activity. The oft-cited Dunning-Kruger effect [15] poses some uncertainty for the validity of self (and therefore peer) assessment at lower experience levels; if a student does not know what competence is, how can they consciously know if they are competent or not? How can they assess competence?

Having said that, some studies have identified a reasonable level of validity in peer-review added to technical exercises. In [16], Cassidy presents a study that concludes that even inexperienced students have a capacity for self-assessment, although errors tended to be in the form of underestimation rather than overestimation of their performance. In [17], a study to understand the validity of peer, self, and tutor assessment commented on how combining these can help students develop more "realistic perceptions of their own abilities". In [18], Orooji and Taghiyareh present an interesting study based on multi-label voting, which identified that peer-voting tended to be positively biased (potentially because of reciprocal effects when the voters are not anonymous), but self-voting and expert-voting tended to match, suggesting that students were more honest with themselves than with each other. However, technical exercises, particularly at earlier stages of undergraduate education, tend to be accompanied by "correct" answers – for example, we would expect the validity of peer marking on a mathematics assignment to be relatively good since objective model answers can be provided to the students. Whether this is true for more subjective "soft" skills is less certain. In [19], an oral presentation was assessed through peer marking, to which the authors found no significant difference between student and lecturer assessment. In [20], peer review was used for assessing writing from students just before their capstone (final) year in a course that is focused on technical writing, measuring both the reliability and validity of the student peer-review contributions. In our work, we seek to provide some understanding of not only the validity of peer-marking in terms of comparing against the expert markers but also against self-evaluation and perception of the student's own work. We also discuss some of the other extenuating factors, such as timing and engagement of the students, and their impacts on the validity of peer assessment.

III. Background and Motivation

At The University of Auckland, within the Electrical and Computer Engineering Department, undergraduate Electrical and Electronic Engineering (EEE), Computer Systems Engineering (CSE), and Software Engineering (SE) programmes intersect in COMPSYS 201. The three disciplines take the same course under the belief that fundamental computer-related knowledge is applicable across all three cohorts. As a result, this course is one the largest common courses in the department, with approximately 220-250 students each year. We examined the class in 2017, and the course's primary technical content focused on:

> "*Digital systems and binary coding; binary numbers; Boolean algebra and computer logic; combinational logic circuits; sequential logic circuits; hardware description language; digital design flow; register transfer level descriptions and design; data paths and control units; from circuits to microprocessors; basic computer organisation; introduction to modern microprocessors; timers and interfacing; C and assembly language for microprocessors; designing digital systems using microprocessors.*"

Our idea was to introduce a writing assignment that encouraged students to find out more about how these technical topics are applied in the real-world. We planned out the assignment activity and associated study during the typical course (re)-development period before the semester commenced. Our aspirations were:

1. To introduce substantial reading, writing, and peer review exercises earlier in the three specialisations
2. To encourage student-led identification of themes, trends, and pathways where students could make use of their own individual interests, and make links between the fundamental concepts of the course and aspirational links to cutting-edge applications
3. To enhance the course with a holistic activity that linked to the University's Graduate Outcomes, specifically around the 'Scholar' Aspiration [21] (our emphases in bold):

> *"They are aware of contemporary **research** in their field of specialisation and able to **conduct their own** research and investigations. They are **excited by ideas, discovery and learning** and are conscientious in their endeavours to understand the complexities of the worlds they encounter at work and in society."*

And the Communication/Engagement Graduate Capability:

> *"Graduates of the University are expected to be able to **receive and interpret information, express ideas and share knowledge with diverse audiences** in a range of media and formats. They are expected to be able to establish a rapport and build collaborative relationships with individuals and groups."*

Given the fast-paced evolution of technology in engineering, we felt it important to help contextualise their studies, particularly in the fundamentals of computer engineering, through an activity that would help foster a broader awareness of both technological trends and current "hot" application domains. Extrinsic motivation (marks) encouraged students to engage in an assignment that might reveal cross-disciplinary challenges and opportunities. We thought that it might help students understand why they are learning what they are being taught.

In the context of the wider degree programme, students are often tasked with writing reports in Part III and IV of their studies, but do not usually have any formal access to other student writing. Additionally, in our experience, we have found that the assessment criteria for these reports in Part III and IV generally assign few marks to presentation or writing quality, instead focusing on content quality, even though that can be implicitly affected by the presentation/writing quality. Understanding that there are gradations of how *well* something is being communicated is lost in these types of assessments.

We set the assignment to take place in the earlier stages of the course. Our twelve-week semester runs as six weeks of contact time, two weeks of mid-semester break, and a further six weeks of contact before the examination period. The assignment was released alongside the initial lecture on "Trends in Digital Systems" in Week 1, and then ran in parallel with lecture content on combinational and sequential logic design. The assignment featured two components; an Executive Summary with self-review (due in Week 4), and a peer review (due in Week 6).

IV. Assessment and Research Design

Students were asked to write a short executive summary of between 400 to 500 words, summarising a technical topic with a target audience of their other classmates. The relatively short length was chosen to avoid making the assignment too arduous for students. The summary topics were restricted to five application domains relevant to computer engineering that we thought might help foster an interest in computer engineering, while also providing opportunities for the separate specialisations to relate their disciplinary backgrounds in these technologies: autonomous vehicles, artificial intelligence/machine learning, robotics, mobile computing, and virtual/augmented reality. We chose to restrict the topics to help students avoid analysis paralysis from identifying and forming their own topics. Students were instructed to include the following details in their executive summary:

- What are underlying computing trends in this domain, and why
- Recent developments in this area (both in terms of research and commercial advances)
- Some insights into the future and what might be coming next

These three items were intended to ask students to explore the past, the current, and the future state of each application domain, thus providing a wider overview of current/next generation engineering challenges. We also provided students with a short prompt of about 100 words on each topic, giving a few examples of each technology to help students know what to look for in their research and help constrain the domains.

Additionally, to help direct students towards reliable sources and avoid students getting stuck in lengthy academic research papers, instructions were given to refer to at least three recent articles from the previous three years from IEEE Spectrum, IEEE Potentials, IEEE Computer, and/or Communications of the ACM. These publications were chosen because they were highly respected industry magazines and contained more "journalism-style" articles rather than academic publications that might be too technical and difficult for second-year undergraduate students to read and understand. These publications also served as examples of the level and quality of written language expected from the students. Beyond these limitations, the structure and formatting of the writing were left up to the students.

A marking rubric, shown in Fig. 1, was released to students at the same time as the assignment instructions with four assessment criteria: identification of trends (/1.5), critical thinking (/1), writing (/1.5), and referencing (/1), summing to five marks in total. Each criterion was marked on a scale of 0-8, with the shown weighting indicating the relative importance of each criterion.

In addition to writing an executive summary, each student was required to submit a self-evaluation using the same marking rubric. After the submission deadline, each student

	Excellent (8-7)	Very Good (6-5)	Good (4)	Weak (3-2)	Poor (1-0)
Identification of trends (out of 1.5)	Explicitly draws from several references to identify recent trends and contextualises insights into future trends and challenges	Identifies and explains technology changes, the current state of the art, and comments on future trends	Identifies and explains how technology has changed	Attempts to identify a technological theme	Does not identify any technology themes or trends
Critical Thinking (out of 1)	Makes a critical evaluation of the importance of the technology, and their justified opinion on the trends, and the role of computer engineering in addressing the challenges in the technology	Reflects on the importance of the chosen computing area, and addresses the role of computer engineering in the technology trend	Clearly identifies the impact for the chosen computing area	Gives some indicator for the impact of the chosen computing area	Does not demonstrate critical thinking
Writing (out of 1.5)	Writing is very well organised, with a clear and consistent argument presented	Writing is well structured, easy to read and understand	Writing is readable and fluent, has infrequent errors	Writing is generally readable, may have fluency issues	Writing is unclear, or has frequent errors of expression
References (out of 1)	Refers to multiple relevant articles	Refers to three or more relevant articles	Refers to three recent articles	Refers to two recent articles	Does not have any references

	Excellent	Very Good	Acceptable	Poor
Quality of feedback	Provides high quality feedback with reference to the criteria, and thoughtful suggestions for improvement	Gives remarks and evaluations of several aspects of the work, with reference to the criteria	Identifies at least one aspect of the work that is positive, and one area that could be improved	Does not attempt to give feedback, or feedback is simplistic

Fig 1. The assessment rubrics, for the summary (above), and the quality of peer review feedback (below)

was assigned two other summaries to read and peer-mark. Peer marking was done anonymously (i.e. double-blind), and students undertook this activity without any pre-calibration (there was no prior opportunity for students to be coached in how to communally interpret the assessment criteria, beyond being provided with the criteria at the beginning of the assignment). The executive summary itself was worth 5% of the course grade, while the peer marking exercise was worth 2% of the course grade. The peer marks given did not have any impact on the marked student's grade - the mark for the executive summary was based solely on the expert marker's assigned grade. Two trained teaching assistants conducted all of the marking, including some overlap and moderation between the markers, in order to minimise individual marker bias effects. There is, of course, an argument to be made about whether the teaching assistants were genuine experts - both markers had significant writing experience themselves, and for the purposes of this research, we assume that they are expert enough. Expert marking was done independently of peer- and self-marks to avoid being influenced by student perceptions or expectations (i.e. tutors could not see the peer and self-marks while they did the expert marking). All marking included a numerical score based on the marking rubric along with a written comment (at least 50 words for peer marks and at least 100 words for expert marks). At the conclusion of the assessment, the teaching assistants evaluated the quality of written feedback that each student provided (using the second rubric in Fig. 1), awarding the remaining 2%. All submission and marking were done electronically through the Canvas Learning Management System (LMS).

There was a total of 223 students in the class in 2017, with a wide range of ability as indicated by their Grade Point Averages (GPAs). Our study of student perceptions focused on data collected through the assignment. As such, our research methodology was embedded in the assignment design. The appropriate human participant ethics approval was sought and received, with the caveat that students would have to opt-in to having their assessment data included in this analysis. This left 182 students in our sample, with 182 expert-assessments, 279 peer-assessments, and 181 self-assessments (some students did not complete peer- or self-assessments). The identities of students who had consented to participate in the research (or not) were kept confidential from the teaching staff until after the end of the semester to avoid any perception from the students of bias by the markers and researchers due to participation or non-participation. All marks were immediately anonymised after they were extracted out of the LMS. All students had met minimum English requirements for entering the university, although English fluency still varied.

To get a better understanding of student expectations of assessment, we formed the following research questions (RQs):

- RQ1: Do student expectations of their own marks differ from the marks given by expert markers, and if so by how much? (self vs. expert-assessment)
- RQ2: Do students mark each other's work accurately in comparison to expert markers, and if not by how much? (peer vs expert-assessment)
- RQ3: Do student expectations of their own marks differ from how they assign marks to other students, and if so by how much? (self vs. peer-assessment)

V. Data Analysis

A. Overall Assignment Grades

Before delving into the analysis of the data in terms of answering the research questions, it is useful to first understand the overall performance of the students according to the marks from the expert assessors. All data shown are the raw scores without any scaling. Table I shows that the overall average was 3.53 out of 5, approximately 75%. Fig. 2 shows the overall grade distribution of the class (out of those that opted-in to having their data included in our research), with a relatively Bell-shaped curve centred on the mean. Many students interpreted this class average as being low, relative to their expectations from first-year engineering courses. Looking more closely at the different criteria in Table I, it becomes clear that the writing score, sitting at approximately 60%, has had a large impact on reducing that class average. Referencing, which is arguably one of the more procedural elements of writing (especially when aided by modern referencing tools), had the best performance, at approximately

TABLE I. SUMMARY STATISTICS FOR THE ASSIGNMENT GRADES

Criterion	Scored Out Of	Average (mean)	Standard Deviation
Identification of Trends	1.5	1.12	0.26
Critical Thinking	1.0	0.70	0.17
Writing	1.5	0.89	0.31
Referencing	1.0	0.82	0.19
Overall Grade	**5.0**	**3.53**	**0.75**

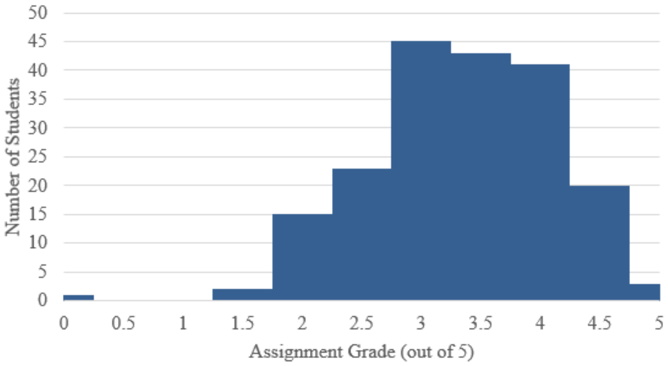

Fig 2. Overall assignment grade distribution

82%. The standard deviations show that the majority of the grades varied relatively tightly around these means, representing approximately 1-2 levels on the original 0-8 scale of the marking rubric (about 20%). We determined that this was reasonable for an assessment activity of this scale with a large number of students, so no additional scaling was applied.

B. Analysis of Assessment Mark Pairs

We answered the first part of each research question (i.e. is there a difference in perceptions) by conducting two-tailed paired t-tests between each set of assessment marks, as shown in Table II. Mark pairs where one of the marks was missing due to non-participation were removed from the dataset, hence the sample sizes in the table are slightly lower than the total collected. The results show that there was no statistically significant difference between the expert- and self-assessments, indicating that the overall marks were reasonably similar and that there was insufficient evidence that students were inaccurate in their self-assessments (assuming that we treat the expert-assessments as the ground truth). However, there were statistically significant differences between expert- and peer-assessments and between self- and peer-assessments, pointing towards meaningful inaccuracies in the peer-assessments. The overall mean for the peer-assessments tended to be higher than the expert- or self-assessments by a small amount as shown in Table II, but when we look more deeply into the differences, we find that the errors are not actually unidirectional, as shown in Table III. There were a significant number of students overestimating and underestimating the expert mark, with more students giving peer-marking grades that were lower than the expert or self-assessment grades. The mean difference in both directions is reasonably large (between 10-20% of the overall mark), indicating that the magnitude of the inaccuracies is large enough to be of concern. This result indicates that while students are generally able to apply the rubrics to their own work that they are familiar with, when it comes to applying

TABLE II. TWO-TAILED PAIRED T-TESTS FOR EACH ASSESSMENT PAIR

Pair	N	μ_1	μ_2	p-value
Expert vs. Self	180	3.56	3.61	0.487
Expert vs. Peer	275	3.54	3.88	<0.01
Self vs. Peer	273	3.59	3.88	<0.01

TABLE III. DIRECTIONAL COUNTS AND MAGNITUDE DIFFERENCES FOR EACH ASSESSMENT PAIR, WHERE M_1 IS THE FIRST MARK IN EACH PAIR

Pair	Directional Counts			Differences	
	$< M_1$	$= M_1$	$> M_1$	$\mu<M_1$	$\mu>M_1$
Expert vs. Self	91	2	87	-0.75	0.69
Expert vs. Peer	195	4	76	-0.69	0.55
Self vs. Peer	161	6	106	-0.99	0.75

marking rubrics to unfamiliar texts a lot more uncertainty is present. A secondary factor may be the amount of time that students spend on conducting peer-review; at least anecdotally, students spent many hours on writing their own executive summaries but tried to complete the peer-review as quickly as possible. We discuss this further in Section VI. This is not to say that self-assessment was perfect; there were still a small number of cases with large errors, such as a student scoring their own work as 1.58/5 when the expert marker assessed it as 4.44/5, and another where the student scored their work as 5/5 when the expert marker assessed it as 2.64/5.

To get a better understanding of how inaccurate the students were as a cohort, we visualised the absolute errors (ignoring the direction) for each assessment pair as shown in Fig. 3. The graph shows that mark pairs are generally within 20% of each other (equivalent to 1 mark out of the total 5 marks available through the marking rubric), but non-negligible proportions of mark pairs have a difference of more than 20%. In one case, the peer-mark was 46.2% lower than the expert-mark. This is a demonstration of how most students either had very different opinions to the assessors, or a lack of ability to evaluate quality "correctly". To answer each research question specifically:

- RQ1: There is insufficient evidence that student expectations of their own marks differ significantly from the expert marks that they eventually receive.
- RQ2: There is strong evidence that students conduct peer-marking in a way that is generally inaccurate, both under and over-estimating the expert-mark by about 14% on average.
- RQ3: There is strong evidence that student peer-marking yields different results to self-marking, which is coherent with the previous two statements since the self-marks are not significantly different from the expert-marks.

C. Analysis of Individual Criteria

We decided to extract further insights by looking at each criteria item separately, rather than only focusing on the overall mark. We ran two-tailed paired t-tests again for each criterion for each assessment pair, as shown in Table IV. While many of the p-values are small (indicating statistical significance), the

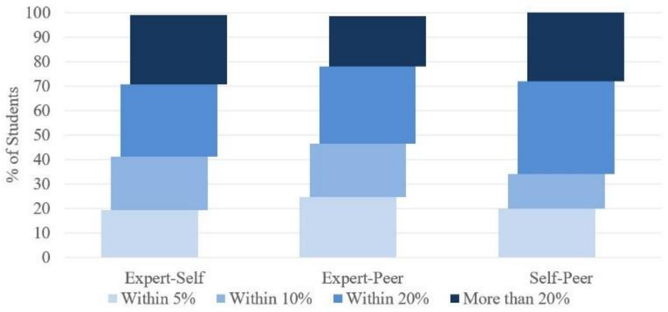

Fig 3. Proportions of students with $M_2 – M_1$ absolute error by assessment pair

TABLE IV. P-VALUES FOR TWO-TAILED PAIRED T-TESTS BY CRITERIA AND ASSESSMENT PAIR

Pair	N	Identification of Trends	Critical Thinking	Writing	Referencing
Expert vs. Self	180	0.018	0.463	<0.01	0.003
Expert vs. Peer	275	<0.01	0.032	<0.01	0.012
Self vs. Peer	273	<0.01	0.008	0.022	<0.01

writing criterion had the most statistically significant differences between the expert markers and student perceptions (in the order of 10^{-9} for expert vs. peer and 10^{-24} for expert vs. self). Critical Thinking was the only criterion that had no substantial evidence for differences in any of the pairings. For all criteria there was a reasonable balance of underestimation and overestimation of the expert-marks by students, except for in the writing criterion, where the vast majority of students overestimated (101 thought their writing was better than what the markers awarded them, 26 agreed with the markers, and 53 thought their writing was worse). The students who overestimated their writing scores did so by 2-3 levels on average (out of 9), or approximately 25%. This is reasonably strong indication of a form of unconscious incompetence, where students don't know that they are not very good at writing. This is of great concern to the instructors and is a sign that courses need to include more effort on teaching writing, including writing organisation and techniques for structuring clear, consistent arguments.

At the same time, students were very likely to hold their peers to different standards in comparison to themselves when it came to the identification of trends and referencing criteria, as evidenced by very low p-values. This could be attributable to confusion about the expected standards from the marking rubric, that the students have internal biases towards or against their own work, or that again students did not spend much time on peer assessment. The generally low p-values for the referencing criterion, which was reasonably objective in comparison to the other criteria, do suggest that the teaching staff could have been clearer about the requirements (or that students should have read the marking rubric more carefully before submission).

D. *Excellent vs Poor Students*

We also investigated whether or not accuracy in peer- or self-marking was correlated with the expert-mark, in order to determine whether students who perform better and receive higher grades are better at estimating their marks than students who receive lower grades. In this data, no statistically significant correlation relationship was found between the student's expert-mark and the absolute error between their self-mark and the expert-mark. This indicates that the student's performance is not a strong predictor of their ability to identify quality. However, we found that students who achieved a high expert-mark were more likely to underestimate their self-mark, and students who achieved a low expert-mark were more likely to overestimate their self-mark. This makes logical sense since the marks have inherent minimum and maximum values and the directionality of the error is largely dependent on how close the mark was to the minimum or maximum value.

E. *Writing Self-confidence*

At the beginning of the semester, we also asked the students to fill out a diagnostic survey that included questions about their confidence on pre-requisite content concepts. One of the questions asked about their confidence with writing; the intention was to collect information about writing confidence, independent of the requirements of the writing assignment, by asking the question before the details of the assignment were revealed to students. This was answered on a scale of 1 to 5 and acted as a very light self-reported benchmark of past writing ability/experience. A total of 164 students completed both the diagnostic survey and the executive summary. There was some weak evidence that the self-confidence ratings and self-assessment scores for writing were statistically different ($p=0.077$ based on a two-tailed paired t-test). This is somewhat interesting because it could indicate that the students' perception of their own writing ability changed between the beginning of the semester and after they completed the writing assignment. This could be explained by the hypothesis that the students' prior understanding of "technical writing" changed once they had actually completed a writing assignment that included researching and reading technical magazine articles as examples of good writing. The self-confidence scores were on average 0.165 marks higher than the final self-mark for their assignment, meaning that their perceptions of their own writing ability generally fell, indicating some form of improved awareness about their own abilities. However, this should be interpreted with caution as the statistical evidence is weak.

F. *Peer-marking Quality and Accuracy*

Finally, the peer-marking written feedback was also marked based on whether the feedback was constructive and if it made reference to the specific marking criteria. The markers were reasonably generous in terms of assessing the quality of the peer-marking, without trying too hard to determine the actual correctness of the feedback given its subjective and opinion-based nature. In general, the cases where the quality of the written feedback was assessed to be lower were due to students focusing solely on the writing quality (i.e. grammar) and not engaging with the actual content, or providing feedback that seemed unrelated to the marking criteria. In a few cases, students simply wrote that the work they reviewed was good or not and did not provide substantive feedback.

It could be hypothesised that students who were assessed to be giving higher quality written feedback would also be closer to the expert-markers in terms of their quantitative marking. We analysed this relationship as well and found no statistically significant correlation. We argue that students who understand the marking rubric and identify quality quantitatively according to that rubric should not also be assumed to be good at providing high quality written feedback qualitatively for students (and vice versa), and that these two marking tasks may require different types of skills. This may be relevant for educators and assessors in general; we are often expected to provide both quantitative scores as well as qualitative written feedback to students, which quite reasonably require different types of thinking and skills but are commonly treated as a single activity of "marking". Mackiewicz [22] demonstrates how important written feedback is in engineering contexts, and how the effectiveness of this feedback can be significantly impacted by a lack of appropriate skills.

VI. Qualitative Feedback and Discussion

In this section, we want to lightly touch on the feedback that we received from students about the assignment and reflect upon how we ran the assignment. Some of the student feedback was provided informally through office hours and Piazza (an online collaborative question/answer platform commonly used in education contexts), but also formally through the course evaluation at the end of the semester. Where students chose to tell us about the writing assignment, they generally did not like doing it. There is an inherent self-selection bias, and students who thought that the assignment was a valuable part of the course may not have felt the need to say anything.

A. Common Student Queries

Something that we noticed during the assignment was how student queries were mostly fixated on the more procedural elements of the assignment. The most common questions were around the word limit and the referencing. Our assignment stated clearly that the summary must be between 400-500 words, but there were still many questions about whether footnotes and references are included, how markers will count the words, how initialisms are counted, and whether 400-500 was a hard limit or if there was a further 10% leniency on top of that. Also, while we provided a list of recommended sources for finding supporting material in the assignment brief, we still had many questions about whether other internet sources were considered sufficiently reliable, and how many references were needed (even though this was also given in the marking rubric). Students were also confused by the term "computing trend", even after we provided an explanation in the assignment brief, in lectures, and on Piazza. This may have been an unfamiliar term to many students if they had not encountered it before, and when rephrased as "changes over time", more students understood what we were looking for.

B. Peer-marking Timing and Engagement

The timing of the peer reviews in week 6, towards the mid-semester break, had some adverse effects on student participation and engagement. Since other courses had scheduled tests and assessments around this time as well, students tended to spend as little time as possible on the peer review. It was only worth 2%, in contrast to the 10-15% assessments for other courses that naturally demanded more time commitment. Additionally, students commented that in previous (first year) courses, peer review tasks were rewarded with participation marks only and that the quality of peer review was not assessed. Some students admitted to putting little effort into the peer review exercise, with some commenting that it seemed like a box-ticking exercise and not a valuable learning experience. This would have had obvious impacts on the validity of the peer-reviews, as demonstrated in our statistical analysis.

C. Assignment and Course Feedback

The negative feedback post-assignment mostly revolved around students commenting that they did not do an engineering degree to do more writing and that they couldn't see how a writing assignment was supposed to help them. One student wrote that "the assignment felt totally pointless and I didn't take anything from it." Another wrote that "I'm not sure what actual skills it was requiring of us." This is a valuable commentary on how our intention to help students develop their professional communication skills and meet the university's Graduate Profile and the Washington Accord's Graduate Outcomes was not effectively communicated to the students. It also speaks to the lack of awareness amongst the students about the importance of these "soft" non-technical skills and how they fit into employer expectations. The ongoing challenge of understanding how to engage engineering students on the importance of technical communication and writing, and how to bridge the gap between instructor intentions to teach writing and how students perceive such efforts, is not new [23]. We learnt through this process that simply adding more writing assessment does not solve this problem - it should be accompanied with better support on helping students improve their skills through teaching, rather than merely giving students more opportunities to practice their skills.

A few students commented that the context/application focus was not coherent with the technical content of the course, in the sense that writing an executive summary about autonomous vehicles or robotics seemed to have little to do with the logic gates (and other content) they were learning about in lectures. While we intended to provide some context for how the technical content the students were learning in class could be applied to real-world applications, we hypothesise that the gap between these was simply too large.

D. Reflection

Did the assignment actually work? Did the students improve their writing skills? On those questions, we can offer no definitive answer without further study. In our conversations with students, we hypothesise that it has encouraged students to think about writing as a communication tool earlier in their education. However, in subsequent courses, we have not yet observed (at least informally) any improvement in report writing skills. As may be expected, the primary constraint on student writing ability appears to be time - when completing a six-week or semester-long design project, the associated report tends to be the last task that is completed, often a few hours before the assessment deadline. This obviously limits the quality of the writing, so we cannot determine what the baseline writing ability of the students might be. Arguably, communicating under time pressure is also an important skill, but it is one that is even harder to teach. Convincing students to take report writing seriously and to devote the appropriate time to it remains challenging.

VII. Conclusions and Future Work

In this paper, we sought to do four things. Firstly, we described the rationale behind designing a writing assignment for a second-year undergraduate engineering course, with the aim of helping students not only improve their writing skills but also better understand the broader context in which their studies sit. Secondly, we presented and discussed the idea of integrating self, peer, and expert assessment of the writing assignment to see how these reflective practices work in the context of more subjective or softer skills, rather than more traditional mathematics-based assignments. Thirdly, we presented a statistical analysis of the self, peer, and expert assessments to determine the validity of student marking, helping us understand how wide the gap is between student perceptions and assessors. We found that while self-assessment was relatively close to the expert marks and therefore valid, the peer review grades were statistically significantly different. This leads us to suggest that student familiarity with the work that they are reviewing may have an impact on how they recognise quality and apply rubrics fairly. Taking the qualitative feedback into account, simply put, students have spent a lot more time on their own work, making them more accurate at self-assessment, but spend far less time on peer assessment, making them less accurate. Lastly, we engaged with the student feedback on the assignment and reflected on our practice, including potential ways to improve the assignment in the future.

Regarding future work, this style of research has the potential for allowing academics to develop a greater understanding of how student expectations shift over time, by applying the same peer/self-review activities across multiple year levels. One of our ideas is to implement a writing assignment at each year level, with content that is relevant to the technical course that it is in, but with the same or very similar marking rubrics each time. This would allow teachers to evaluate both how the students are improving in terms of their skills and identify if the peer/self-review marks are converging towards the expert marks. However, in a university context this requires co-operation between many people in order to create a cohesive approach across year levels and courses/papers. Additionally, a similar experimental design can also be used by teachers to determine if expectations were not clear in their classes, i.e. if the self-assessment is very different to the expert-assessment, then it may be an indication that the student expectations did not match the teacher's expectations. There is also interesting research to be done towards identifying disengaged/disinterested students, and how they perceive and perform in self/peer-marking exercises.

References

[1] Engineering New Zealand. *Requirements for accreditation of engineering education programmes* [Online]. Available: https://www.engineeringnz.org/documents/122/Programme_Accreditation_Manual.pdf. [Accessed 16 June 2018].

[2] S. A. Male, M. B. Bush, and E. S. Chapman, "Perceptions of competency deficiencies in engineering graduates," *Australasian J. of Eng. Educ.*, vol. 16, no. 1, pp. 55–68, Jan. 2010.

[3] A. Keane and I. S. Gibson, "Communication trends in engineering firms: implications for undergraduate engineering courses," *Int. J. Eng. Educ.*, vol. 15, no. 2, pp. 115–121, 1999.

[4] S. Jenkins, M. K. Jordan, and P. O. Welland, "The role of writing in graduate engineering education: a survey of faculty beliefs and practices," *English for Specific Purposes*, vol. 12, no. 1, pp. 51–67, 1993.

[5] W. Zhu, "Faculty views on the importance of writing, the nature of academic writing, and teaching and responding to writing in the disciplines," *J. Second Language Writing*, vol. 13, no. 1, pp. 29–48, 2004.

[6] D. D. Nulty, "Peer and self-assessment in the first year of university," *Assessment & Evaluation in Higher Educ.*, vol. 36, no. 5, pp. 493–507, Aug. 2011.

[7] K. J. Topping, "Peer assessment," *Theory Into Practice*, vol. 48, no. 1, pp. 20–27, 2009.

[8] D. J. Boud and W. H. Holmes, "Self and peer marking in an undergraduate engineering course," *IEEE Trans. Educ.*, vol. 24, no. 4, pp. 267–274, Nov. 1981.

[9] E. Wheeler and R. L. McDonald, "Writing in engineering courses," *The Research J. Eng. Educ.*, vol. 89, no. 4, pp. 481–486, 2013.

[10] D. Kmiec and B. Longo, *The IEEE Guide to Writing in the Engineering and Techincal Fields*, Hoboken, NJ: John Wiley & Sons, 2017.

[11] F. Dochy, M. Segers, and D. Sluijsmans, "The use of self-, peer and co-assessment in higher education: a review," *Studies in Higher Educ.*, vol. 24, no. 3, pp. 331–350, 1999.

[12] O. E. Tsivitanidou, Z. C. Zacharia, and T. Hovardas, "Investigating secondary school students unmediated peer assessment skills," *Learning and Instruction*, vol. 21, no. 4, pp. 506–519, Aug. 2011.

[13] D. Boud, R. Lawson, and D. G. Thompson, "Does student engagement in self-assessment calibrate their judgement over time?," *Assessment & Evaluation in Higher Educ.*, vol. 38, no. 8, pp. 941–956, 2013.

[14] J. Sande and A. Murthy, "Including peer and self-assessment in a continuous assessment scheme in electrical and electronics engineering courses," in *Proc. IEEE Frontiers in Education Conf. (FIE)*, 2014.

[15] J. Kruger and D. Dunning, "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments," *J. Personality and Social Psychology*, vol. 77, no. 6, pp. 1121–1134, Dec. 1999.

[16] S. Cassidy, "Assessing "inexperienced" students' ability to self-assess: exploring links with learning style and academic personal control," *Assessment & Evaluation in Higher Educ.*, vol. 32, no. 3, pp. 313–330, 2007.

[17] L. A. Stefani, "Peer, self, and tutor assessment: relative reliabilities," *Studies in Higher Educ.*, vol. 19, no. 1, pp. 69–75, 1994.

[18] F. Orooji and F. Taghiyareh, "Peer assessment and self-assessment in social learning environments through a new crowd-sourced mechanism," *IEEE Access*, vol. 6, pp. 7321–7339, 2018.

[19] J. M. Meseguer-Dueñas, A. Vidaurre, J. Molina-Mateo, J. Riera, and R. M. Sala, "Validation of student peer assessment of effective oral communication in engineering degrees," *IEEE J. Latin-American Learning Technologies*, vol. 13, no. 1, pp. 11–16, Feb. 2018.

[20] P. A. Carlson, F. C. Berry, and D. Voltmer, "Incorporating student peer-review into an introduction to engineering design course," in *Proc. IEEE Frontiers in Education Conf. (FIE)*, 2005.

[21] The University of Auckland. *Graduate profiles* [Online]. Available: https://www.auckland.ac.nz/en/students/forms-policies-and-guidelines/student-policies-and-guidelines/graduate-profile.html. [Accessed 16 June 2018].

[22] J. Mackiewicz, "The effects of tutor expertise in engineering writing: a linguistic analysis of writing tutors' comments," *IEEE Trans. Prof. Commun.*, vol. 47, no. 4, pp. 316–328, 2004.

[23] J. Nelson, "This was an easy assignment: examining how students interpret academic writing tasks," *Research in the Teaching of English*, vol. 24, no. 4, pp. 362–396, 1990.