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

Andrew Tzer-Yeu Chen  Benjamin Tan  Kevin I-Kai Wang
Overview

• Introduction
• Research questions
• The Writing Task
• Key Insights
• Conclusions and future work
Introduction

• Perception is **everything**
  ➢ Student response pre-, during, and post- activity

• What do (novice) students **think**?
  ➢ Technical stuff is easier (right/wrong)
  ➢ “Soft” stuff is more difficult

• Effective communication is critical in engineering
  ➢ Graduate outcomes: “**excited by ideas, discovery and learning**”
  ➢ Graduate capability: “**receive and interpret information, express ideas and share knowledge with diverse audiences**”

• Future engineers need to be able to deal with nuances, subjective requirements
Research Questions

• The idea (in brief):
  ➢ Make students write earlier in the degree, assess themselves and each other – then see what happens

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)
Aims

• Gain insights into student perceptions and responses to a “softer” context
• Explore a potential model for running similar ‘low cost’ writing activities
  ➢ i.e., if valid, then a workable design for other courses

Not enough time to ‘teach’ writing...

...but student writing is getting worse...
The Writing Task

• COMPSYS 201
  ➢ Fundamentals of Computer Engineering
  ➢ First year of specialisation (2\textsuperscript{nd} year students)
  ➢ 220-250 students each year
  ➢ Intersection of EEE, CSE, and SE programmes

• We set a Writing Task:
  ➢ 400 to 500 word summary of a technology
  ➢ Autonomous vehicles, AI/ML, Robotics, Mobile Computing, VR/AR
The Writing Task

• Three questions for focus:
  ➢ What are the underlying computing trends?
  ➢ Recent developments
  ➢ Insights into the future

• At least 3 recent articles from:
  ➢ IEEE Spectrum, Potentials, Computer, and/or Communications of the ACM
  ➢ “Journalism-style”

• Encourage students to find links between technical topics and the real world
The Writing Task

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

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

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

- Self-assessment and double-blind peer assessment
- Expert marking (i.e., by teaching assistants)
- 5% for the summary, 2% for contributing to peer assessment
Key Insights

- We examined the data from:
  - 182 expert-assessments
  - 279 peer-assessments
  - 181 self-assessments

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.
Key Insights

• Peers hold others to different standards in comparison to themselves
• No statistically significant correlation between student’s mark and the absolute error between self-mark and expert-mark
• Weak evidence that student self-confidence fell after the assignment
• No link between quality of student qualitative feedback and their own achievement
Key Insights

• Some students fixated on procedural elements
  ➢ Word counts, number of references, “leniency”
• Some students did not recognise any learning value in undertaking peer review, nor in providing quality peer review comments for others
• Some students didn’t like the assignment
  ➢ “did not do an engineering degree to do more writing”
  ➢ “I’m not sure what actual skills it was requiring of us”
Conclusions and Future Work

• Investigated the gap between peer, self, and expert assessment of writing
  ➢ Self and Expert gap is okay, but other gaps not so much

• Students spend a lot more time on their own work, and much less time on the work of others

• In future:
  ➢ Investigate the shift in the gaps over time
  ➢ Does writing actually improve?
  ➢ Can we find students that are disengaged?
Supplementary Slides
Data analysis and reflections

- 223 students enrolled
  - ethics approval received for opt-in
  - 182 expert-assessments
  - 279 peer-assessments
  - 181 self-assessments
- Class average of ~75%

**TABLE I. SUMMARY STATISTICS FOR THE ASSIGNMENT GRADES**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Scored Out Of</th>
<th>Average (mean)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of Trends</td>
<td>1.5</td>
<td>1.12</td>
<td>0.26</td>
</tr>
<tr>
<td>Critical Thinking</td>
<td>1.0</td>
<td>0.70</td>
<td>0.17</td>
</tr>
<tr>
<td>Writing</td>
<td>1.5</td>
<td>0.89</td>
<td>0.31</td>
</tr>
<tr>
<td>Referencing</td>
<td>1.0</td>
<td>0.82</td>
<td>0.19</td>
</tr>
<tr>
<td>Overall Grade</td>
<td>5.0</td>
<td>3.53</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Fig 2. Overall assignment grade distribution
Data analysis and reflections

Two-tailed t-tests between pairs of assessment marks

- Expert vs. self – no statistically significant difference
- Expert vs. peer / self vs. peer – statistically significant differences
- Inaccuracies in both directions