Department of Electrical and Computer Engineering

University of Auckland, New Zealand

TALE 2018

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 (2nd 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

	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)

- 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

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**

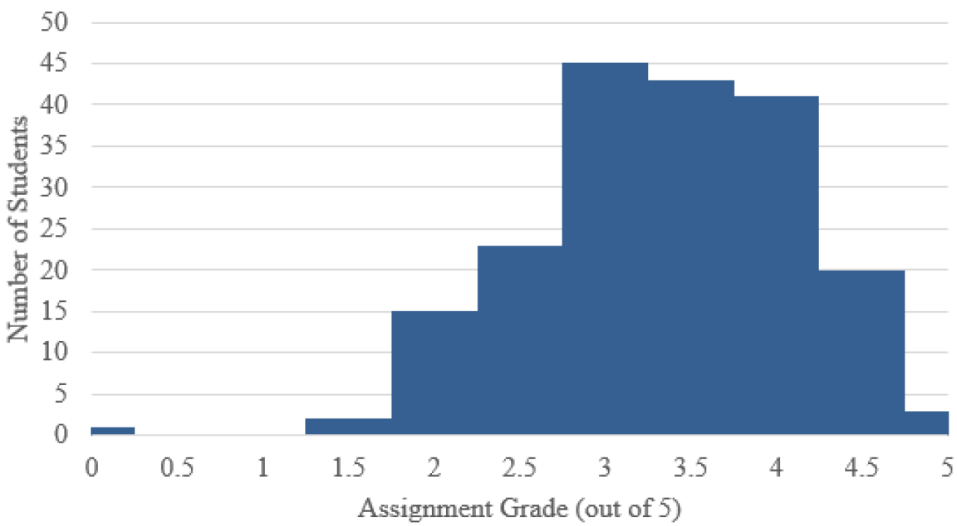

Fig 2. Overall assignment grade distribution

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

Data analysis and reflections

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

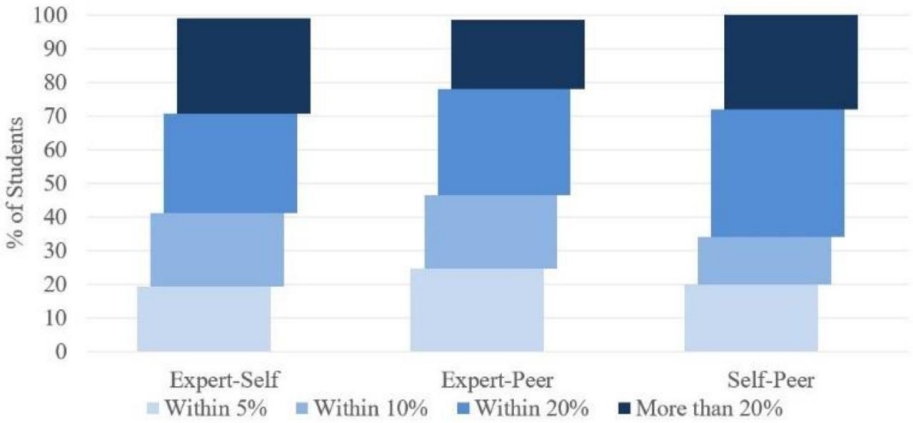

Fig 3. Proportions of students with $M_2 - M_1$ absolute error by assessment pair

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

- 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