33rd Annual Conference of the Australasian Association for Engineering Education (AAEE 2022)
4 - 7 December 2022 – Sydney, NSW

Scalable Vivas: Industry Authentic Assessment to Large Cohorts through "Debate Club" Tutorials

Ashleigh Kirkland[a], Cornelius Paardekooper[a], Joshua Flynn[a], Dylan Cuskelly[a], Elena Prieto-Rodriguez[a], William Mcbride[a], Alexander Gregg[a].
University of Newcastle[a]
Corresponding Author Email: Ashleigh.Kirkland@uon.edu.au

ABSTRACT

CONTEXT

In the context of engineering education, Industry Authentic Assessment (IAA) – where students are assessed in a manner that replicates the professional environment – has gained traction and been the subject of much study in recent years (Wiggins, 1990; Wellington, Thomas, Powell, & Clarke, 2002; Soares, 2013; Andrew L Guzzomi, 2017; Sotiriadou, 2020; Ullah, 2020). This discussion is particularly pertinent in the current climate, with the Federal Government now providing strong incentives for universities to produce 'job ready' graduates (Department of Education, Skills and Employment, 2020). The Viva Voce (oral assessment) has long been viewed as an authentic assessment technique in engineering, also favoured for its inherent academic integrity and capability for deeper, unscripted discussion (Sotiriadou, 2020). This technique also indirectly addresses a range of EA Stage 1 Competencies around communication and professionalism – 3.2, 3.4 and 3.5. (Engineers Australia, 2012)

PURPOSE OR GOAL

One-on-one staff/student vivas do not scale well to large cohorts or to frequent/micro assessment. In this paper, we critically analyse and detail our experiences utilising peer marking as a means to achieve Viva Voce micro assessment and feedback for a large (300+) cohort through a 'debate club' style tutorial. In this paper, we evaluate whether;

1. This was possible, practical and economical.
2. Students found this class style engaging.
3. The style aided improvements to communication skills as per the EA competencies.

APPROACH OR METHODOLOGY/METHODS

We address our research questions through analysis of qualitative student and instructor feedback. This is supported with attendance/engagement/grade analytics.

ACTUAL OR ANTICIPATED OUTCOMES

Peer marking provided a means for practical, economical, and scalable vivas, with weekly presentations by 300+ students managed and graded entirely contemporaneously. Students found the class/assessment style highly engaging, and a notable improvement in stage 1 communication competencies was observed throughout the semester.

CONCLUSIONS/RECOMMENDATIONS/SUMMARY

Through this methodology the Viva Voce has proven to feasible and economical at scale. The culmination with industry focus has created an engaging and enjoyable assessment as noted by students, with the benefits of encouraging growth in communication skills and professionalism.

KEYWORDS

Industry authentic assessment; Viva Voce; Debate;

Introduction

Over the past decade, the completion rate among tertiary engineering programs has been slowly increasing for both domestic and international students (Engineers Australia, 2020). At the same time, institutions face an increased and government-incentivized demand to provide graduates who are 'job ready' (Department of Education, Skills and Employment, 2020) - tailoring courses and programs to provide an authentic training environment representative of the workplace. These, taken together, pose a significant challenge for engineering educators – who must provide industry authentic experiences to large cohorts while maintaining academic integrity throughout assessment.

One well-established and industry-authentic assessment technique (IAA) is the Viva-Voce (oral examination), where students demonstrate their understanding in a discussion with an assessor. This is highly authentic to the engineering industry – where design meetings/pitches/etc. are an everyday practice. This technique also ensures academic integrity and provides students an opportunity to demonstrate understanding through deeper discussion outside the rigid constraints of e.g. a written/multiple-choice/short answer assessment (Sotiriadou, 2020).

Classical Viva Voce assessment is time-intensive – often one-on-one with a student and assessor. In this paper, we discuss our experience implementing a Viva Voce IAA to a large cohort through weekly, micro-assessed 'debate style' tutorial exercises. This approach was implemented in a first-year engineering materials design course at a regional Australian university. We discuss the feasibility/practicality of this approach and analyse instructor and student feedback on this implementation, particularly pertaining to engagement with the tutorial classes and assessments, and improvement of key EA competencies around communication skills and professionalism.

Background

MECH1750 is a new engineering materials design course at the University of Newcastle in Australia. This course provides fundamental material science background alongside manufacturing and material design/selection content. With 300+ students across a wide range of disciplines (i.e. Mechanical, Mechatronics, Aerospace, Medical), the course leverages a flipped classroom model. Content is delivered through pre-recorded lecture videos, and these are supported by engaging weekly tutorial classes with embedded assessment.

In these classes, students work in teams to develop and 'pitch' solutions to engineering materials design problems (e.g. "What material/manufacturing process is best for a Frying Pan?"). These problems are intentionally broad/vague to allow for creativity in both scope and solution.

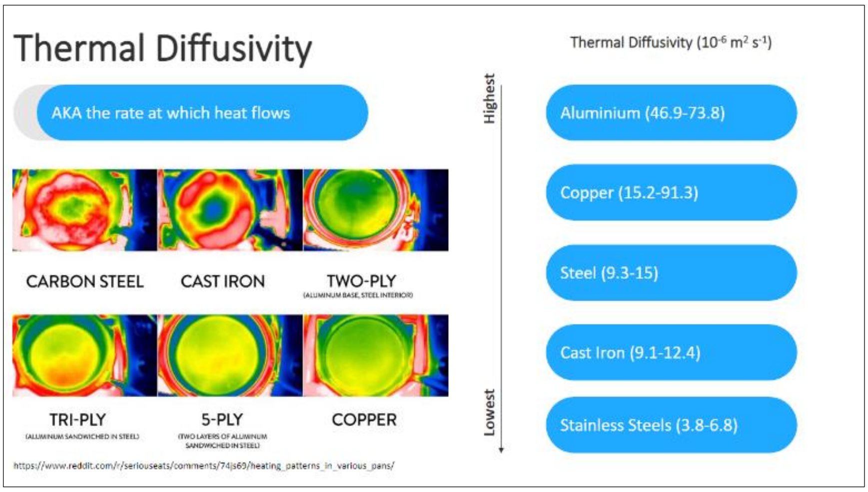

Figure 1: One slide from a student presentation. Students were encouraged to scope their own problem – identifying which material properties were important and ranking solutions accordingly.

Throughout the class, student teams were paired up and a representative from each presented a 4-minute pitch on their teams' behalf. After both pitches have been presented, teams participate in a discussion on their solution, providing an opportunity for rebuttal, validation, or deeper investigation of the topic. This sequence of activities the authors term a debate, which is acknowledged that this does not conform to the traditional notion of a debate.

Spectating members from both teams then provided peer feedback on these debates, rating on a four-point Likert scale:

- Presentation Skill,
- Technical Understanding,
- Professionalism,
- and performance in a free-form discussion

Additionally, each spectator was required to provide a short comment on each presentation, noting both an outstanding feature of the pitch, and a possible area of improvement.

An example debate was presented in the first session, to benchmark the Likert scale and teach students how to provide helpful comments.

Tutorials consisted of each team participated in 4 debates, with their opponents being shuffled each time. Each member of the team was required to present at least once.

Grades were awarded in part for satisfactory engagement with each aspect of the process:

- Presenting in at least one debate,
- Giving constructive feedback as a spectator,
- Actioning their peer provided feedback as being "helpful"," neutral" or "unhelpful",
- and in part as a function of their peer-assessed performance in the debate.

With (typically) 140 students/35 groups/17 simultaneous debates in each class and three staff demonstrators, only a fraction of debates were instructor-observed each week. Peer feedback across the entire class was moderated/scaled to align the average peer-provided grades with instructor-provided grades in those debates that were observed. Over the course of the semester, each group was observed several times.

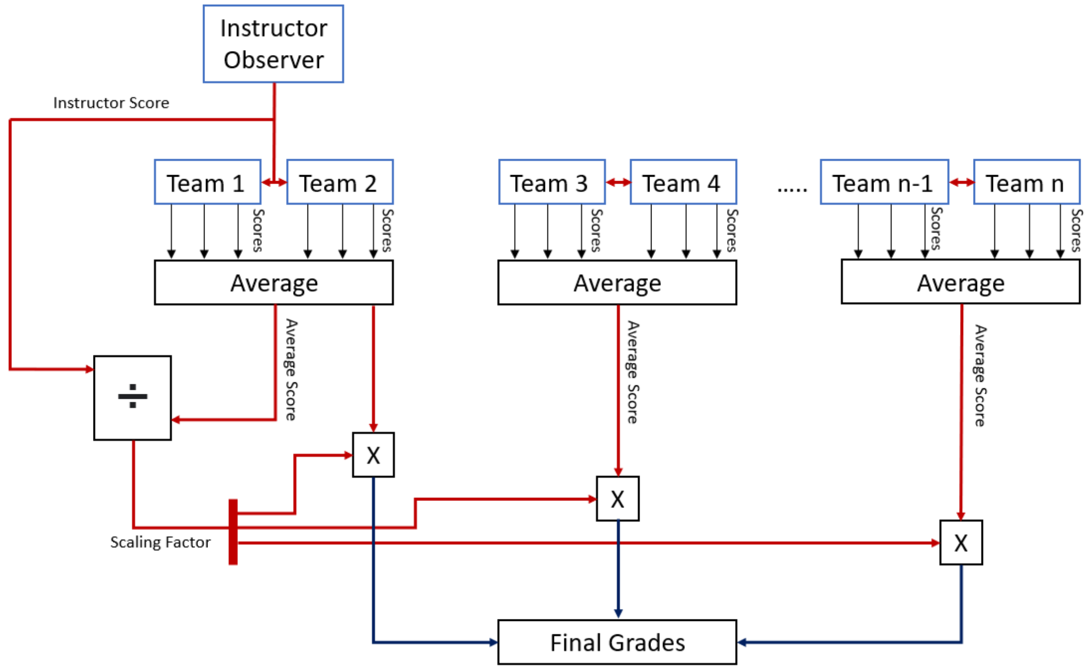

Figure 2: Flow chart of grading/scaling process. One/several instructor-observed debates informs a scaling factor that moderates peer-graded debates.

Scores are moderated based on scaling which is applied to every student in the session. The scaling factor is calculated via the differences between the instructor and student given performance scores. The performance score is a normalised sum of the feedback given on the Likert scale (1 being very bad and 4 being very good), which is then compared to the instructor given performance score for the students observed. The differences between these scores are averaged to create the scaling factor for that session.

If the scaling factor is greater than 1 – meaning that the staff rated the students higher than their peers, every student's peer performance mark gets scaled by this factor to provide the performance portion of the grade for that tutorial. If the scaling factor is less than 1, the marks remain as per the student's average.

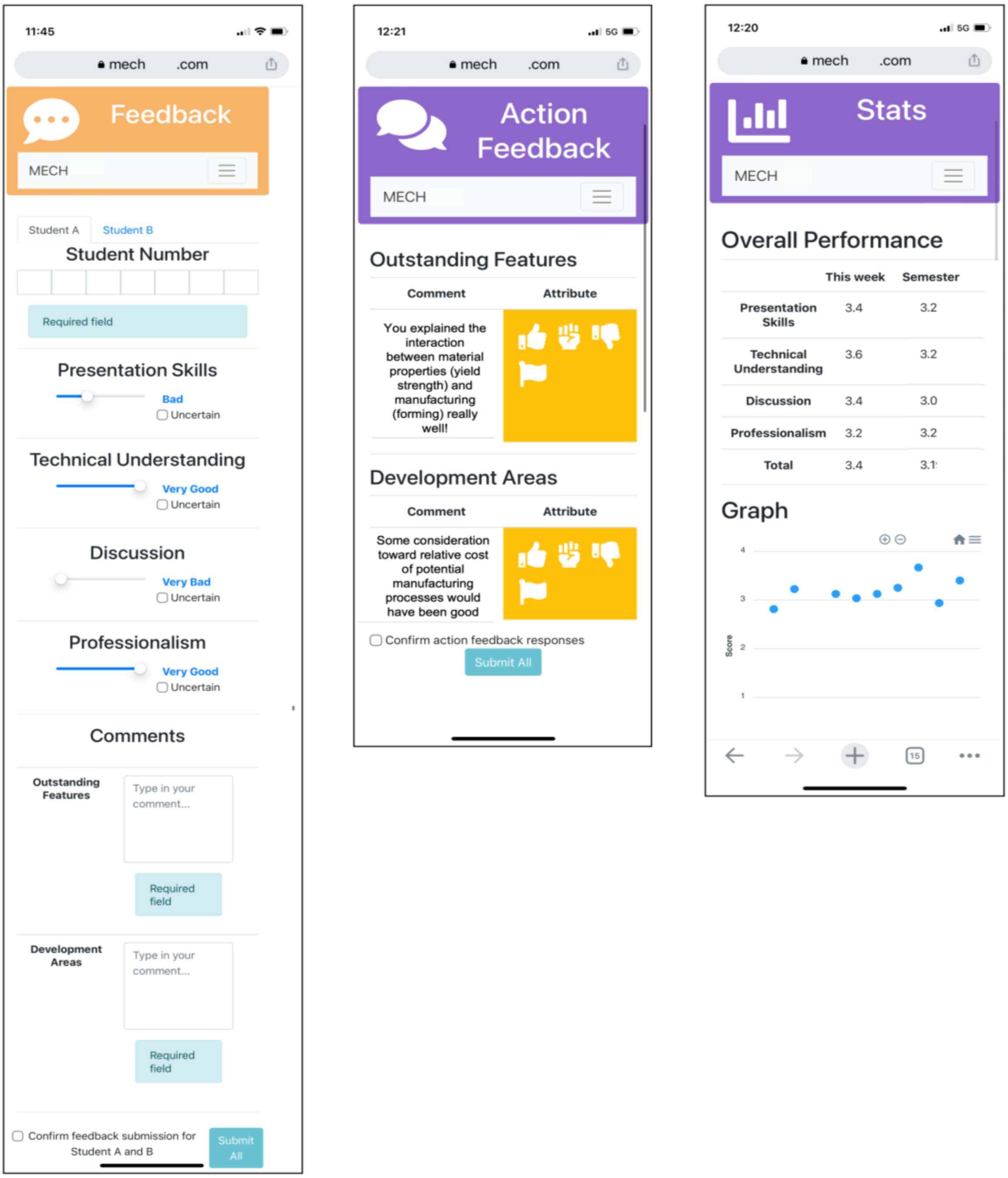

Figure 3: Excerpts from the student feedback system. Left: the feedback input/submission form. Centre: the feedback output/viewing/review page. Right: historical performance/self-review page.

The actioning of the peer provided feedback allowed for unhelpful feedback to be identified. This led to more active moderation of problem individuals/groups by ensuring an instructor observed more of their debates.

Students were provided access to an online dashboard (Paardekooper, Flynn, Kirkland, Cuskelly, & Gregg, 2022), where they could check their most recent and historical debate performance, as well as review the outstanding feature/area of improvement comments from their peers. These comments were anonymous to the students, and a thumbs-up/thumbs-down/flag comment review functionality was provided to disincentivise unprofessional feedback and allow instructors to moderate/track inappropriate comments.

Over the course of the semester, approximately 2000 debate sessions took place, with each student nominally providing one presentation each week. By the end of semester, more than 12,000 individual pieces of feedback were provided.

Methodology

To answer our research questions, we collected and analysed four data sets:

1. Responses to the standard, optional, end-of-course experience questionnaire (CES) – namely those pertaining to the tutorial assessments within ‘best aspects’/’needs improvement’ questions.
2. A separate, voluntary survey centered on the tutorial assessments.
3. A series of follow-up semi-structured interviews with survey respondents.
4. A series of informal interviews/debriefs with teaching staff.

Thematic analysis was undertaken on comments from both surveys and interview transcripts to uncover patterns within the data. An inductive approach was used to drive this process (Braun & Clarke, 2006), with comments first coded in-line with the research questions (i.e. “Classes were Engaging/Unengaging”, “Classes helped/did not help improve EA Competencies”, “Classes helped/did not help me learn") and sub-themes then identified from these. Two researchers collaborated in this coding process and debated alignment of comments until a consensus was reached for each case.

Results and Discussion

Thematic Map

An overwhelmingly positive sentiment was clear across the survey responses and interview data, with three strong themes emerging. These, and their sub-themes, are shown in Figure 4.

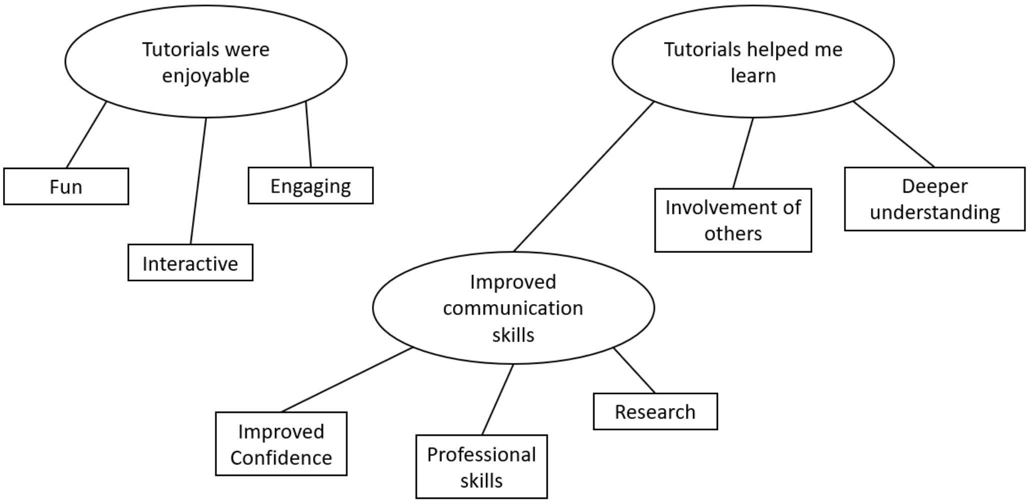

Figure 4: Final Thematic map.

Assessment Feasibility

Student and staff perceptions around this class/assessment were positive, with the overwhelming sentiment being that this represented a feasible approach to 'scalable vivas'. Each week, 250 students on-average presented in a debate and provided/received peer feedback.

This peer-to-peer interaction was essential to success of this approach. Through an automated web system that managed the submission, collation, moderation, scaling and provision of feedback (Paardekooper, Flynn, Kirkland, Cuskelly, & Gregg, 2022), marking workload was essentially limited to contemporaneous observation of a subset of debates, and ~1 hour of weekly review/moderation of a small number of flagged comments.

The web system developed allowed students to both provide feedback in real time and later review the feedback they received as a function of their debate performance. It also displayed historical debate performance, allowing students to track their improvement over the course of the semester. This system was central to the feasibility of the class – manual tracking/submission/return of feedback on this scale would not have been feasible.

This approach resulted in an extremely efficient casual teaching budget – with 18 hours of demonstration per week (three demonstrators over three, two-hour classes) and no additional marking required for this combined tutorial/ 30% assessment in a 300+ student course. This class type also scales well with increased resources as additional demonstrator hours will simultaneously improve marking statistics and result in more student facetime.

The assessment style was well received with students, who noted that sessions were run effectively and were a valuable use of their time:

> *"The weekly debates were a new format that I hadn't experienced before but overall I really enjoyed them **and they always made me learn a lot more about that weeks content that I otherwise would have**."*

> *"The debates every week made the course so engaging and **made learning the content so much easier and more efficient**"*

In general, students found the workload associated with preparing and delivering a viva-voce style debate presentation each week to be manageable, and in-line with the 3% per week assessment weighting. Students indicated that they spent 1-2 hours per week on average preparing for the tutorial debates, and that this preparation was good study for the other assessments.

Some concerns around workload and equity of assessment – particularly in terms of timetabling - emerged from the student feedback. Classes scheduled early in the week had less time to consume the lecture content and prepare a solution for the tutorial. While this was managed by adjusting instructor expectations for these 'early' classes, it should be noted that the classes scheduled later did not outperform earlier classes. However, it is still recommended that efforts are made to ensure sufficient time between content release and assessment.

Throughout the semester instructors noted clear improvements in the quality of the debates presented. Additionally, it was noted that students tended to recalibrate themselves as to what constituted placement in each point on the Likert scale. More data is needed to quantify this notable improvement.

Engagement

All data suggests that students found the debate-style tutorials to be fun, engaging learning experiences. On average, attendance to weekly tutorials exceeded 78%, in spite of challenges posed by online delivery/COVID.

Students noted enjoying all aspects of the process – development of a solution with a small team of engineers, presenting this solution as a speaker, watching and learning from the presentations of others (both within and outside their team), and giving/receiving helpful peer feedback. Both student and instructor feedback suggested that building a supportive culture among the cohort reduced anxiety/apprehension around public speaking and strongly contributed to student enthusiasm for the class.

> *"The tuts [were the best aspect of the course]! They were so much fun to do each week after our team started flowing. It was great to not only make our own solutions but to see and give feedback to others. You guys did such a good job mustering everyone into such a positive community."*

> *"I also really enjoyed the debating aspect where you could see how other teams think and the other ideas that I had completely glossed over."*

> *"Honestly weekly group based public speaking debates sounds on paper like the worst course possible, but it turned out to be one of my favourites."*

Improvement of Skills

Students noted the value of this assessment in improving both their technical (materials) knowledge and communication skills:

> *"The debates were amazing at developing my understanding of the course material as well as the bonus experience of improving presentation skills and 'close-to-real-world' experience…"*

These classes were designed to improve key Engineers Australia stage 1 competencies around professionalism, namely 3.2 - *Effective oral and written communication in professional and lay domains*, and 3.5 - *Orderly management of self, and professional conduct*.

Both teaching staff and the student cohort observed an appreciable improvement in these areas throughout the semester. Students not only grew more comfortable/confident as presenters, but as more thoughtful and reflective engineers – recognising the nuance of real-world engineering design problems and better communicating these complexities and the challenges associated with solving these in their presentations.

“As the Semester progressed my presentation skills increased dramatically, as did my skills in the discussion after the presentations. The consistent practicing of these skills throughout the Semester definitely played a role in my improvements.”

While this was explicitly encouraged through minor weekly interventions by teaching staff at the beginning of each class (e.g. “this week we’d like you to focus on avoiding strawman arguments”), the data suggests that the predominant driver of improvement was peer-peer interaction – peer observation of others debates and feedback provided by others each week.

This improvement could also be seen in a net increase in average scores for instructor-observed debates throughout the semester – from 2.96/4 in week 5 to 3.41/4 in week 10. Interestingly this trend was not as strongly visible in raw peer scores. Students it seemed raised their expectations throughout the semester as they themselves improved.

Students also noted that the process of providing peer feedback encouraged reflection around their own presentations/solutions and motivated self-improvement – primarily through adoption of positive aspects/practices from other debates, but also in that listening to other perspectives on the problem was good revision of the course content.

“The tutorials are a great way to get people engaging with the content and also help to develop professional skills in the form of researching appropriate materials and presenting to other people. They are also a great tool to expose everyone to different ideas and frames of thought or ways to tackle a problem that an individual or group may not have considered.”

Other Themes

In addition to key themes already previously identified, there were a handful of smaller tropes that emerged. Chief among these was industry authenticity. Students both acknowledged and appreciated the relevancy of the assessment task to industry, noting benefits to their confidence surrounding job preparedness.

“The consideration of industry within the tutorials was excellent with really great choice of tutorial problems which struck a balance between difficult and rewarding… “

“[best aspect of the course:] Real emphasis on real-world experience and applications”

Conclusion

Materials Science/Engineering is typically taught in a very traditional fashion. The design of ‘debate-club’ viva-voce tutorial assessments for MECH1750 represented a significant departure from this. Nevertheless, student and instructor feedback suggests that this IAA was not only feasible/economical, but an engaging useful learning experience. Students improved not only their technical understanding but also their communication skills.

Further work is required to probe the integrity of peer grading – while a simple instructor/student moderation/calibration was applied here, there is potential for further analysis/comparison/validation as more data is collected. The contribution of these classes to improved learning should also be quantified, and future work may entail cross-correlation of debate performance with that in other assessment. Finally, direct consultation with industry is planned to validate that an IAA has actually been achieved.

In conclusion: Viva Voce assessment is not only possible but can be economical at scale; students find it engaging and enjoyable, and it can help improve communication skills and professionalism.

References

Andrew L Guzzomi, S. A. (2017). Students' responses to authentic assessment designed to develop commitment to performing at their best. *European Journal of Engineering Education, 42*(3), 219-240.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 77-101.

Department of Education, Skills and Employment. (2020). *Job-ready Graduates.* Australian Government.

Engineers Australia. (2012). *Stage 1 Competency Standard for Professional Engineer.*

Engineers Australia. (2020). *Australia's Next Generation of Engineers.* Engineers Australia.

Fun Man Fung, S. B. (2021). Adopting a Gamified Approach of Conducting Viva Voce in an Undergraduate Lab Module. *Journal of Chemical Education*, 98-100.

Jon Pearce, R. M. (2009). *Involving students in peer review.* Melbourne: Centre for the Study of Higher Education, University of Melbourne.

Paardekooper, C., Flynn, J., Kirkland, A., Cuskelly, D., & Gregg, A. (2022). Facilitating Scalable Vivas through Purpose-Built Software. *Australasian Association for Engineering Education.* Sydney, NSW.

Sharmila Torke, R. R. (2010). The impact of viva-voce examination on students' performance in theory component of the final summative examination in physiology. *Journal of Physiology and Pathophysiology Vol. 1*, 10-12.

Soares, S. C. (2013). Authentic assessment in Software Engineering education based on PBL principles a case study in the telecom market. *2013 35th International Conference on Software Engineering (ICSE)*, 1055-1062.

Sotiriadou, P. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, 1-2.

Ullah, S. N. (2020). Examples of Authentic Assessments in Engineering Education. *2020 IEEE Global Engineering Education Conference (EDUCON)*, (pp. 894-897).

Wellington, P., Thomas, I., Powell, I., & Clarke, B. (2002). Authentic Assessment Applied to Engineering and Business Undergraduate Consulting Teams. *International Journal of Engineering Education*, 168-179.

Western Sydney University. (2020, October). *Designing and Assessing Vivas.* Retrieved from Designing and Assessing Vivas: https://www.westernsydney.edu.au/__data/assets/pdf_file/0011/1757837/Designing_and_Assessing_Vivas.pdf

Wiggins, G. (1990). The Case for Authentic Assessment. *Practical Assessment, Research, and Evaluation: Vol. 2*, Practical Assessment, Research, and Evaluation: Vol. 2.

Copyright statement

Research papers are invited for AAEE 2022. The review criteria for each category are presented in the Call for Papers and Workshops.