ASCILITE 2024

Navigating the Terrain:
Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies

Leveraging Oral Assessments to Enhance Learning and Integrity in Teaching Data Analytics

Hasti Samadi, Geela Venise Fabic Chee
The University of Melbourne

The rapid advancement of generative AI tools, like ChatGPT, has significantly impacted academic integrity in higher education. This paper explores the integration of oral assessments with traditional project evaluation in data analytics courses to address these challenges. While oral assessments cannot completely prevent cheating, they enable examiners to probe students' understanding more deeply. We present an overview of our assessment design and processes rather than detailed student results. We compare the implementation of oral assessments in a fully online professional course and a face-to-face undergraduate course. Moreover, we compare results before and after oral assessment training and explore AI's role in efficiently generating individualised questions. Our findings demonstrate that oral assessments reduce academic dishonesty, enhance comprehension, and increase assessment rigour.

Keywords: oral assessment, data analytics, academic integrity, generative AI, higher education, professional learners, undergraduate students

Introduction

Assessment is pivotal in shaping both what and how students learn in any educational setting. Students' approaches to learning are often influenced by their expectations of how they will be assessed (Sambell et al., 2017). In recent years, the integrity of assessments has become a growing concern in higher education, particularly in data analytics education (Tu et al., 2023). The introduction of generative AI tools, such as OpenAI's ChatGPT, has aggravated these issues. These tools can automate complex tasks such as data cleaning, model building, interpretation, and report writing (Tu et al., 2023). Students may misuse these tools to generate content that appears original, making it difficult for educators to ensure that the submitted work truly reflects the student's understanding and effort.

Research has shown that oral assessments can strengthen academic integrity, increase inclusivity, and enhance higher-order thinking skills (Huxham et al., 2012; Mahendra, 2023). However, concerns about the validity and fairness of oral assessments persist, particularly regarding examiner biases based on age, gender, race, or socioeconomic status (Gardner & Giordano, 2023). Addressing these biases is crucial, especially for non-native English speakers. Additionally, fear of public speaking should be considered (Grieve et al., 2021). On the other hand, when implemented correctly, oral assessments can also improve fairness in the assessment, as neurologically diverse students, such as those with autism or dyslexia, often perform better in oral formats (Accardo et al., 2020).

In response to the challenges of maintaining assessment integrity, we introduced an oral assessment component alongside traditional project-based assessments in data analytics courses. This new component mitigates the potential misuse of generative AI by requiring students to explain their work verbally in a supervised setting. Students present their projects and answer questions in an interview-style format, ensuring a deeper engagement with the material. This paper outlines the pilot implementation of this new assessment design in two subjects at the University of Melbourne: Data Analytics with Python and Elements of Data Processing. We also introduce a trial to streamline question creation for a larger cohort using a university-implemented web version of ChatGPT-3.5 (hosted locally by the university) called Spark AI. To ensure that privacy and ethical considerations are prioritized, we exclusively utilized Spark AI, a university-approved platform designed with robust data security and ethical safeguards. Following consultations with the university's Human Ethics Committee, we confirmed that our approach falls under quality assurance evaluation rather than human research, and therefore does not require formal ethics approval. In line with

quality assurance practices, we present only aggregated outcomes, ensuring that no individual student details are disclosed.

Our study aims to address the following research questions: (a) How effective are oral assessments in mitigating academic dishonesty in data analytics courses? (b) Can oral assessments enhance students' understanding and engagement with their work? (c) How can AI tools be utilised to increase efficiency in oral assessments?

Oral assessment for professional learners

The subject Data Analytics with Python is a 12-week online course offered three times a year with 30-35 enrolments for each offering. One of its assessments involves a data analytics project where learners work in groups of 3-4 to analyse real-world datasets. They clean the datasets, perform statistical analysis, develop machine learning models, and document their process in a technical report. In the 2023 first offering, 85% of learners scored 80% or higher on this project which raised academic integrity concerns. Despite improvements in the second offering, 77% still scored 80% or higher. Given that Data Analytics with Python is fully online, there is a clear need for invigilation for this assessment. To address this, an oral assessment component was introduced in the third offering of the subject. Each group gave a short oral presentation, and each member answered individualised questions about their project. The oral assessment made up 20% of the project grade, while written reports and code comprised 80%. This resulted in a more balanced distribution of marks and reduced the proportion of learners with high scores on the project to 43%. The introduction of oral assessments significantly helped mitigate academic dishonesty by requiring students to demonstrate their understanding in real time. This approach made it more difficult for students to depend on external sources or group members without having a comprehensive grasp of the material.

A lead academic carefully crafted individualised questions for 27 students (across 9 reports), spending approximately 4.4 minutes per question. The same academic also conducted oral presentations and interviews along with a second marker. This approach worked well for the small cohort. The teaching team observed that learners were enthusiastic about sharing their work. The marks also revealed that students with low participation in live classes performed worst in the oral assessment. Additionally, we identified potential indications of generative AI usage, as some students who were unable to justify their reasoning during the oral assessment had employed techniques in their reports that were not covered in the course material.

Oral assessment for undergraduate students

Insights gained from the Data Analytics with Python course significantly shaped the design of oral assessments for a larger cohort – Elements of Data Processing subject. This face-to-face undergraduate course is offered twice per year and accommodates 400-500 students. This necessitated the involvement of several junior teaching staff as examiners. These examiners had limited to no experience in conducting and assessing oral evaluations. Crafting individualised questions for each student by a lead academic proved impractical for such a large cohort. The following sections detail our strategies to address these challenges and ensure effective implementation.

In Elements of Data Processing, students work in groups of 3-4 on a similar assignment to Data Analytics with Python. To reduce the chance of AI misuse, an oral assessment component was added in Semester 2, 2023. Each group gave a short oral presentation, and each member individually answered questions about their findings. Each member was expected to be able to answer any question about their group project. Based on lessons from Semester 2, 2023, we improved the oral assessment in Semester 1, 2024. The new format included a longer time allocated for both the group presentation and the interview assessment which occurred right after their presentations.

Preparation

Instead of resisting an emerging technology, we used AI to create personalised questions for each student. Generating individual questions for 500 students is time-consuming and risks reusing questions. In this trial,

examiners were asked to enter student reports into Spark AI with the prompt: "Generate 20 challenging, conceptual questions for an oral assessment". Examiners were instructed to select 2-3 relevant questions per student and encouraged to pose their own follow-up questions based on students' responses to further challenge their understanding of the task. They were also advised to ensure that all questions aligned with the learning objectives of the assignment. University-hosted Spark AI ensured confidentiality. Students also had the option to opt out of AI-generated questions, but none chose to do so.

To ensure fairness, each oral assessment session had two markers, with final grades determined by averaging their marks. In the second round of implementation of oral assessments, we provided students with an improved and more detailed rubric and a 1-hour of presentation training session, including video examples of responses with varying levels. A 2-hour workshop was held for examiners, to discuss the rubric. Examiners marked a sample presentation individually and we facilitated a group discussion towards consistent marking. They were also trained to use Spark AI for question generation while maintaining confidentiality.

Oral assessment results

The data from three consecutive deliveries of the subject (Semester 1 2023, Semester 2 2023, and Semester 1 2024), after applying min-max normalization, is illustrated in Figure 1.
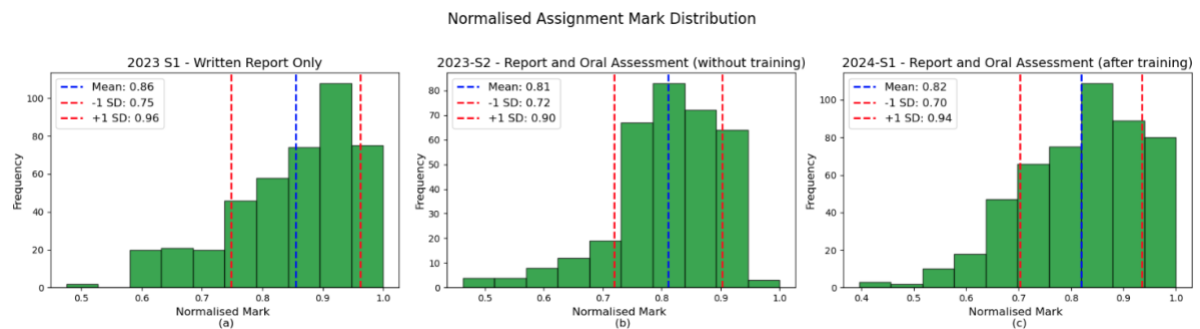

Figure 1. The distribution of normalised assignment marks (a) Semester 1 2023 (No Oral Assessment), (b) Semester 2 2023 (Oral Assessment without training) (c) Semester 1, 2024 (Oral Assessment after training).

In Semester 1, 2023, the assignment only included a written report, with a mean score of 0.86 (SD = 0.11), and 72% of groups scoring 80% or above, raising concerns about academic misconduct. About 10% of groups reported inactive members or members who admitted to using generative AI. In Figure 1(b), after introducing an oral assessment worth 25% of the total mark (21% for presentation, 4% for Q&A), the mean score dropped to 0.81 (SD = 0.09), with 61% scoring above 80%. Reports of inactive members dropped to around 6%. In Figure 1(c), after providing training and adjusting the oral assessment (12.5% each for presentation and Q&A), the mean score was 82% with a standard deviation of 0.14, 62% scoring above 80%, and only 3.5% reporting inactive members. The drop in inactive members indicates increased individual accountability and active engagement, reducing reliance on others' contributions. It suggests that oral assessments can improve engagement in group projects. Figure 2 compares the two oral assessment implementations.

In Figure 2(a) the first round of oral presentations (Semester 2, 2023), the mean score was 0.84 (SD = 0.17), with 72.7% of students scoring 80% or higher, indicating potential leniency or marking uncertainty. Some students missed the presentation, likely due to public speaking anxiety. In the second round, improved rubrics, as well as staff and student training reduced absences to zero. This may suggest that students felt better equipped to handle the oral component after the training. As shown in Figure 2(b), the marks distribution became more balanced, with a mean of 0.67 (SD = 0.26) and only 38.29% scoring 80% or above. Improved training and clearer rubrics made oral assessments more reliable, with fewer high scores reflecting consistent grading. Although there was a reduction in oral assessment marks the overall assignment marks for this project were higher. This could indicate the potential of oral assessments to encourage deeper engagement

with the subject matter as students were required not only to present their findings but also to defend their understanding under direct questioning.
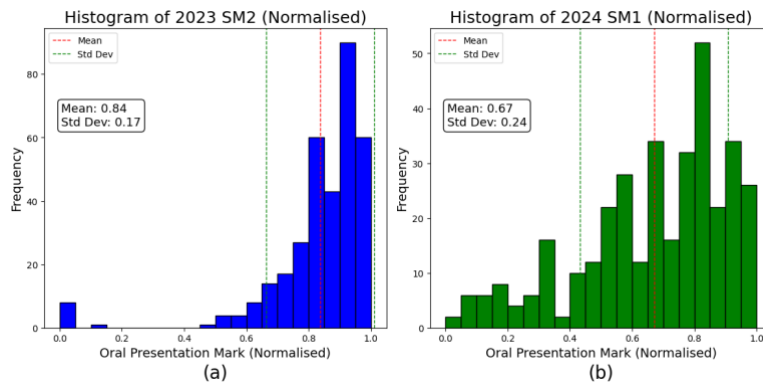

Figure 2. The distribution of oral presentation marks (a) 2023-S2 (without training), (b) 2024-S2 (after training)

In line with our quality assurance evaluation, we surveyed 16 examiners involved in oral assessments about the use of Spark AI in generating questions for students. According to the survey, 60% of examiners chose to combine AI-generated questions with their own, primarily to save time (75%) and ensure consistency (38%). However, 67% of examiners felt that AI-generated questions were too generic to meet assessment requirements, and 64% noted that students found AI-generated questions to be easier as they were more predictable and less probing. This may have contributed to the skewed marks. To address these issues, a clearer framework and instructions in prompt engineering are needed to generate more suitable and challenging questions.

Discussion

During the oral assessments, multiple observations and dialogues with students revealed that many found the experience valuable. They suggested that this assessment prepared them for real-life skills and deepened their understanding of the subject material. The oral assessment required them to comprehend the entire project, providing a holistic insight into the data analysis pipeline. Additionally, students reported that the oral assessment encouraged ongoing team discussions and prompted them to seek justifications for various decisions made by different members. For professional learners and undergraduate students alike, despite the mode of teaching, we observed that conducting oral assessments improved the learning experience of students. This also improved marking distributions and decreased the potential misuse of generative AI.

The use of Spark AI improved the efficiency of oral assessments by generating individualised questions based on students' project reports. This streamlined question generation allowed examiners to focus on assessing responses, especially in large cohorts. Privacy and confidentiality were protected using a university-hosted version of generative AI. Our findings and observations indicate that incorporating oral assessments alongside traditional written reports can play a crucial role in upholding academic integrity and the overall learning experience for students. Overall, we found that oral assessments can help improve student engagement and ensure that the work genuinely reflects the students' understanding. These results also confirm that adequate training for both students and examiners is essential in conducting a fair oral assessment.

Since we used generative AI as a trial, this presents some limitations. Firstly, we did not create a formal framework for generating questions. This contributed to some inconsistencies with the questions which were manually managed by examiners. To achieve better results, a possible solution is to train a GPT model specifically on the subject content to extract in-depth and probing questions (Zheng, 2023). Secondly, we have not yet obtained the human ethics approval which limits our capacity to perform more in-depth analysis. Further analysis can reveal some correlations between the potential misuse of generative AI and the oral assessment results.

Conclusion

The integration of oral assessments has proven valuable in maintaining academic integrity and enhancing students' communication and critical thinking skills in data analytics education regardless of the cohort and mode of teaching. While conducting oral assessments is time-consuming, both students and instructors find value in oral assessments, fostering deeper understanding and authentic engagement with the subject material. To facilitate oral assessment in a large cohort, incorporating generative AI for personalised question generation has improved efficiency. Further refinement of AI tools and comprehensive training for examiners can improve the process. Additionally, exploring various assessment formats, such as more interactive oral assessments, and addressing potential biases will aid in improving future implementations. The continued integration of technology to facilitate oral assessments can help educators identify common knowledge gaps and adjust teaching methods accordingly towards enriching the educational experience.

References

Accardo, A. L., Finnegan, E. G., Kuder, S. J., & Bomgardner, E. M. (2020). Writing interventions for individuals with autism spectrum disorder: A research synthesis. *Journal of Autism and Developmental Disorders, 50*, 1988-2006. https://doi.org/10.1007/s10803-019-03955-9

Gardner, D. E., & Giordano, A. N. (2023). The challenges and value of undergraduate oral exams in the physical chemistry classroom: A useful tool in the assessment toolbox. *Journal of Chemical Education, 100*(5), 1705-1709. https://doi.org/10.1021/acs.jchemed.3c00011

Grieve, R., Woodley, J., Hunt, S. E., & McKay, A. (2021). Student fears of oral presentations and public speaking in higher education: a qualitative survey. Journal of Further and Higher Education, 45(9), 1281-1293. https://doi.org/10.1080/0309877X.2021.1948509

Huxham, M., Campbell, F., & Westwood, J. (2012). *Oral versus written assessments: a test of student performance and attitudes*. Informa UK Limited. https://doi.org/10.1080/02602938.2010.515012

Mahendra, E. (2023). ORAL ASSESSMENT IN TRIGGERING STUDENT HIGHER-ORDER THINKING SKILLS (HOTS). *JISAE: Journal of Indonesian Student Assessment and Evaluation, 9*(1), 8-14. https://doi.org/10.21009/jisae.v9i1.32928

Sambell, K., Brown, S., Graham, L., Sambell, K., Brown, S., & Graham, L. (2017). Engaging students with positive learning experiences through assessment and feedback. *Professionalism in Practice: Key Directions in Higher Education Learning, Teaching and Assessment,* 139-187. https://doi.org/10.1007/978-3-319-54552-3_5

Tu, X., Zou, J., Su, W. J., & Zhang, L. (2023). What should data science education do with large language models. https://doi.org/10.1162/99608f92.bff007ab

Zheng, Y. (2023). ChatGPT for teaching and learning: an experience from data science education. Paper presented at the *Proceedings of the 24th Annual Conference on Information Technology Education,* 66-72. https://doi.org/10.1145/3585059.3611431