ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society

CritiQAL: Critiquing via questions and answers for assurance of learning

Thommen George Karimpanal
Deakin University

CritiQAL is a structured viva-based assessment design built using Generative Artificial Intelligence (GenAI) to support assurance of learning. The key idea behind CritiQAL is to exercise students' evaluative judgement – that is, the ability to assess and evaluate the quality of explanations or solutions. It was implemented in Trimester 1, 2026 in SIT796 Reinforcement Learning, a technical unit in Artificial Intelligence at Deakin University with a cohort of 139 postgraduate students. The tool was trained on unit materials to generate questions and suggest answers aligned to unit content, with parts of a suggested answer intentionally constructed to be incorrect. Students were tasked with answering questions and critiquing suggested answers, shifting the task from simple recall towards evaluation, justification and explanation. CritiQAL was used to conduct a viva task from week six onwards, with a full, half or zero-point rubric, where students were required to accumulate five points by Week 11. This paper outlines the rationale, design and implementation of CritiQAL as a practical response for assurance of learning in a context increasingly shaped by GenAI.

Keywords: evaluative judgement, assurance of learning, viva, generative AI, oral assessment

Introduction

The increased accessibility and fluency of generative artificial intelligence (GenAI) tools have created significant challenges for assessment in higher education (Bearman et al., 2024). In many disciplines, students can now obtain polished written explanations, summaries and even technically plausible responses with minimal effort. While these tools may sometimes support learning, they also risk reducing the productive struggle through which deeper understanding develops. In practice, this creates a growing gap between what students can submit and what they can actually explain, defend and critique independently. This challenge is especially salient in technically demanding domains such as artificial intelligence (AI), where student understanding is often not adequately revealed through polished written responses alone. Also, as GenAI can readily produce plausible but flawed explanations, it becomes increasingly important to design assessments that reveal whether students can judge the quality of an answer rather than merely present one.

This paper describes the design and implementation of a GenAI-based tool, CritiQAL (Critiquing via Q&A for Assurance of Learning), a structured viva-based assessment approach used in the postgraduate unit, SIT796 Reinforcement Learning at Deakin University in Trimester 1, 2026. We designed CritiQAL to assist with assurance of learning (Tertiary Education Quality and Standards Agency, 2025) by explicitly exercising students' evaluative judgement (Tai et al., 2018), which is the capacity to determine the quality of work, reasoning or explanations. CritiQAL presents students with randomly generated questions and partially erroneous suggested answers, engaging them in identifying, evaluating and explaining the flaws in those responses (Dieterich et al., 2025). In addition to separate written submissions, students were required to complete a CritiQAL-assisted viva task which tested core conceptual ideas in the unit.

This paper focuses on the pedagogical rationale, design and implementation of CritiQAL. As ethics approval was pending, this paper does not present formal analysis of student data or perceptions. It positions CritiQAL as an approach that brings together interactive oral assessment, evaluative judgement and learning through the critique of flawed examples to support assurance of learning in the age of GenAI.

Pedagogical rationale

The central idea behind CritiQAL is that students should not only be able to produce answers but should also be able to make judgements about the quality of an answer. CritiQAL deliberately shifts part of the emphasis

towards answer evaluation, as opposed to answer production. This shift is important because a student may know enough to state a familiar definition yet still struggle to recognise when a related explanation is partially wrong, misleading or overly simplistic. In technical subjects, this distinction matters, as concepts may be reproduced at surface level without being properly understood at the level needed to critique alternative formulations (Carless & Boud, 2018; Tai et al., 2018). By requiring students to inspect and evaluate a suggested answer, CritiQAL makes this dimension of learning visible. The design also aligns with the broader goal of helping students engage more actively with standards of quality. Instead of treating correctness as something determined solely by the teacher, the assessment asks students themselves to identify what is accurate/inaccurate and why. In this sense, it helps ensure that students are not merely recalling content and are instead reasoning about the adequacy of explanations.

A second pedagogical rationale for CritiQAL lies in the use of oral interaction (Nallaya et al., 2024), which provides opportunities for follow-up questioning, clarification and probing that are not easily replicated in written work (Tan et al., 2022). In CritiQAL, follow-up questions were an integral part of the design. This allowed tutors to explore whether a student genuinely understood a concept, whether they could refine an initial answer, and whether their reasoning remains coherent when question framing changed. This was particularly useful in distinguishing between superficial fluency and more robust conceptual understanding.

A third rationale concerns the educational implications of GenAI. If students can easily access answers generated by external systems, then educationally meaningful assessment needs to move beyond asking whether an answer can be produced. A more valuable question is whether students can recognise when an answer is incomplete, overly generalised or simply wrong. In that sense, CritiQAL treats evaluative judgement as a necessary capability in a learning environment increasingly saturated with AI-generated output.

Overall, CritiQAL aimed to make the viva itself a learning activity, where students practised explanation, critique and conceptual refinement over time. By distributing the viva across multiple weeks, it became a recurring site at which students could demonstrate and develop their understanding.

Context and tool design

CritiQAL was implemented in SIT796 Reinforcement Learning, a postgraduate unit at Deakin University, in Trimester 1, 2026. The cohort comprised 139 students. The tool was developed using custom GPT with unit materials such as lecture slides so that the generated materials remained closely aligned with the unit content and terminology. A tutor, in plain language, could ask the tool to generate questions and answers from a particular topic or week's content, at different levels of difficulty. This grounding in local unit content ensured that the assessment focused on the intended curriculum rather than on generic AI knowledge.

The key feature of CritiQAL is that suggested answers contain a mixture of correct, incorrect, incomplete or misleading elements. Students critically inspect these responses, identify flaws and justify their judgements, enabling tutors to probe conceptual understanding.

The suggested answers were deliberately kept short and cohesive, and appeared as brief paragraphs or bullet points. This was done to make the task manageable in a viva setting while also encouraging students to evaluate the full answer carefully. Importantly, the incorrect elements were varied in their nature and location in the answer. By varying where the flaw appeared, the design discouraged superficial answer-scanning and instead required students to engage critically with the answer.

Questions were also designed to be focused and conceptually coherent. Each question targeted a single core concept or a small cluster of closely related ideas rather than multiple unrelated sub-parts. Questions were not framed as true/false or other binary formats. This was pedagogically important because binary formats would reduce the demand for explanation and judgement. Instead, the questions required students to explain, distinguish, interpret or reason about reinforcement learning concepts in a more meaningful way.

The design also incorporated levels of question difficulty. "Pass" (P) and "Credit" (C) level tasks primarily focused on foundational understanding, terminology and basic conceptual ideas. These questions were

intended to check whether students had grasped the core ideas necessary for further learning in the unit. "Distinction" (D) and "High Distinction" (HD) level questions, by contrast, introduced hypothetical, counterfactual or "what if" scenarios. These scenarios tested whether students could transfer principles to unfamiliar contexts and demonstrate reasoning and application of principles. This ensured higher-grade performance expectations were less amenable to rote preparation and were based on sophisticated conceptual reasoning. We note that while CritiQAL facilitated the viva by generating questions across different topics, concepts and levels of difficulty, tutor judgement remained central to validating the difficulty, correctness and suitability of the generated content and evaluating student understanding.

Implementation

The CritiQAL viva task was introduced in SIT796, an 11-week unit, from week six onwards and focused on content from weeks one to five. Content up to week five was specifically chosen as it focused on the foundational concepts for the unit. Students were required to complete five questions, typically designed to take a total of 8-15 minutes per student, although the exact duration varied depending on the level of probing required. Students could make as many attempts as needed to complete the viva. The assessment was implemented in a way that reflected the mixed-mode nature of the cohort. On-campus students were recommended to complete their sessions in person, while online students completed the viva during online sessions. Additional individual viva slots were made available for students unable to attend regular sessions or for those uncomfortable doing the viva before their peers. This provided flexibility while preserving consistency in the assessment structure.

Responses were awarded a full point if the student answered correctly (as determined by the tutor) or a half point when the student demonstrated the core idea but the answer was incomplete or imprecise. A zero point was awarded when the student got the basic concept wrong. Students were required to accumulate a minimum of five points by week 11. The half-point category served an important pedagogical purpose of making visible the difference between partial understanding and fundamental misunderstanding. It served as a measure of the ease with which a student completed the viva, which could be considered a proxy for their conceptual clarity. Table 1 depicts hypothetical examples of student progression.

Table 1.
Hypothetical examples of student progression.

Student	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Total Points
Student 1	1P(full), 2P(full), 3P(full), 4C(full), 5HD(full)						**5**
Student 2	5P(full)	1P(full), 2C(full)			4D(full)	3HD(full)	**5**
Student 3	2P(half)	5P(zero), 5P(half)	1P(full), 3P(full), 2C(zero)	2P(half)	5C(full), 3P(zero)	3P(full)	**5.5**

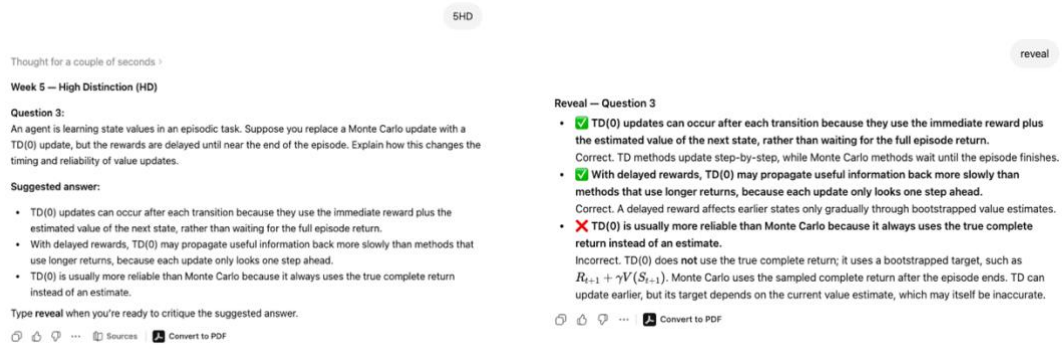
As seen in Table 1 above, all students achieve the minimum threshold of 5 points. P, C, D and HD denote question levels; and full, half and zero indicate degree of success of their attempts. For example, 1P(full) indicates a successfully attempted Pass level question from week one, 2C(zero) indicates a Credit level question from week five attempted unsuccessfully, and so on. Student one achieves the minimum threshold of five points in a single week (week 6), whereas student two makes progress across multiple weeks. Although student three achieved the minimum threshold of five points, they did so via several attempts, which indicates a possible lack of conceptual clarity. Students are not explicitly graded on these scores. Instead, their progression, including question level, number of attempts and degree of success, provided supplementary evidence of conceptual understanding that could be considered when assessing their separately graded written submissions.

Follow-up questioning was a central part of the implementation. After the student responded to a question and critiqued the suggested answer, the teacher could ask for clarification, request further justification or pose

variations on the original scenario. This helped determine whether the student was reasoning through the concept or relying on memorised language. It also allowed the teacher to see how students could defend or refine their explanation under mild conceptual pressure. To some extent, this approach may also have mitigated uncritical reliance on GenAI aids, as conceptually demanding questioning and follow-up probing made such reliance both harder to sustain and easier to detect.

As the questions were generated from unit material and the suggested answers were visible, follow-up questioning was especially important. Students could potentially prepare for the visible form of the task using external tools, including GenAI. The teacher's ability to probe the student's reasoning was therefore essential to the integrity and pedagogical value of the approach. As this could be particularly demanding to do, an optional "reveal" command was included as part of the tool as a tutor aid, which when used, revealed the correct and incorrect elements of the suggested answers.

Illustrative example



(a) (b)

Figure 1. Examples of the outputs of CritiQAL. (a) depicts an example of a week five HD question and answer generated through the tool. (b) depicts the "reveal" tutor aid functionality

As observed in Figure 1 above, HD questions focus on hypotheticals ("Suppose you replace…"). If needed, the teacher can then ask follow-up questions to further gauge the students' level of understanding.

Design implications and practical considerations

Several implementation considerations emerged from the design of CritiQAL. The first is that teacher expertise matters. The tool can generate questions and suggested answers, but the quality of the viva depends heavily on the teacher's ability to judge responses and ask productive follow-up questions. This is particularly true for tasks involving hypothetical scenarios, where deeper conceptual probing is often required and the teacher must be able to recognise whether a student's reasoning is genuinely sound. The "reveal" aid can mitigate this issue partially, but the quality of assessments still depends heavily on tutor expertise.

The second is that the method is labour-intensive. Even though each student's viva is relatively short, the cumulative workload of managing and scheduling viva slots, evaluating multiple responses per student and sustaining attention across many interactions is substantial. So, although CritiQAL may help generate rich information about understanding, it does so at a significant cost in terms of time and cognitive effort.

The third is that the method accounts for partial understanding. By using full, half and zero-point scoring, the design avoids collapsing all non-ideal responses into a single incorrect category. This enables a more fine-grained view of how students are engaging with core ideas. Although this paper does not report student data, the structure of the assessment itself reflects a view of learning as developmental rather than binary. An unexpected outcome of the viva was that it stimulated discussions even among students who had already completed their vivas. A significant increase in attendance numbers was also observed compared to previous years, although this effect was not systematically studied, and thus cannot be attributed to CritiQAL alone.

Lastly, while the implementation here was grounded in a unit-specific context, the core structure of asking students to answer and critique partially flawed responses could be adapted to many other disciplines.

Conclusion

CritiQAL was designed as a structured viva-based approach to assurance of learning in a postgraduate AI unit. Its distinctive feature is that students are required not only to answer discipline-specific questions, but also to critique suggested answers that contain both correct and incorrect elements. This shifts the task from answer production alone towards evaluation, justification and conceptual explanation.

The design is pedagogically significant because it directly exercises evaluative judgement, a capability that is increasingly important in a higher education context shaped by GenAI. If students can easily obtain plausible answers from external systems, then the capacity to recognise quality, diagnose flaws and defend reasoning becomes a more meaningful marker of understanding. CritiQAL was developed to address precisely this issue.

The implementation in SIT796 demonstrates how such a design can be integrated into a unit as a distributed oral viva, with foundational questions at pass/credit level and hypothetical scenario-based questioning at distinction/high distinction level. The approach also highlights important practical issues, particularly the need for teacher expertise and the workload required to implement the method effectively.

This paper has focused on the design, rationale and implementation of CritiQAL as a teaching innovation. Formal analysis of student perceptions and outcomes is outside the scope of this paper. More broadly, CritiQAL demonstrates how GenAI-generated content can be used as an object of critique rather than simply as a source of answers. This principle may be applicable across disciplines where students are expected to think critically by evaluating and questioning information presented to them.

References

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, *49*(6), 893-905. https://doi.org/10.1080/02602938.2024.2335321

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Dieterich, S., Rumann, S., & Rodemer, M. (2025). Conditions for effective learning from erroneous examples: A systematic review. *Educational Psychology Review*, *37*(4), 94. https://doi.org/10.1007/s10648-025-10071-x

Nallaya, S., Gentili, S., Weeks, S., & Baldock, K. (2024). The validity, reliability, academic integrity and integration of oral assessments in higher education: A systematic review. *Issues in Educational Research, 34*(2), 629–646. https://search.informit.org/doi/10.3316/informit.T202412180001400112109728 4

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education*, *76*(3), 467-481. https://doi.org/10.1007/s10734-017-0220-3

Tan, C. P., Howes, D., Tan, R. K., & Dancza, K. M. (2022). Developing interactive oral assessments to foster graduate attributes in higher education. *Assessment & Evaluation in Higher Education*, *47*(8), 1183-1199. https://doi.org/10.1080/02602938.2021.2020722

TEQSA. (2025). *Gen AI and learning assurance – new resource available*. https://www.teqsa.gov.au/about-us/news-and-events/latest-news/gen-ai-and-learning-assurance-new-resource-available

ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society

Karimpanal, T.G. (2026) CritiQAL: Critiquing via questions and answers for assurance of learning. In Cochrane, T., Narayan, V., Poole, T., Lai, P., Lake, P., Moore, O., & Li, L. (Eds.), *The Future is Connection: Collaboration, Creativity and Compassion in a Digital Society*. Proceedings ASCILITE 2026. Brisbane (pp. 654–659). https://doi.org/10.65106/apubs.2026.3084

Note: All published papers are refereed, having undergone a double-blind peer-review process.
The author(s) assign a Creative Commons by attribution license enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© Karimpanal, T. G. 2026