When Reports Become "Too Good": Rethinking Assessment in the Age of Generative AI

Carvalho, P.
Department of Computer Engineering (DEI)
School of Engineering, Polytechnic of Porto (ISEP)
Technical University of Porto (UTP)
Porto, Portugal
pbc@isep.ipp.pt

Abstract—Generative artificial intelligence (GenAI) is changing the conditions under which academic writing is produced and, with them, some of the assumptions that have traditionally supported its assessment. This paper discusses the implications of GenAI for assessment, drawing on experience supervising and assessing undergraduate Project/Internship reports and Master's dissertations in Computer Engineering. In recent years, particularly since GenAI tools became widely available, student reports have tended to become more polished, better structured, and more consistent in academic style. This observation does not establish that GenAI caused the change, nor does the quality of a report reveal whether or how AI was used. It does, however, raise an assessment question: how much can be inferred about a student's independent understanding and competence from the formal quality of a finished text when sophisticated writing support is readily available? The paper discusses this question by distinguishing communication quality from evidence of learning and considers how assessment can draw on a broader set of evidence, including the development process, implementation artefacts, critical reflection, and oral defence. The purpose is not to reduce the value of academic writing, but to reconsider what can reasonably be inferred from the final written product when AI-assisted writing is increasingly accessible.

Keywords—generative artificial intelligence; engineering education; academic writing; assessment; oral defence; evidence of learning.

I. Introduction

The arrival of widely accessible generative AI has changed academic writing remarkably quickly. In higher education, much of the first discussion understandably focused on authorship, plagiarism, and academic integrity. Those questions remain important, but they are not the only ones. There is another, more ordinary problem that becomes visible when one spends time reading and assessing student reports: the text itself has changed.

From the author's experience in supervising and assessing student work, this change has become particularly noticeable in undergraduate Project/Internship reports and Master's dissertations in Computer Engineering. Before GenAI tools became commonplace, weaknesses in writing were often easy to see. Reports could contain grammatical and syntactic errors, repetitive vocabulary, abrupt transitions, poorly organised sections, or technically correct ideas that students struggled to express clearly. These weaknesses did not necessarily mean that the underlying engineering work was poor. Nevertheless, they formed part of the submitted work and therefore influenced supervision, assessment, and sometimes the oral examination.

More recent reports often look different. The writing tends to be smoother, the structure more regular, and the vocabulary more sophisticated. This observation is not presented here as evidence of a causal relationship between GenAI use and writing quality. The paper is not based on a controlled longitudinal study, nor does a polished report indicate how much AI assistance was involved. The point is narrower and, for assessment, more important: the formal quality of a finished text has become much easier to improve with readily available tools.

This raises a central question: when sophisticated writing support is readily available, what can the formal quality of a finished report reveal about the student's knowledge and competences? This paper addresses that question from a teaching and assessment perspective, drawing on sustained experience in supervising and assessing undergraduate Project/Internship reports and Master's dissertations in Computer Engineering. It identifies a distinction that has become increasingly important in the GenAI context: the distinction between the formal quality of academic writing and the evidence it provides of the student's own understanding and competence. It then considers how assessment can draw on a broader range of evidence, with particular attention to the role of the oral defence.

Against this background, the paper focuses on three aspects: the assessment problem that becomes more visible in the GenAI context; the distinction between the formal quality of academic writing and evidence of disciplinary understanding; and the value of combining different sources of evidence, including the oral defence.

II. What Has Changed in the Written Product?

The change is easiest to describe by considering the kinds of problems that used to require considerable attention during supervision. Language errors, limited vocabulary, difficulties in structuring technical discourse, and problems of clarity and coherence were common enough to be part of the normal work of helping students prepare a final report. In some cases, discussion during the defence also returned to these issues.

Today, many of those visible weaknesses are less prominent. A student can use GenAI to reformulate a paragraph, improve transitions, suggest a more academic expression, correct grammar, reorganise material, or turn rough notes into fluent prose. Used carefully, these functions can be genuinely helpful. They may be especially valuable for students who understand their technical work but have difficulty expressing it clearly in academic English.

The educational benefit should not be dismissed. Clear writing matters in engineering. Graduates need to document systems, explain decisions, communicate with different audiences, and produce professional technical material. If GenAI helps a student communicate an idea more clearly, that can be useful.

The difficulty begins when attention moves from the quality of the product to the interpretation of that quality. In the past, a well-organised explanation could reasonably contribute to an assessor's judgement about the student's ability to organise and communicate knowledge. That inference is now less straightforward. The final wording may represent the student's own formulation, extensive AI-assisted revision, or something in between. In many cases, the document itself does not allow the assessor to know which.

This is why the issue is not simply whether GenAI makes writing "better". The more relevant question is whether the same features of the final report still carry the same evidential weight in assessment.

III. The Assessment Paradox

There is a paradox here. Better written reports are desirable. No teacher should wish for a return to unclear, poorly structured academic writing simply because it made individual authorship easier to infer. Yet a better text is not automatically better evidence of learning.

Reports and dissertations may appear better written without students' underlying knowledge and competences necessarily improving in the same proportion. This matters because written reports have often been asked to do several jobs at once. They communicate the work, document the process, demonstrate disciplinary knowledge, and provide evidence of skills such as analysis, synthesis, judgement, and critical thinking.

Recent literature reaches a similar concern from different directions. UNESCO's guidance discusses the longer-term implications of GenAI for education and the need for human-centred pedagogical responses [1]. In the assessment domain, the TEQSA expert guidance argues that GenAI has intensified existing assessment challenges and that assessment practices need to be reconsidered so that they can both manage risks and make constructive use of AI [2].

This does not make the written report irrelevant. It does, however, suggest the need for greater precision about what can reasonably be assessed through it.

IV. Text Quality And Student Competence

A useful way to make the problem concrete is to separate two dimensions that are often blended in marking.

The first is the quality of the text: grammar and syntax, structure and organisation, vocabulary, fluency, and academic style. These remain legitimate assessment criteria where communication is a learning outcome.

The second concerns the student's knowledge and competences: whether the student understands the work, can justify technical decisions, can explain the implementation, can interpret results, recognises limitations, considers alternatives, and can respond critically when assumptions are challenged. These two dimensions should not be treated as equivalent.

GenAI makes this distinction more important because it can support many of the visible features of the first dimension. It can improve a sentence without understanding the system being described. It can propose a coherent structure without having made the engineering decisions behind it. It can produce plausible explanations that the student may or may not be able to defend.

This is not an argument for assuming that polished writing is AI-generated, nor for penalising students because their writing is good. Such an approach would be both unfair and pedagogically unhelpful. Rather, it is an argument for avoiding an overly strong inference from polished writing to independent competence.

The practical implication is that assessors need additional opportunities to see the student thinking, deciding, explaining, and responding.

V. Why the Oral Defence Matters More

For Project/Internship reports and Master's dissertations, the oral defence already provides such an opportunity. In an AI-mediated writing environment, its role becomes even more valuable.

A defence allows the assessor to move beyond the wording of the report. The student can be asked why a particular architecture was selected, what alternatives were considered, how a component works, what caused an unexpected result, or what would be changed if the project were started again. Follow-up questions are particularly informative because they require the student to reason in real time rather than reproduce a prepared paragraph.

This does not mean turning the defence into an interrogation about AI use. In fact, repeatedly asking whether a student used a particular tool may reveal less about learning than asking good technical questions. The aim is to establish understanding, judgement, and ownership of the work.

The oral defence can therefore become a particularly valuable moment for assessing understanding: students can explain and justify technical decisions, discuss implementation and results, answer unexpected questions, reflect on limitations and alternatives, and demonstrate critical thinking and autonomy.

There is growing support for this direction. Hartmann reports the use of oral examinations as a response to GenAI-related concerns and discusses practical issues such as student anxiety, fairness, and instructor time [3]. More recently, Dao and Hwang propose a streamlined interactive oral assessment model intended to make individual reasoning visible through spontaneous explanation and responses to prompts [4]. These approaches do not remove the practical challenges of oral assessment, but they show why it can provide evidence that a finished written product alone may not reveal.

VI. From a Single Product to Multiple Sources of Evidence

Taken together, these observations suggest that too much evidential weight should not rest on one polished artefact. Rather than trying to determine, sentence by sentence, whether AI was involved, assessment can be strengthened by drawing on several sources of evidence.

For engineering work, these sources are already available in many courses: intermediate versions, design

choices, code or other implementation artefacts, demonstrations, testing decisions, project logs, supervision discussions, reflective commentary, and the oral defence. None of these is perfect on its own. Together, however, they provide a richer picture of what the student can do.

A broader approach shifts attention from the written report as a single product towards evidence of learning drawn from the report, development process, implementation and artefacts, oral defence, and critical reflection. The important change is not the removal of writing from assessment, but a reduction in the amount of evidential work expected from the final text alone.

Research on GenAI and assessment is increasingly moving in this direction. International survey evidence from three universities found that both educators and students perceived GenAI as affecting assessment, with educators showing a stronger preference for adapted assessments that assume AI use and emphasise critical thinking; the authors also argue for greater attention to the learning process, higher-order thinking, and authentic applications [5]. Foung, Lin, and Chen examined a redesigned writing assessment in which students used GenAI and other online tools and reflected on that use, illustrating an approach that makes tool choice and reflection part of the assessed learning process [6].

For engineering education, this approach is particularly appropriate. Engineering competence is demonstrated not only through description, but through decisions, implementation, testing, interpretation, and the ability to explain trade-offs. Assessment should make those activities visible.

VII. Implications for Assessment Practice

For assessment practice, this leads to a few direct consequences.

First, writing should continue to be assessed when written communication is part of the intended learning outcomes. GenAI does not remove the professional importance of clear technical communication.

Second, criteria need to distinguish more clearly between communication quality and evidence of disciplinary understanding. If both matter, both should be assessed, but they should not be allowed to stand in for one another.

Third, the process deserves greater visibility. Asking students to retain meaningful intermediate artefacts, explain important revisions, or discuss the evolution of a design can provide useful context without creating an excessive surveillance culture.

Fourth, oral questioning should be aligned with the competences that the assessment is intended to verify. A few well-chosen questions about decisions, alternatives, implementation, limitations, and results can reveal more than a general request to "summarise the project".

Finally, the objective should not be to make assessment "AI-proof". That goal is likely to become less realistic as tools become embedded in ordinary academic and professional work. A more sustainable objective is to design assessment in which appropriate tool use can coexist with credible evidence of individual learning.

This position is consistent with recent work that frames the GenAI challenge as a question of assessment validity and design, rather than only one of academic misconduct [2], [7].

VIII. Limitations and Next Steps

This paper should be read as a teaching reflection, not as an empirical study. Its observations come from accumulated experience of supervising and assessing student work; they were not collected through a predefined research protocol, systematic coding procedure, or dataset designed to compare cohorts before and after the widespread availability of GenAI. The paper therefore provides neither prevalence estimates of AI use nor evidence that GenAI caused the observed changes in writing quality. Other factors may also have contributed, including changes in student cohorts, supervision practices, institutional guidance, writing support, and students' increasing familiarity with digital tools.

A further limitation is that the discussion concerns a particular assessment context: undergraduate Project/Internship reports and Master's dissertations in Computer Engineering. The balance between written product, process evidence, and oral examination will differ across disciplines and assessment formats. The considerations discussed here should therefore not be read as a universal assessment model. Future work could compare report characteristics across cohorts, document how students use GenAI during the writing process, and examine whether changes in written quality correspond to oral performance or other measures of technical understanding. It would also be useful to investigate students' perceptions of fairness and transparency when greater weight is placed on process evidence and oral defence.

The immediate point is practical rather than causal: one familiar assessment assumption deserves closer scrutiny. When sophisticated writing assistance is readily accessible, a polished final document should be interpreted with greater care.

IX. Conclusion

GenAI does not make academic writing unimportant. Clear technical communication remains a legitimate educational and professional outcome, and writing support can help students' express ideas more effectively.

The issue is narrower: formal writing quality should not be asked to provide more evidence than it can support. In assessment contexts where GenAI can substantially assist the production and revision of text, stronger judgements about individual understanding are better grounded in several sources of evidence than in the finished report alone.

For Project/Internship reports and Master's dissertations in Computer Engineering, this means retaining the report while giving appropriate weight to the development process, implementation artefacts, critical reflection, and meaningful oral questioning. This is not presented as a tested assessment model, but as a practice-based direction that warrants systematic evaluation.

The central argument can therefore be stated simply: "Better writing is welcome. Better evidence of learning is essential."

References

[1] F. Miao and W. Holmes, Guidance for Generative AI in Education and Research. Paris, France: UNESCO, 2023.

[2] J. M. Lodge, S. Howard, M. Bearman, P. Dawson, and Associates, Assessment Reform for the Age of Artificial Intelligence. Australia: Tertiary Education Quality and Standards Agency (TEQSA), 2023.

[3] C. Hartmann, “Oral Exams for a Generative AI World: Managing Concerns and Logistics for Undergraduate Humanities Instruction,” College Teaching, published online Sep. 11, 2025, doi: 10.1080/87567555.2025.2558563.

[4] L. Dao and J. Hwang, “Quick Fix: Validating Learning in the AI Era Through a Streamlined Interactive Oral Assessment Model,” College Teaching, published online Aug. 10, 2026, doi: 10.1080/87567555.2026.2715403.

[5] R. F. Kizilcec, E. Huber, E. C. Papanastasiou, A. Cram, C. A. Makridis, A. Smolansky, S. Zeivots, and C. Raduescu, “Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States,” Computers and Education: Artificial Intelligence, vol. 7, Art. no. 100269, Dec. 2024, doi: 10.1016/j.caeai.2024.100269.

[6] D. Foung, L. Lin, and J. Chen, “Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices,” Computers and Education: Artificial Intelligence, vol. 6, Art. no. 100250, Jun. 2024, doi: 10.1016/j.caeai.2024.100250.

[7] B. H. R. Mohamed and A. Temimi, “Assessment design under imperfect information: generative AI, disclosure, and student response in higher education,” Assessment & Evaluation in Higher Education, published online Sep. 9, 2026, doi: 10.1080/02602938.2026.2729791.