ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society

When AI talks back: A cross-disciplinary study of educator design judgment in AI-assisted case development

Xuanxuan Zhang, Mary Jesselyn Co

Monash University

Educators are increasingly turning to generative AI when developing teaching materials, yet most research to date has centred on student-facing applications. This paper presents a work-in-progress study that asks a different question: what happens to an educator's design judgment when AI enters the case-writing process? Two business educators – one in accounting, one in management – will each develop a teaching case study using AI-assisted workflows, documenting their prompts, editorial decisions, and reasoning through structured process logs. Drawing on Schön's (1983) reflective practitioner framework and Abell's (1997) 10 criteria for effective teaching cases, the study examines where human expertise proves essential and where AI contributes meaningfully. The pairing of two sub-disciplines, management accounting and management, allows us to examine whether disciplinary knowledge shapes these design decisions in distinct ways, although the study's central concern remain the design judgement itself. Early observations from the instrument development phase point to a useful distinction: AI appears better suited to qualities that require evaluative judgment (structure, data accuracy, length) than those demanding generative judgment (ambiguity, controversy, authentic voice). We expect further patterns to emerge once data collection begins. This paper outlines the research design and its anticipated contributions.

Keywords: Generative AI, case-based teaching, reflective practice, curriculum design, human-centred learning design, cross-disciplinary, qualitative

Introduction

Much of the conversation about generative AI in higher education has, understandably, focused on students – how they use it, whether they should, and what it means for assessment integrity (Crawford et al., 2023; Lodge et al., 2023; Zhou & Luo, 2025). But there is another side to this shift that has received far less scrutiny: educators themselves are quietly adopting AI as a design tool, using it to draft materials, generate scenarios, and build resources for their units. A growing body of work now examines how educators themselves use AI in their teaching. Lodge et al. (2023) map a research agenda for generative AI in tertiary education and observe, from their own practice, that materials produced by AI require review by an experienced teacher to be usable, since instructional design still depends on human input. Al-Ali and Miles (2025) show that educators need targeted training to integrate AI into their teaching effectively, foregrounding the educator rather than the student. Within accounting, Mcbride and Rose (2026) document educators' responses to generative AI and their interest in using it to create case study materials. This work establishes that educators are turning to AI and that their judgment matters to the result, yet it centres on perceptions, training needs, and pedagogical responses rather than on the design judgment exercised in producing a case, and none of it evaluates the resulting materials against an explicit quality framework. Similarly, Panke (2025) offers an autoethnographic account of using generative AI across research, teaching and instructional design. That account, however, documents the process without evaluating the resulting materials against explicit quality criteria, and stops short of theorising what the educator's judgment contributes. We build on this work by pairing process evidence with a criterion-based assessment of case quality, and by reading educator AI use through Schön's reflective practitioner lens.

Teaching case studies offer a particularly revealing context for exploring this gap. A good case – one that sparks genuine debate in a classroom – needs more than accurate content and clean structure. It needs a compelling hook, characters that feel real, ambiguity that resists easy resolution, and enough tension that students disagree with each other (Abell, 1997; Herreid, 1997). These are qualities rooted in human judgment and lived experience, and they sit uneasily alongside the fluency and agreeableness of AI-generated text. Early

practitioner accounts of AI-assisted case writing have begun to appear, but they tend to document prompt-engineering technique rather than the underlying design judgment, and they rarely evaluate the resulting case against any explicit quality framework (Singh et al., 2023). This leaves open the question this study pursues: not how to prompt, but what the educator's judgment contributes and how good the output actually is.

This paper presents a work-in-progress study that investigates the following question: How do business educators exercise design judgment when using generative AI to develop teaching case studies, and how does disciplinary context shape this process? We situate the work within the conference sub-theme of human-centred learning design. Our interest is not in whether AI can write a case study – it can produce something that looks like one readily enough – but in what the educator brings to the process that the AI cannot. Every time an educator reads an AI-generated draft, notices something missing, and intervenes, they are making a design decision that protects the pedagogical quality of the learning experience. Those decisions are the focus of this study.

Conceptual framing

Two frameworks anchor the study (refer to Figure 1). The first is Schön's (1983) concept of the reflective practitioner. Schön argued that professional expertise operates not through the mechanical application of theory, but through a *reflective conversation with the situation* – the practitioner shapes the materials, the materials resist or respond in unexpected ways, and the practitioner adjusts. What makes AI-assisted design theoretically interesting is that the AI "talks back" very differently from traditional design situations. It is fluent and accommodating. It rarely pushes back. It produces plausible text without the productive friction that, in Schön's account, is what actually drives professional learning. Whether the reflective loop still functions in the same way under these conditions is an open question – and one this study is designed to explore.

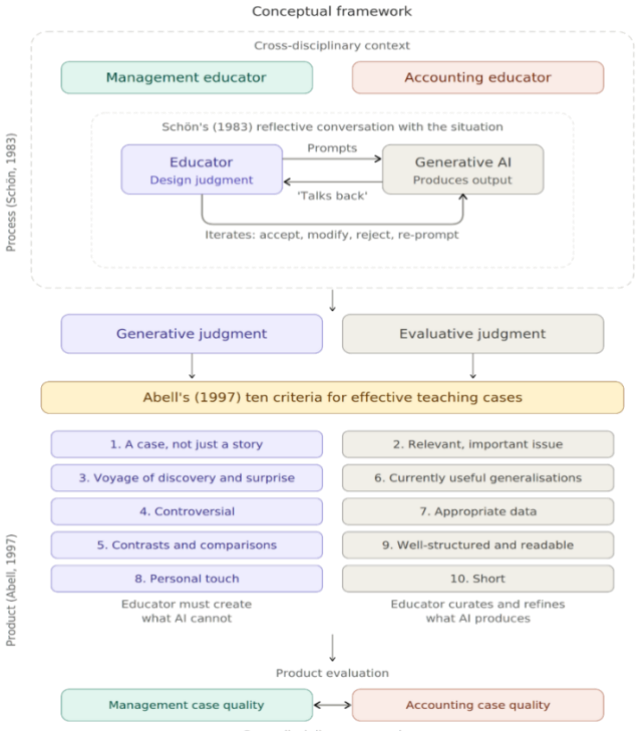


Figure 1. Conceptual framework based on Schön (1983) and Abell (1997)

The second framework is Abell's (1997) 10 criteria for effective teaching cases, drawn from his widely cited IMD technical note. Abell identified the qualities that distinguish excellent cases from run-of-the-mill ones: a genuine managerial issue with a hook; relevance to the audience; layered problems that unfold like an onion; genuine controversy where intelligent people disagree; productive contrasts; transferable generalisations; appropriate data (not too much, not too little); a personal touch through real characters and direct quotations; clear structure; and brevity. In our study, these criteria do double duty. They serve as the evaluation rubric for assessing the quality of the finished cases, and they provide an analytical vocabulary for categorising the kinds

of design decisions educators make as they work with AI. When an educator rewrites an AI-generated character description because it feels "flat," for instance, that's a criterion eight decision – a judgment about personal touch that the Abell framework helps us name and compare across cases. Although the framework predates generative AI by decades, it remains widely used in case-method scholarship. That scholarship continues to develop in response to internationalisation, a changing student body, and new technology (Mu & Hatch, 2025). Its plain, discipline-neutral vocabulary for case quality is what makes it useful here. It lets us see where AI-assisted writing succeeds and where it falls short, whatever the tool. We apply the criteria to the written case itself. Abell links some of them to the teaching note and companion videos, too, but those lie outside the scope of this study. It is also worth noting that Abell treats case writing as a form of research in its own right, since a series of contrasting cases can provide the building blocks of a study. This is how we approach AI-assisted case writing, not simply as producing materials, but as something worth studying.

Methodology

The study uses a comparative reflective practice design, an approach in the tradition of educators studying their own practice. Two educator-researchers from a large Australian business school – one in management, one in accounting – will each develop a teaching case study for use in their own units, using whatever AI-assisted workflow they would normally adopt.

Both educators will use the same AI tool throughout the case development process – Anthropic's Claude Opus 4.8, accessed through the claude.ai interface. This is a deliberate methodological choice rather than a convenience. Because we are interested in how disciplinary context shapes design judgment, we need to be confident that differences in the process logs reflect differences in what each educator does with AI output, not differences in what the AI produces in the first place. Standardising the tool keeps the AI consistent across both cases: the same model responds to both of us, in the same interface, with the same underlying capabilities. This also simplifies our analysis within the Schön (1983) framing. If both educators are having reflective conversation with the same interlocutor, we can compare those conversations directly, the AI "talks back" in the same way to both of us, and the variation lies in what each educator notices, questions, and decides to change. When the accounting educator rejects an output that the management educator might have accepted, or vice versa, we can attribute that to disciplinary knowledge and pedagogical reasoning rather than platform idiosyncrasies. We will document the model version, interface settings, and any changes to the platform that occur during the data collection period to ensure transparency and replicability.

The cross-disciplinary pairing is deliberate rather than incidental. A management accounting case and a management case place different demands on the writer. The accounting case combines quantitative structure, where figures must be internally consistent, with judgment about which qualitative factors matter. The management case leans more wholly on qualitative, open-ended material, where the same evidence can support several readings. This maps onto the study's central distinction: the accounting case rests on a mix of elements that can be evaluated against a method and elements that call for judgment, while the management case rests more heavily on what must be generated through judgment. Existing evidence is consistent with this, suggesting that generative AI manages lower- and mid-order tasks more readily than the higher-order judgment work that resists it (Tharapos et al., 2025). Comparing the two allows us to examine whether the nature of human design judgment varies depending on the disciplinary knowledge the educator brings to the table. Because the researchers are also the participants, several risks must be acknowledged. We may tend to see what we expect to see; recording our own decisions in real time may itself influence them; our prior familiarity with the rubric may shape the choices under study; and the two authors share an institutional setting. To address these, we code independently, retain version-tracked drafts that support each claim, evaluate one another's cases, and deliberately flag instances that run counter to our expectations. We recognise, however, that because we already know what the study is looking for, these interaction effects can be acknowledged but not fully removed. Our observations therefore speak to our own experience rather than that of educators less familiar with these tools or their theoretical underpinnings.

Data collection

The study draws on four data sources. The first and richest is a set of *structured process logs*, completed in real time during each AI interaction episode. For every prompt-response-decision cycle, each educator records

what they asked the AI, what it produced, what they did with the output (accepted, modified, partially used, rejected, or re-prompted), and – critically – why. Two additional fields capture discipline-specific notes and in-the-moment reflective observations. The second source is the *draft iterations* themselves, preserved with version numbering to make visible how each case evolved from raw AI output to finished teaching material. Third, both educators independently evaluate both completed cases using an *operationalised Abell rubric* with four-point scales for each criterion, producing a 2×2 matrix of self-evaluations and cross-evaluations. Finally, a *semi-structured reflective interview* between the two educators captures retrospective insights that only emerge after the process is complete.

Analysis

We will use a hybrid deductive-inductive thematic coding approach. The deductive layer draws on Schön's reflection categories and Abell's criteria; the inductive layer allows patterns we have not anticipated to surface from the data. Both researchers will code independently, with intercoder reliability checked on a 20-25% subset of coded units, where the unit of analysis is the individual prompt-response-decision episode recorded in the process logs (rather than the case, of which there are only two). The rubric evaluations will be analysed for criterion-level patterns – which aspects of case quality scored highest and lowest, and where did self-evaluations and cross-evaluations diverge? Findings will be triangulated across all four data sources.

Preliminary observations and anticipated contributions

Data collection has not yet begun, but the process of building the instruments has already surfaced something worth noting. When we operationalised Abell's 10 criteria into a rubric – defining observable indicators and scale descriptors for each one – we found that some criteria felt straightforward to assess while others resisted clean operationalisation. That resistance turned out to be informative. The criteria that were hardest to pin down in rubric language – voyage of discovery (criterion 3), controversy (criterion 4), and personal touch (criterion 8) – are also, we suspect, the qualities that AI-assisted workflows will struggle most to produce. They require the kind of judgment that depends on knowing what a real classroom argument sounds like, or what makes a character feel authentic rather than assembled from plausible attributes.

At the other end, criteria like appropriate data (criterion 7), structure (criterion 9), and brevity (criterion 10) were far easier to operationalise and seem more naturally suited to AI assistance – they involve well-defined, pattern-recognisable qualities where the educator's role shifts from creating to curating. Appropriate data (criterion 7) is harder to place. Assessing how much data a case contains is straightforward, but deciding what to include and what to leave out is not, since Abell ties that choice to what the case is meant to teach. We therefore expect this criterion to draw on both kinds of judgment rather than sitting cleanly on either side. This early observation points toward a distinction that may prove analytically useful: the difference between *generative* design judgment, where the educator must produce something the AI cannot – genuine ambiguity, lived authenticity, calibrated tension – and *evaluative* design judgment, where the educator assesses and refines what the AI has produced. This resonates with Wen and Wen (2024), whose study of educators building teaching materials with AI documents uncovered a recursive cycle of prompting, evaluating, and revising output; our distinction sharpens that account by separating the judgment that generates what the AI cannot do from the judgment that evaluates what it produces. If this distinction holds in the data, it reframes how we think about the educator's role in AI-mediated curriculum design. The educator is neither a wholesale author nor a passive editor. They are a reflective practitioner whose most critical contribution lies in precisely those qualities that make learning materials worth engaging with – qualities that are, at their core, human-centred.

We anticipate three contributions from this study. First, empirical evidence about how educators – rather than students – interact with AI in their teaching practice, addressing a gap in the current literature. Second, a transferable reflective protocol and case quality instrument that colleagues in other disciplines can adapt for their own AI-assisted material development. At this early stage, this instrument is the contribution best positioned to travel, and its value will be more fully realised as educators beyond the research team apply it in their own settings. The third is a comparison across the two sub-disciplines. It examines whether the nature of human design judgment varies across accounting and management, and, if so, what accounts for the

difference. The answer has practical value, since it would help business schools shape the support and AI guidance they offer staff in each discipline.

Finally, we note the limits of this design. With two educators, two cases, a single Australian business school, and a focus confined to case writing, the study is exploratory and is not intended to generalise in a statistical sense. Its value lies in the depth of what it reveals and in a reflective protocol that others can adapt, together with the generative/evaluative distinction it explores. Future work can examine these ideas with larger and more varied samples, further disciplines, and other kinds of teaching material. A particularly promising direction is to apply the protocol with educators who are not already immersed in the underlying frameworks, and to ask how familiarity with the work of Schön and Abell shapes the design decisions educators make when using AI.

References

Abell, D. (1997). *What makes a good case?* (IMD-3-0731). International Institute for Management Development. https://www.thecasecentre.org/products/view?id=12023

Al-Ali, S., & Miles, R. (2025). Upskilling teachers to use generative artificial intelligence: The TPTP approach for sustainable teacher support and development. *Australasian Journal of Educational Technology, 41*(1), 88–106. https://doi.org/10.14742/ajet.9652

Crawford, J., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence. *Journal of University Teaching and Learning Practice, 20*(3). https://doi.org/10.53761/1.20.3.02

Herreid, C. F. (1997). What makes a good case? *Journal of College Science Teaching, 27*(3), 163–165. https://www.ecsb.org/wp-content/uploads/2016/09/What-Makes-a-Good-Case.pdf

Lodge, J. M., Howard, S., Bearman, M., Dawson, P., & Associates. (2023). *Assessment reform for the age of artificial intelligence*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology, 39*(1), 1–8. https://doi.org/10.14742/ajet.8695

Mcbride, K., & Rose, J. (2026). Let's chat about ChatGPT: Exploring accounting educators' perceptions of nascent generative AI. *Accounting Education*. Advance online publication. https://doi.org/10.1080/09639284.2026.2657407

Mu, F., & Hatch, J. (2025). Innovations in the case method of teaching management, 2000–2024: A scoping review. *The International Journal of Management Education*, *23,* Article 101150. https://doi.org/10.1016/j.ijme.2025.101150.

Panke, S. (2025). How can (A)I research this? An autoethnographic exploration of generative AI in research, teaching and instructional design. *Journal of Teacher Education, 76*(3), 230–24. https://doi.org/10.1177/00224871251325065.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Tharapos, M., Lau, K. H., Peszynski, K., Nguyen, L. A., Magdziarz, S., Morton, E., Borg, V., & Duan, S. (2025). Generative AI in accounting education: Evaluating ChatGPT's role in assessment and skill development. *Accounting & Finance*, *65*(3), 3231-3244. https://doi.org/10.1111/acfi.70051.

Wen, W., & Wen, H. (2024). Bridging generative AI technology and teacher education: Understanding preservice teachers' processes of unit design with ChatGPT. *Contemporary Issues in Technology and Teacher Education, 24*(4), 582–611. https://citejournal.org/volume-24/issue-4-24/general/bridging-generative-ai-technology-and-teacher-education-understanding-preservice-teachers-processes-of-unit-design-with-chatgpt/

Zhou, A., & Luo, Y. (2025). Exploring the impact of generative AI on student learning in accounting. *Journal of Accounting Education, 72,* Article 100982. https://doi.org/10.1016/j.jaccedu.2025.100982.

ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society

Zhang, X. & Co, M. J. (2026). When AI talks back: A cross-disciplinary study of educator design judgment in AI-assisted case development. 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. 889–894). https://doi.org/10.65106/apubs.2026.3128

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.

© Zhang, X. & Co, M. J. 2026