LEARNING LETTERS

2024
Volume 4
Article 33
https://doi.org/10.59453/ll.v4.33

A technology-agnostic framework for designing assessments in the era of artificial intelligence

Jen Scott Curwood[1], Nick Kelly[2], Kazjon Grace[1], and Karly Lazarou[1]

[1] The University of Sydney

[2] Queensland University of Technology

There is an urgent need to understand and benefit from artificial intelligence within schools. However, government policies that focus on academic integrity and duty of care do not address how students can leverage AI to enhance their learning nor how teachers can intentionally design assessments to account for AI. The Australian Framework for Generative Artificial Intelligence in Schools suggests that assessments need to clearly state "how generative AI tools should or should not be used" while also permitting a "clear and unbiased evaluation of student ability". This is a worthy aspiration, yet there are presently few tools and examples to guide teachers in creating assessments. This conceptual paper draws from the field of design to articulate a framework for developing assessments that focuses on AI dialogue and trace-augmented critical reflection.

Keywords: artificial intelligence, composition, assessment, pedagogy

Corresponding author: Jen Scott Curwood, js.curwood@sydney.edu.au

Recommended citation: Curwood, J. S., Kelly, N., Grace, K., & Lazarou, K. (2024). A technology-agnostic framework for designing assessments in the era of artificial intelligence. *Learning Letters*, 4, Article 33. https://doi.org/10.59453/ll.v4.33

Introduction

Digital technologies have become an inextricable part of modern education. The advent of artificial intelligence (AI) tools now provides extended text responses to natural language prompts, which can pass for students' original writing. In fact, a recent report from TurnItIn analysed 200 million student papers and found that one in 10 may contain AI-written language in 20% of their content (Hoover, 2024). While assessment ostensibly provides students with an opportunity to represent their knowledge, school-based assessments also enable schools to make standards-referenced judgements, support selection decisions, and align with national curriculum (Newton, 2007). Because AI can produce high-quality responses to common educational tasks, it raises "fundamental questions about what educators worldwide should be teaching and how students should be assessed" (Bower et al., 2024 p. 2).

Policies that focus on academic integrity and duty of care do not address how students can leverage AI within disciplinary learning or how teachers can intentionally design assessments to account for AI. Polices also fall short of considering how "speculative capture" (Nichols et al., 2024, p. 5) entangles textual production with an AI platform's governing logics. As human and artificial cognition continues to increase in capability (Siemens et al., 2022), it can be challenging to discern, within an assessment response, what has originated from human cognition. Consequently, human-AI interactions need to be accounted for *by design* within the assessment process to give educators insight into the interplay amongst students' cognition, reflection, and composition.

We begin with the premise that the rapidly evolving digital landscape requires that school-based assessment fully accounts for students' learning practices, which are reflected within

their composition processes. Process-focused assessment allows teachers to “evaluate the process of learning as well as the outcome” (Dixson & Worrell, 2016, p. 157), which shifts the focus away from assessing outputs as the primary artefacts that represent student knowledge. The landscape of AI tools, capabilities, and approaches is so broad and fast-moving that treating it like a fixed technology with a fixed set of skills and a fixed curriculum will be insufficient. Instead, for inspiration on how to concurrently assess both AI skills and acquisition of relevant knowledge in such a dynamic environment, we turn to design pedagogy.

The creative design domains – including architecture, interaction design, product design, and more – have a tradition of practice-based studio education where formative assessment is in the form of "crits”, or constructive critical review by practitioner-educators (Tovey, 2015). In domains where each student or group may be solving a completely different problem in a completely different way using a completely different methodology, feedback focuses as much on the acquisition of self-analysis, autonomy, critical thinking, and other metacognitive strategies as it does on demonstrating proficiency in the specific domain of relevance (Tovey, 2015). The key idea here is that *design is a practice*, and a student’s development of that practice, can only be assessed through observing and reflecting on the process of design, not its outputs. Daniel et al. (2023) have called for a “new form of evaluation” of the “process and products” (p. 34) of AI-augmented writing, and inspired by design pedagogy, we propose this can be achieved through leveraging AI-human dialogue and promoting trace-augmented critical reflection.

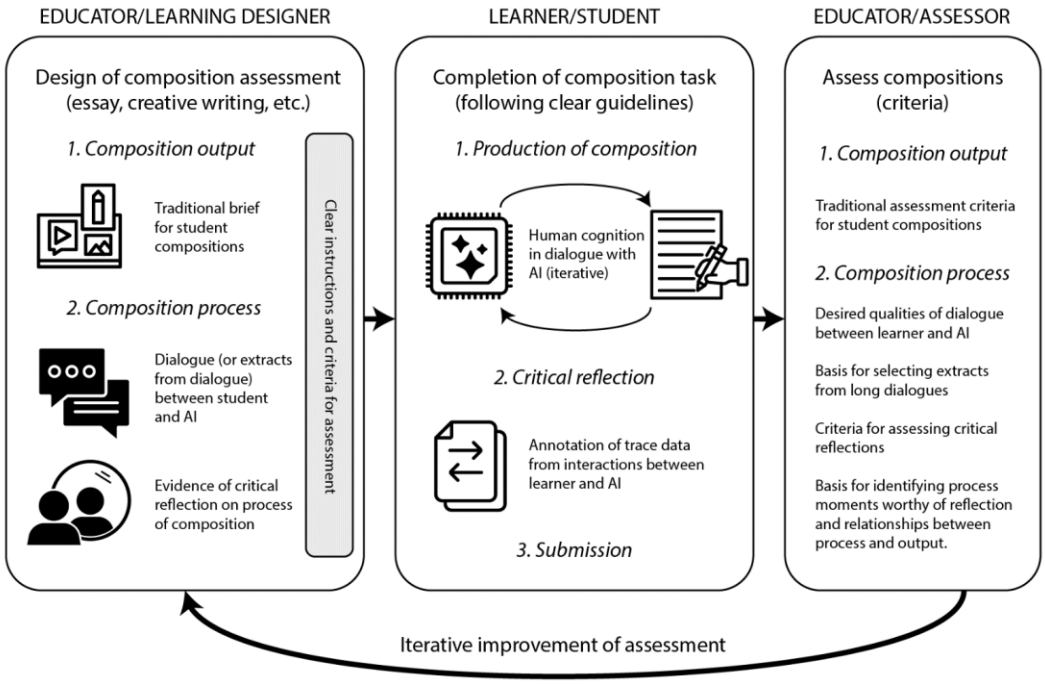
A framework for developing assessments

Rather than positioning AI tools as a hindrance or a threat to fair and ethical student assessment, teachers need an approach to assessment which recognises that AI technologies will evolve and proliferate, likely in ways that current education policies and systems do not envision, let alone account for. We offer a framework that is technology-agnostic, in that it is designed for the current wave of AI tools and will continue to be relevant as those tools evolve (see Figure 1). The framework builds upon traditional composition assessments by introducing two interrelated measures for assessing the *process* of composition:

1. The *dialogue* that occurs between learner and technology; and
2. The need for *critical reflection* upon this creative process, of which dialogues with technology are a part.

The framework responds to the Framework for Generative Artificial Intelligence in Schools’ (Australian Government, 2023) call for teachers to *show how AI tools should be used* and to ensure *clear and unbiased assessment*. These elements are expanded upon below and then applied within an illustrative example from the discipline of English. The framework allows us to envision an arc for students’ engagement with AI throughout high school as they learn how to utilise AI to support their development as writers within specific disciplines.

Figure 1: A technology-agnostic framework for developing composition assessments



Dialogue between learners and AI

The output produced by an AI tool emerges from an ongoing interaction between the user's original contributions and the AI's data-derived responses (Tang et al., 2024). This interaction between human and technology can be conceived of as a dialogue, involving both words and actions from each party (Hornbæk & Oulasvirta, 2017). The submission of the *dialogue between learner and AI* as an artefact of the student learning process provides teachers with insights into the pathways of their students' thinking and learning. AI-learner dialogues reveal where the student started, which party contributed what, and what strategies were used to conceive, develop, refine, and finalise work. This is critical for establishing academic integrity, but it also encodes content knowledge and directs teacher feedback. Boud and Molloy (2013) argue that teachers need a "richer conception of what feedback is and a broader notion of its scope" (p. 5) in order to overcome simplistic notions or outright misconceptions about the function and value of feedback. Whenever a student decides to integrate AI-generated content, they are making a judgement of that content's quality, and that judgement will be reflected in the dialogue, which provides a basis for teachers to offer targeted, meaningful feedback.

Human-AI collaboration leaves procedural traces: the "chat history", or record of messages between the student and the tool that led to the eventual artefact, and it is this dialogue which we propose educators focus on for both critical reflection and assessment. In an LLM, this might be a literal chat history, and in a text-to-image context this might be a family tree of captioned images, connected by links indicating changes to prompts, and so forth (as in Secretan et al., 2008). Agnostic of AI modality or capability, these traces represent the AI-learner dialogue in a way that can be employed to augment reflection, learning, and assessment alike. While it has been suggested that passive approaches to interacting with AI may reduce learning performance (Wang et al., 2023), learner-centric active engagement with AI may do the opposite (Long & Magerko, 2020). Specifically, it can support what Hattie and

Timperley (2007) conceptualise as error detection skills and self-regulatory proficiencies by encouraging students to use AI within self-feedback.

Categorisations of learner-AI interactions are beginning to emerge in the literature. Shibani (2023) suggests that LLM-assisted writing includes planning/ideation, information seeking/evaluation, and writing/presentation, each of which can be demonstrated deeply, shallowly, or not at all. Within each of these categories (and perhaps more as the field develops), there are strategies for how to engage effectively and critically with the tool, most of which require the learner to demonstrate knowledge of the content. For example, whenever a student asks an AI tool to revise content, we know good AI practice is to provide specific, actionable feedback on how the content should be revised (Diao et al., 2023, Wang et al., 2024). The AI dialogue represents a window into a student's writing process not unlike finding the old, discarded journals of a famous author, full of half-baked ideas and discarded first-drafts. Just as in design pedagogy, the best window into a student's practice is encouraging reflection on their process (Coorey, 2012), and class-level reflective discussion of AI-augmented writing has been shown to help students develop their notions of writing, authorship, and academic integrity (Fyfe, 2023). Building on this emerging practice, classroom activities could be constructed around modelling, critiquing, sharing, and discussing strategies exhibited in AI-learner dialogues.

Trace-augmented critical reflection

By leveraging the traces of human-AI dialogue, students can engage in critical reflection to enhance understanding of their writing processes and disciplinary learning practices. Again, taking inspiration from design pedagogy, this can take the form of *annotations* on the dialogue (Ball et al., 2009) or else another kind of reflective activity (e.g., reflective writing task about the dialogue, video discussion of the dialogue, presentation, etc.). By including such activity within the assessment, students are encouraged to reflect upon their co-composition process with the technology by considering key moments that were instrumental in shaping the resulting output. This serves purposes for both student and educator. It aids the student's learning by needing to reflect upon their composition process and their own role within this. It also aids the educator in making a (likely unwieldy) dialogue more easily assessable by highlighting salient features in a dialogue, aiding judgements about which parts of a composition are AI-initiated or human-initiated, and offering insight into students' evolving understanding of how they represent their disciplinary knowledge within a given assessment task.

This kind of activity can be customised to feed into further exploration of AI capability, academic integrity, or content. It can also be construed as a *peer* reflection activity in which students peer review one another's AI-learner dialogue. In this kind of assessment-as-learning activity, students might be asked to critique the AI dialogue strategies of their peers: How did they use AI, and in what ways was/wasn't that successful? What alternatives might be tried? What content knowledge could have been used better in prompts? How did the AI dialogue address key elements of the assessment task? In addition, teachers can actively model how to generate and reflect upon human-AI dialogue, and learners can share transcripts in small groups, identifying and categorising successful (and unsuccessful) ways that working with the AI tool helped their learning.

Worked example: Applying the framework to the English curriculum

In the context of English, the inclusion of dialogue and/or critical reflection alongside composition not only highlights the pathways to the conclusions that students arrive at, but

also reveals the specific choices they make in refining and representing their disciplinary knowledge. The framework accounts for student agency in their interactions with AI, but also requires accountability in terms of how AI contributes to their processes and outputs, which allows students to demonstrate their ability to grapple with disciplinary understanding, analysis, and evaluation within a given output. By including the AI dialogue, or the critical reflection, or both, within an assessment task, teachers obtain evidence of students' processes in synthesising knowledge and developing skills that sit at the convergence of artificial and human cognition (Markauskaite et al., 2022).

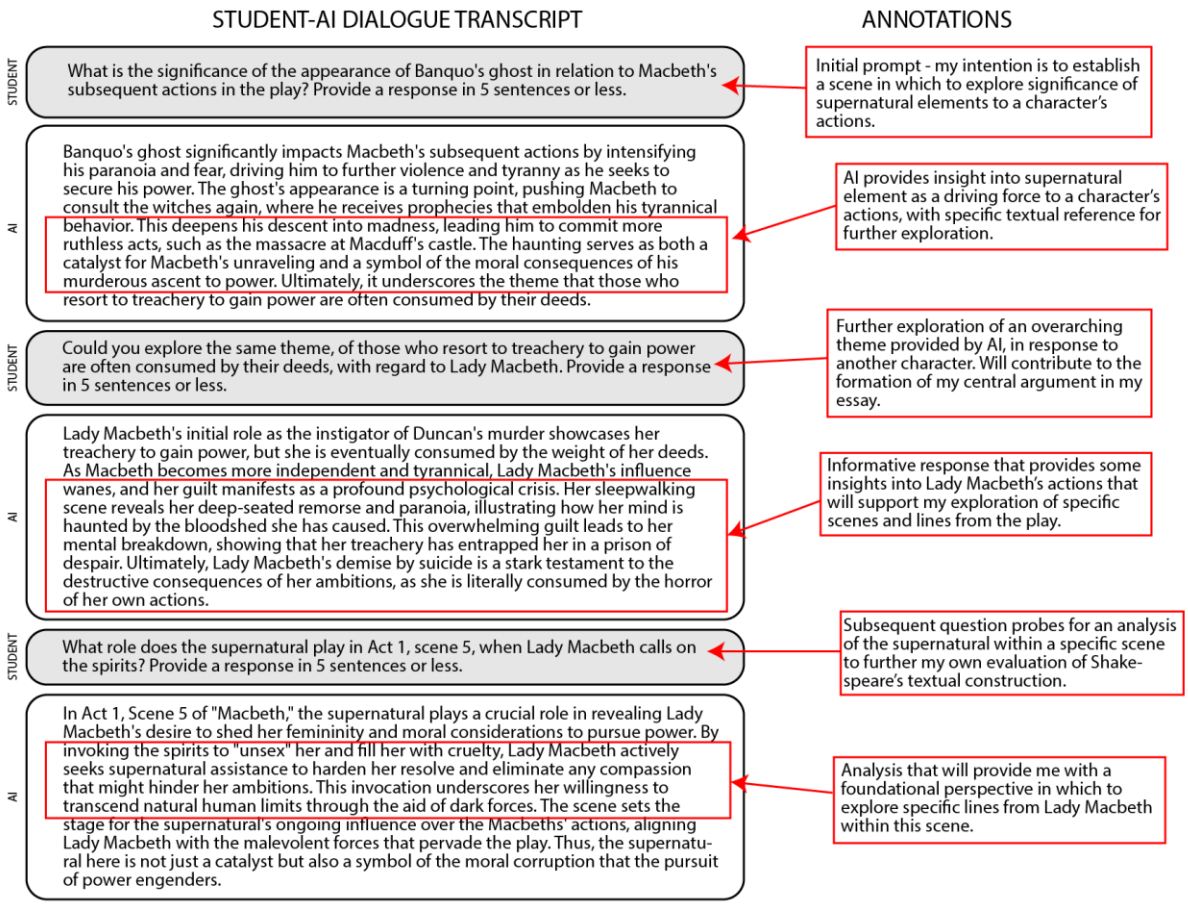
The framework encourages students to critically engage with AI-generated content and fosters a comprehensive understanding of the tool. To demonstrate, we examine a common task type: an analytical essay on a Shakespearean drama, a text selection which aligns with the requirements of the New South Wales Stage 5 English syllabus (New South Wales Education Standards Authority, 2024a) for students in Years 9 and 10. The proposed task asks students to use AI and compose an analytical response:

> In *Macbeth* every supernatural event either provokes an action or results from one. Considering the above statement, analyse how Shakespeare constructs his play *Macbeth* with events of supernatural intervention that shape the decisions, actions and ultimately the fate of key characters in the play.
>
> Within your response, analyse Shakespeare's use of language, dramatic and structural features, submit an annotated transcript of your AI dialogue, and/or include a critical reflection evaluating your use of AI.

Our conceptualisation for the student submission for an assessment such as the one above, would be a two-part submission of both *product*, the analytical essay, and the *process*, namely the AI dialogue and/or a critical reflection that evaluates their use of the tool, and the value it contributed to the submitted product. This AI-augmented composition that incorporates the framework would enable teachers to effectively and accurately assess student achievement while also gaining insight into how students developed their literary knowledge through utilising an AI tool. In teachers' marking criteria for the AI dialogue and critical reflection, they can include specific criteria for assessing students' *processes* to examine how they:

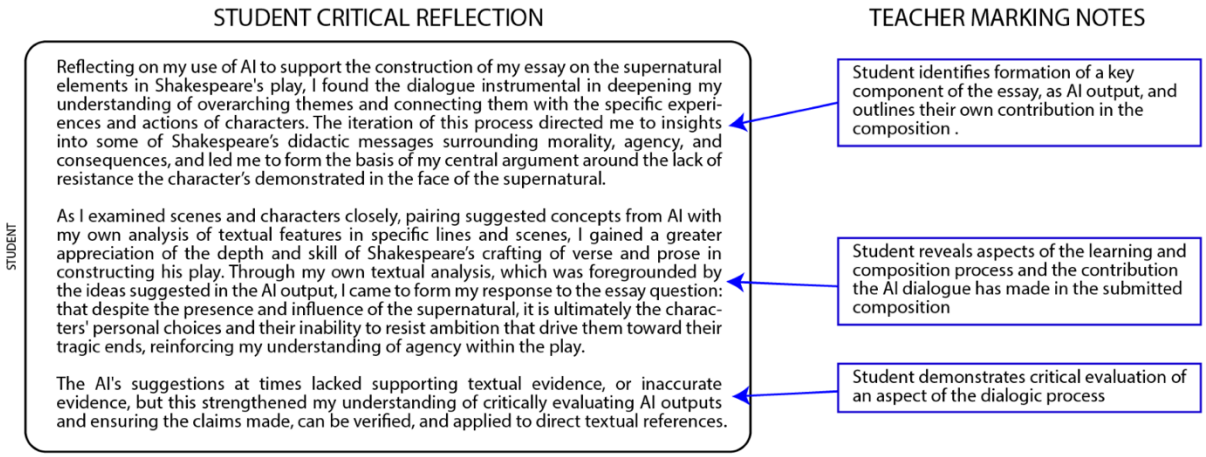
- Demonstrate effective ability to engage in AI dialogue with accuracy, relevance, and iteration.
- Show evidence of meaningful understanding and continuation of AI prompts (McTear & Ashurkina, 2024).
- Provide evidence of divergent thinking (Markauskaite et al., 2022) and inferencing (McTear & Ashurkina, 2024).
- Reflect critically on how the AI dialogue deepens understanding of disciplinary knowledge.
- Consider thoughtfully how engagement with AI shapes development of concepts and arguments within the composition process.

Figure 2: Example of an annotated dialogue between student and AI in response to brief



We propose that a task structured by teachers in this way could facilitate a more precise evaluation of students' abilities to monitor, revise, and reflect in order to refine text composition (New South Wales Education Standards Authority, 2024b), particularly through the annotated AI dialogue (see Figure 2). The practice of critical reflection can identify the emergence and development of ideas around the supernatural elements within the play, and the way in which these were then furthered by the student through a close examination of textual evidence and Shakespeare's construction of scenes, events, and characters (see Figure 3). The development of a central argument, and the ways in which the student came to this may also be highlighted, alongside choices in vocabulary, register, and paragraph or sentence structures. By looking to the field of design for inspiration, our framework offers conceptual and practical ideas for teachers to gain new insights into how students learn as well as how they can effectively and ethically represent their knowledge.

Figure 3: Example of student critical reflection and teacher marking notes.



Discussion and conclusion

There is an urgent need to understand and benefit from artificial intelligence within schools; however, there is a lack of congruence between the approaches of educational institutions (as seen in policies and assessments) and the technological capacity of current (not to mention future) AI platforms readily accessible to students. While tertiary design education has the luxury of regular in-person critique as the primary means of formative assessment (Michela, 2022), the realities of secondary schooling, including curriculum demands and time constraints, have often led to an emphasis on product or performance-based assessments as key measures of students' content learning. The technology-agnostic framework offers teachers a way to leverage AI though incorporating AI dialogues and trace-augmented critical reflections into the assessment process.

Moving forward, there would be benefit from a substantial investment in teacher professional development so teachers can develop capacity to utilise AI within their disciplinary pedagogy and assessment in line with the Australian Framework for Generative Artificial Intelligence in Schools. By carving out time and space for teachers to explore how AI will impact programming, planning, and assessing within their discipline, schools might take a positive and proactive approach to learning within the rapidly shifting digital landscape rather than a negative and reactive one. Thoughtful, intentional application of AI in schools can work to combat the negative impacts of teachers' workload and work intensification on their health, wellbeing, and attrition (Creagh et al., 2023).

Future theoretical and empirical work can explore pedagogically aligned applications of AI to augment how teachers offer constructive feedback and create meaningful assessment tasks. The framework also raises questions around teacher workloads, especially around assessment in the context of AI. There would be benefit from AI tools that are explicitly designed to aid teachers in assessing student work to guide them in implementing the proposed framework.

Funding

No funding was received for the conduct of this research.

Disclosure statement

The authors report no potential conflict of interest.

Disclosure of the use of AI-assisted technologies during writing

No AI-assisted technologies were used during the writing process.

About the authors

Jen Scott Curwood is an Associate Professor of English Education and Media Studies in the Sydney School of Education and Social Work at the University of Sydney, Australia. Her research explores the intersections of literacy, creativity, and technology in order to advance social change and build more sustainable futures. As a university teacher educator and former secondary English teacher, she examines how teachers can leverage innovative technologies to cultivate students' digital and critical literacies. She is co-author of the book *Qualitative Methods for Researching Online Learning*.

ORCID: https://orcid.org/0000-0003-0995-2078

Nick Kelly is an Associate Professor of Design Science in the School of Design at Queensland University of Technology, Australia. He is an interdisciplinary scholar working at the intersection of Design and Education. He has published over 100 scholarly works centred on themes of design cognition (how designers think), metacognition for learners, and places where these two things come together (design pedagogy, design for learning, learning by design, design of learning technologies, design of schools, co-design). He is an Australian Endeavour Fellow, co-author of the recent book *The Art of Co-Design: A Guide to Creative Collaboration*, and a Chief Investigator in the project *Thriving in Vertical Schools* https://verticalschools.org.

ORCID: https://orcid.org/0000-0001-8621-105X

Kazjon (Kaz) Grace is a Senior Lecturer in the School of Architecture, Design, and Planning at the University of Sydney, Australia. His research brings together what we know about human creativity with expertise in AI and machine learning. Specifically, with his team at the Designing with AI Lab, he is exploring how computers can be partners in the creative process rather than just tools, helping us develop, refine, and evaluate our ideas.

ORCID: https://orcid.org/0000-0002-0096-899X

Karly Lazarou is a PhD candidate at the Sydney School of Education and Social Work at the University of Sydney, Australia. She is an English teacher educator and a secondary school English teacher. Her research explores generative AI in subject English and the influence of such technologies on writing pedagogy and assessment and English curriculum design. Her work also considers the role of teacher voice and teacher professional development in response to emerging technologies.

ORCID: https://orcid.org/0009-0003-2280-6275

References

Australian Government. (2023). Australian Framework for Generative Artificial Intelligence in Schools. https://www.education.gov.au/schooling/resources/australian-framework-generative-artificial-intelligence-ai-schools

Ball, E., Franks, H., Jenkins, J., McGrath, M., & Leigh, J. (2009). Annotation is a valuable tool to enhance learning and assessment in student essays. *Nurse Education Today, 29*(3), 284-291. https://doi.org/10.1016/j.nedt.2008.10.005

Boud, D. & Molloy, E. (2013). *Feedback in higher and professional education*. Routledge.

Bower, M., Torrington, J., Lai, J. W. M., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative artificial intelligence? Outcomes of the ChatGPT teacher survey. *Education and Information Technologies*, *29*, 15403–15439. https://doi.org/10.1007/s10639-023-12405-0

Coorey, J. (2012). Reflective methods in design pedagogy. In P. Israsena, J. Tangsantikul, & D. Durling (Eds.), *Proceedings of the Design Research Society international conference, 2012. vol. 1: Research – Uncertainty, contradiction and value* (pp. 349–365). Design Research Society.

Creagh, S., Thompson, G., Mockler, N., Stacey, M., & Hogan, A. (2023). Workload, work intensification, and time poverty for teachers and school leaders: A systematic research synthesis. *Educational Review.* https://doi.org/10.1080/00131911.2023.2196607

Daniel, S., Pacheco, M., Smith, B., Burriss, S., & Hundley, M. (2023). Cultivating writerly virtues: Critical human elements of multimodal writing in the age of artificial intelligence. *Journal of Adolescent & Adult Literacy, 67*(1), 32–38. https://doi.org/10.1002/jaal.1298

Diao, S., Lei, Y., Pan, L., Fang, T., Zhou, W., Keh, S., ... & Zhang, T. (2023). Doolittle: Benchmarks and corpora for academic writing formalization. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 13093–13111). Association for Computational Linguistics.

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into Practice, 55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

Fyfe, P. (2023). How to cheat on your final paper: Assigning AI for student writing. *AI & Society*, *38*(4), 1395–1405.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hoover, A. (2024, April 9). Students are likely writing millions of papers with AI. *Wired*. https://www.wired.com/story/student-papers-generative-ai-turnitin/

Hornbæk, K., & Oulasvirta, A. (2017). What is interaction? In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 5040–5052). Association for Computing Machinery.

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). Association for Computing Machinery.

McTear, M., & Ashurkina, M. (2024). *Transforming conversational AI: Exploring the power of large language models in interactive conversational agents*. Apress. https://doi.org/10.1007/979-8-8688-0110-5

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence, 3*, 100056. https://doi.org/10.1016/j.caeai.2022.100056

Michela, E. (2022). "This uncertain space of teaching": How design studio instructors talk about design critiques along with themselves when giving critiques. *Journal of the Scholarship of Teaching and Learning 22*(1), 48–66.

New South Wales Education Standards Authority. (2024a). *English K-10 syllabus: Course overview.* https://curriculum.nsw.edu.au/learning-areas/english/english-k-10-2022/overview#course-requirements-k-10-english_k_10_2022

New South Wales Education Standards Authority. (2024b). *English K-10 syllabus: Content Stage 5:*

Expressing ideas and composing texts B. *https://curriculum.nsw.edu.au/learning-areas/english/english-k-10-2022/content/stage-5/fa8d3ee0b3*

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice, 14*(2), 149–170. https://doi.org/10.1080/09695940701478321

Nichols, T. P., Thrall, A., Quiros, J., & Dixon-Román, E. (2024). Speculative capture: Literacy after platformization. *Reading Research Quarterly*, *59*(2), 211–218. https://doi.org/10.1002/rrq.535

Secretan, J., Beato, N., D Ambrosio, D. B., Rodriguez, A., Campbell, A., & Stanley, K. O. (2008). Picbreeder: Evolving pictures collaboratively online. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1759–1768). Association for Computing Machinery.

Shibani, A. (2023). Analytic techniques for automated analysis of writing. In O. Kruse, C. Rapp, C. M. Anson, K. Benetos, E. Cotos, A. Devitt, & A. Shibani (Eds.), *Digital writing technologies in higher education* (pp. 317–331). Springer. https://doi.org/10.1007/978-3-031-36033-6_20

Siemens, G., Marmolejo-Ramos, F., Gabriel, F., Medeiros, K., Marrone, R., Joksimovic, S., & de Laat, M. (2022). Human and artificial cognition. *Computers and Education: Artificial Intelligence, 3*, 100107.

Tang, K.-S., Cooper, G., Rappa, N., Cooper, M., Sims, C., & Nonis, K. (2024). A dialogic approach to transform teaching, learning & assessment with generative AI in secondary education: A proof of concept. *Pedagogies: An International Journal, 19(*3), 493–503. https://doi.org/10.1080/1554480X.2024.2379774

Tovey, M. (2015). Developments in design pedagogy. In G. Bingham, D. Southee, J. McCardle, A. Kovacevic, E. Bohemia, & B. Parkinson (Eds.), *DS 82: Proceedings of the 17th international conference on engineering and product design education* (pp. 188-193). The Design Society.

Wang, T., Chen, J., Jia, Q., Wang, S., Fang, R., Wang, H., ... & Zhou, W. (2024). Weaver: Foundation models for creative writing. *arXiv preprint arXiv:2401.17268*.

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education, 194*, 104703. https://doi.org/10.1016/j.compedu.2022.104703