ASCILITE 2026
The Future is Connection:
Collaboration, Creativity and Compassion in a Digital Society

The apprenticeship layer: What assessment reform cannot secure on its own

Carlo Iacono
Charles Sturt University

Australian higher education has responded to generative artificial intelligence with an ambitious program of assessment reform. Structural approaches, most visibly the two-lane model, seek to sustain trust in degrees by securing some assessments while opening others to AI partnership. This conceptual paper argues that the reform consensus, now sharpened by sector attention to assurance of learning and adaptive capabilities, addresses validity more directly than formation. Assessment tasks are not only measurement occasions; they can also be sites of the cognitive work through which capability develops. Drawing on cognitive apprenticeship, desirable difficulties and cognitive offloading, the paper introduces the apprenticeship layer: the stratum of entry-level cognitive work through which expertise has historically been developed and renewed. As AI is increasingly used to perform parts of this work in workplaces and universities, institutions face a risk of certifying capabilities without reliably preserving the practice through which those capabilities are formed. The paper proposes three principles for formation-centred design: friction by design, process as both learning and evidence, and a care floor of guaranteed human attention. It concludes that valid judgement and capability formation are different questions that assessment can and should be designed to answer together.

Keywords: generative artificial intelligence, assessment reform, cognitive apprenticeship, desirable difficulties, evaluative judgement, adaptive capabilities, conceptual analysis

Introduction

Australian higher education has spent three years developing responses to generative artificial intelligence, with assessment reform at the centre. An expert resource commissioned by the Tertiary Education Quality and Standards Agency (TEQSA) shifted attention from detection towards assessment design (Lodge et al., 2023). A second commissioned resource described three implementation pathways (program-wide, unit/subject-level and hybrid), while the June 2026 adaptive-capabilities resource identified digital literacy, distributed cognition, hybrid metacognition and life-long learning as capabilities to be cultivated and assured (Lodge et al., 2025, 2026). These resources are expert advice rather than prescriptive TEQSA Guidance Notes, but they have helped structure sector discussion. The University of Sydney's two-lane approach distinguishes secured assessment of learning from open, AI-enabled assessment for learning (Liu & Bridgeman, 2023). Dawson et al. (2024) place validity ahead of cheating as the primary assessment concern: the central question is whether grades still support the inferences institutions make from them. Corbin et al. (2025) similarly argue that structural redesign of tasks and conditions is more consequential than discursive measures such as rules or declarations alone.

This reform program is serious and evidence-informed. It addresses a public question institutions must answer: can a degree still be trusted? Securing meaningful assessment points while opening other tasks to responsible AI use recognises both that graduates will work with AI and that credentials require defensible evidence. The two-lane model is therefore a useful organising response to the trust problem, even though it is not the only possible architecture.

This paper argues that the trust problem is not the whole problem. A valid judgement that a capability exists does not, by itself, explain how that capability was developed. Current reform increasingly acknowledges this distinction, particularly in its attention to learning processes and adaptive capabilities, but the practical conditions of formation remain vulnerable. The paper introduces the apprenticeship layer: the stratum of first drafts, summaries, reviews and first-pass analyses through which novices can acquire disciplinary judgement. Generative AI is especially capable of assisting with many of these codified, repeated tasks. Where such assistance becomes substitution rather than scaffolding, opportunities for practice may contract in firms and

universities alike. The resulting risk is not that institutions will necessarily certify non-existent capabilities, but that valid assessment may coexist with weaker formation pathways.

Approach and scope

This paper uses critical conceptual analysis rather than systematic review. Australia provides the deliberate policy boundary because the TEQSA-commissioned resources and the two-lane model offer an unusually explicit progression from assessment security to assurance of learning and adaptive capability. The argument reads that policy lineage against scholarship on assessment validity and process, cognitive apprenticeship, desirable difficulties, evaluative judgement and cognitive offloading, together with recent empirical studies of assisted performance, deskilling and non-formation. Sources were selected purposively for conceptual relevance and recency, so the synthesis is illustrative rather than exhaustive. The policy account is Australia-specific; the proposed mechanisms require testing in other systems and settings. As the paper uses published literature and public policy documents and reports no participants, personal data or primary data collection, institutional ethics approval was not required. The paper first examines the validity turn, then develops the apprenticeship-layer concept and proposes three principles for formation-centred design.

The validity turn in assessment reform

Many early institutional responses to large language models emphasised detection. Detection tools, however, could not carry the evidentiary burden placed on them, and assessment-security research had already warned against relying on post hoc detection alone (Dawson, 2021). The subsequent validity agenda reframed the problem. Dawson et al. (2024) ask whether the inferences drawn from assessment still hold, rather than treating student misconduct as the sole or primary issue. A grade supports a claim about learning; if production conditions no longer support that claim, the assessment problem remains regardless of intent.

The 2023 commissioned resource gave this agenda a programmatic form: assessment should support trustworthy judgement across programs, include security at meaningful points, attend to learning processes and involve students in reform (Lodge et al., 2023). The 2025 resource then described three pathways: program-wide reform, assurance within each unit/subject and a hybrid approach (Lodge et al., 2025). This programmatic lineage predates generative AI. van der Vleuten and Schuwirth (2005) argued that assessment quality depends on a coherent program of evidence rather than any single method. The two-lane approach operationalises a related distinction: one lane secures selected demonstrations of learning, while the other permits authentic, AI-enabled work intended to support learning and professional preparation (Liu & Bridgeman, 2023). Corbin et al. (2025) add that changes to task structure and conditions are likely to alter behaviour more reliably than instructions and declarations on their own.

Evaluative judgement, the capability to make decisions about the quality of one's own and others' work, is one constructive response (Tai et al., 2018). In GenAI-rich environments, students need opportunities to develop standards for quality and to judge whether particular outputs are fit for purpose (Bearman et al., 2024). The 2026 commissioned resource places this within four adaptive capabilities: digital literacy, distributed cognition, hybrid metacognition and life-long learning, underpinned by deep disciplinary knowledge (Lodge et al., 2026). This framing points beyond artefact production towards the regulation of learning and human-AI collaboration.

These developments show that validity and formation are neither identical nor competing. They are different questions within the same design: validity asks what inference the evidence warrants; formation asks what practice built the capability being inferred. A task can address both when attempts, feedback, revision and explanation are formative activities and also produce credible evidence. Winstone et al. (2026) propose Black Box Assessment, which assesses process, values error and revision, supports transparency and learner agency, and rewards trajectory rather than the polished product alone. The apprenticeship-layer concept complements that model by asking which constituent practices students must still perform, and with what support, for the trajectory to retain developmental value. The remainder therefore treats validity and formation as analytically distinguishable but interdependent in design.

The formation function of assessment

Biesta (2009) warned that an age of measurement can value what is readily measurable and that qualification is only one educational purpose. The warning remains relevant: a reform organised around trustworthy qualification claims can unintentionally under-specify the formative work that makes those claims possible.

Consider what a task can be doing before it becomes evidence. An essay may involve a fortnight inside a body of material: meeting confusion, making choices among sources, noticing what matters, feeling a paragraph resist and revising it until the argument holds. The submitted artefact is the residue of that work. Similar formative activity can occur in a lab report, problem set, case analysis or annotated bibliography. Assessment tasks are double-purpose instruments: measurement occasions on the surface and formation mechanisms underneath.

The learning-science literature helps specify the mechanism. Bjork and Bjork (2011, 2020) distinguish immediate performance under supported conditions from learning that persists when support is removed. Under appropriate conditions, spacing, interleaving, retrieval practice and generation can make practice feel harder while supporting retention and transfer. Learners can mistake fluency for learning and so may prefer conditions that improve immediate performance without producing equivalent durable learning.

Generative AI can provide unusually strong performance support, so a polished artefact may appear with less of the practice that previously produced it. Early evidence is consistent with this risk, although it comes from different populations, tasks and designs. In a field experiment with nearly 1,000 high-school mathematics students, an unrestricted GPT-4-based interface increased assisted practice performance by 48% but was followed by 17% lower performance than the control group on an unassisted examination; a guardrailed tutor largely mitigated that harm (Bastani et al., 2025). In a retrospective observational study at four Polish endoscopy centres, adenoma detection during non-AI-assisted colonoscopies fell from 28.4 to 22.4% after clinicians had begun using an AI detection system; the design identifies an association and cannot establish that AI exposure caused the decline (Budzyń et al., 2025). A cross-sectional mixed methods study of 666 participants also reported an association between more frequent AI use and lower critical thinking scores, statistically mediated by cognitive offloading, but its self-reported and non-longitudinal design does not support causal inference (Gerlich, 2025). Cognitive offloading itself is longstanding and often adaptive for immediate task performance; its consequences depend on what is offloaded, by whom and for what purpose (Risko & Gilbert, 2016).

Sector survey data indicate that pedagogical integration remains uneven. The Digital Education Council's 2026 survey reports 45,398 responses: 27,284 students and 18,114 faculty across 35 countries. Fifteen percent of students said AI was integrated into many courses, 43% into a few and 43% into none. Among students who had experienced course integration, 42% found it only somewhat helpful and 24% reported no clear learning value (Di Lullo et al., 2026). These self-reported survey results do not measure learning outcomes, but they suggest that adoption and institutional attention have outpaced consistently valuable course design.

The evidence base is young, heterogeneous and not yet capable of establishing long-term effects across higher education. The argument here is therefore a conceptual risk proposition, not a claim that non-formation is already general or inevitable. Where producing an artefact no longer requires sustained engagement with the underlying work, formation can become easier to bypass. A validity-centred architecture can detect the resulting capability at selected checkpoints, but validity alone cannot guarantee that the developmental process occurred.

The apprenticeship layer

How does cognitive expertise develop and renew itself? Several research traditions answer through forms of apprenticeship. Collins et al. (1989) describe cognitive apprenticeship through modelling, coaching, scaffolding and fading. Lave and Wenger (1991) emphasise legitimate peripheral participation: newcomers undertake real but bounded work within a community of practice. Dreyfus and Dreyfus (1986) emphasise the accumulation of situated experience in movement from novice rule-following towards expertise. Ericsson et al. (1993) frame deliberate practice as effortful, feedback-rich activity near the edge of current capability. Polanyi (1966) helps

explain why some expert knowledge cannot be fully transmitted propositionally and must be developed through participation and correction.

Call the developmental space in which these practices occur the apprenticeship layer: an analytic term for supervised, corrigible and relatively low-stakes cognitive work whose immediate output value may be modest but whose developmental value can be substantial. In firms it can include document review, first-pass research, tier-one support and draft memoranda; in universities, essays, lab reports, problem sets and literature reviews can serve analogous functions. University tasks remain proxies for participation, borrowing apprenticeship's developmental logic without reproducing the full community conditions of legitimate peripheral participation. From a narrow efficiency perspective the layer may look expendable. From a formation perspective it can contain a substantial part of the curriculum.

Generative AI is especially capable of assisting with codified and repeated components of this layer, although exposure and effects vary by occupation and task. In a working paper revised in 2026 using ADP payroll data through June 2026, Brynjolfsson et al. (2026) report that employment among workers aged 22–25 in the most AI-exposed occupations stood 19% below where it would have been had it kept pace with their less-exposed peers. The authors caution that other changes in the United States labour market may contribute, and the result does not establish a general causal effect. The narrower structural concern is therefore not mass displacement as such, but whether some tasks from which novices historically learned are being reduced, reorganised or delegated before equivalent developmental opportunities are designed.

Two possible failure modes should be distinguished. One is deskilling: a formed capability is exercised less and may deteriorate. The endoscopy study is consistent with that possibility but does not prove it. The second is non-formation: novices receive fewer opportunities to build a capability in the first place because formative tasks are automated or heavily assisted. Non-formation is harder to observe because there is no prior individual baseline, and it could create succession risks if experienced workers retain responsibility while novices lose access to the work through which expertise was previously developed. These are testable risks, not established sector-wide outcomes.

Higher education often discusses this as an employability issue, but it is also an assessment-design issue. University tasks can function as a pre-professional apprenticeship layer: an essay may rehearse the reasoning of a junior memorandum; a lab report may approximate first-pass analysis; repeated problems can build pattern recognition and judgement. When students delegate substantial portions of this work to AI, opportunities for practice may narrow inside the institution itself. An open lane therefore needs explicit formation design rather than an assumption that authenticity and tool use will automatically produce learning.

Certifying what we no longer form

Validity and formation also operate over different timescales. A secure assessment creates a defensible observation at a point in time; formation accumulates across attempts, feedback, revision and transfer between such points. Process-based assessment can bridge these timescales, but process evidence is not self-validating. Drafts and version histories can record activity without showing who exercised the relevant judgement. The design problem is to identify which processes must remain learner-performed, which may be AI-supported, and how the resulting trajectory will both develop capability and warrant the final inference.

None of this argues against the validity turn. Unverifiable credentials help nobody, and structural redesign is preferable to relying on detection alone. Nor is formation absent from current reform: the 2025 and 2026 commissioned resources explicitly emphasise the process of learning, developmental progression, adaptive capabilities and desirable difficulties (Lodge et al., 2025, 2026). The remaining issue is an observability asymmetry in implementation: secure assessment can be timetabled and audited, while sustained formation is slower, relational and harder to represent in institutional systems. This asymmetry can pull implementation towards security even when policy also values formation. The three principles below address that drift: friction locates essential practice in task design, process evidence makes the trajectory visible, and the care floor makes the provision of required human support auditable.

Formation-centred design

If this analysis is plausible, the next phase of reform should treat valid judgement and capability formation as two questions within an integrated design. At program and task level, designers should specify what evidence will warrant a claim of capability and what repeated, supported practice will build it. The balance between protected practice and AI augmentation is context-dependent. It should vary with the capability and its consequences, students' prior expertise, discipline and professional standards, task structure, cohort size and mode, accessibility, staffing and the availability of valid checkpoints. Students should be partners in setting these formation requirements, including identifying which practices need protection and what support makes them accessible and workable (Lodge et al., 2023). The three principles below are therefore adaptable design tests, not universal prescriptions.

Friction by design

The first principle is deliberate, accessible friction. Generation before consultation, retrieval checkpoints, spacing and interleaving, and staged tasks in which an initial unaided attempt precedes AI-supported revision may support durable learning when aligned with the capability being developed. The 2026 commissioned resource reaches a related conclusion through desirable difficulties and hybrid metacognition: scaffolds should support learners to take increasing responsibility for regulation rather than defaulting that regulation to AI (Lodge et al., 2026). For capabilities not yet established, a formation-before-augmentation sequence may therefore be warranted. The novice–expert distinction is a heuristic rather than a universal rule: assistance that helps an expert may substitute for practice a novice still needs. Two boundaries matter. A difficulty is desirable only when it is surmountable and supports encoding or transfer (Bjork & Bjork, 2011); difficulty that excludes or merely consumes time is not pedagogy. Cognitive friction must also be distinguished from bureaucratic and access barriers. Any added practice requirement must be paid for through curriculum redesign, not simply layered onto unchanged workloads.

Process as both learning and evidence

The second principle is to assess a trajectory rather than only a final artefact. This is where validity and formation most clearly converge. Black Box Assessment operationalises the idea through attention to process, error and revision, transparency and agency, and trajectory (Winstone et al., 2026). Staged drafts, oral explanations, demonstrations and portfolios can require learners to attempt, justify and revise while giving assessors triangulated evidence of capability. The form should be disciplinary: oral defence may suit conceptual reasoning, observed performance may be essential in clinical or laboratory work, and a portfolio may better show progression in design or writing. Process artefacts are not inherently authentic; drafts and reflections can also be generated or curated strategically. The safeguard is dialogic and longitudinal evidence, including live interaction with people who know the student's trajectory, rather than surveillance or a presumption that any single version history verifies authorship.

The care floor

Formation is often relational. The apprenticeship traditions reviewed above include a person or community that notices performance and responds: a coach, an experienced participant or a teacher. Care ethics similarly foregrounds attentiveness and responsiveness in educational relationships (Noddings, 2013; Tronto, 1993). Winstone et al. (2026) likewise treat feedback as a relational practice of care and place AI within, rather than in place of, that relationship. This paper therefore proposes a care floor: a specified minimum of sustained, responsive human attention to each learner's developing capability. It is a normative design condition, not an empirically validated universal threshold. At scale, the floor need not mean individual tutoring on every draft. It can combine stable tutorial or studio groups and a named educator or teaching team; brief live checkpoints on developing work; targeted follow-up when process evidence shows stalled progress or misunderstanding; and a staffed escalation route when peer or automated support is insufficient. Institutions can audit whether this floor is real through workload allocation, group size and continuity, completion of dialogic checkpoints and access to human escalation. Large cohorts change its form; they do not remove the obligation.

Three objections

The strongest objection holds that the capabilities themselves have changed: composing from scratch may be going the way of long division, and curricula should instead form AI-era capabilities such as orchestration, verification and judgement of machine output. The premise is partly right, but it does not eliminate the formation problem. Evaluative judgement likely depends partly on disciplinary knowledge and prior experience producing, revising or closely analysing the kind of work being judged (Bearman et al., 2024; Tai et al., 2018). A graduate with little generative experience may have a weaker basis for recognising characteristic failure modes, while a graduate who only produces without learning to evaluate is also underprepared. Reframing capability around orchestration and judgement changes the required practice rather than removing it.

A second objection invokes calculators, spellcheckers and search engines, which also prompted concerns about learning before education adapted. The analogy cautions against assuming inevitable harm. It also points to sequencing and task analysis: a calculator automates some operations while leaving problem formulation, method selection and interpretation to the learner, whereas generative systems can contribute to a broader portion of many tasks. The design question is less whether a tool is allowed than which constituent capabilities students should first or concurrently practise so that augmentation does not replace the judgement it is meant to extend.

The third objection concerns equity: added friction can burden students with the least time and material slack, while AI can remove barriers for disabled learners, carers and students in paid work. Those gains should be protected. Difficulty that excludes without supporting learning is a barrier, not a desirable difficulty. At the same time, access to protected practice and responsive human attention could itself become stratified if only well-resourced students receive it. A care floor is proposed as an equity safeguard: formation requirements should be accessible, supported and resourced, rather than imposed as additional uncompensated labour or reserved as a premium service.

Conclusion

Assessment reform has addressed an urgent question: can a degree still support trustworthy claims about learning? This paper adds a second: are the capabilities described by the degree being developed through sufficiently rich practice? The apprenticeship layer names a potential site of vulnerability. Entry-level cognitive work can carry developmental value as well as output value, and generative AI can increasingly perform parts of it. If that work contracts in employment, universities may become one of the few structured settings in which many learners can still undertake supervised cognitive practice. This makes assessment design not only a certification issue but also part of the infrastructure through which expertise is renewed.

Formation is slow, relational and resource-intensive, and its effects may be difficult to detect in conventional metrics. The 2026 adaptive-capabilities framework is compatible with the apprenticeship-layer argument: it identifies capabilities whose development depends on learning processes, disciplinary knowledge and human agency. The claim here is narrower than inevitability. Validity and formation are not alternatives: process-focused design can use the same attempts, feedback and revisions to develop capability and support a defensible inference. This convergence must be designed and resourced; it does not follow automatically from securing a final assessment. The next phase of reform should therefore be evaluated both by the trustworthiness of its judgements and by the quality of the practice it secures.

Acknowledgements: Generative AI use

Generative AI tools were used as writing and review aids in preparing this conceptual paper. Anthropic's Claude Fable 5, used within a Claude Project knowledge base containing the author's prior research and writing, assisted with ideation, synthesis, drafting and revision; OpenAI's ChatGPT, using GPT-5.6 Sol, assisted with critical review, citation checking and editing. The paper's central argument, interpretations and conclusions are the author's own. The author reviewed all AI-assisted material and accepts full responsibility for the final manuscript.

References

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

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

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability, 21*(1), 33–46. https://doi.org/10.1007/s11092-008-9064-9

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition, 9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

Brynjolfsson, E., Chandar, B., & Chen, R. (2026). *Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence* [Working paper, revised August 2026]. Stanford Digital Economy Lab. https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine/

Budzyń, K., Romańczyk, M., Kitala, D., Kołodziej, P., Bugajski, M., Adami, H. O., Blom, J., Buszkiewicz, M., Halvorsen, N., Hassan, C., Romańczyk, T., Holme, Ø., Jarus, K., Fielding, S., Kunar, M., Pellise, M., Pilonis, N., Kamiński, M. F., Kalager, M., ... Mori, Y. (2025). Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: A multicentre, observational study. *The Lancet Gastroenterology & Hepatology, 10*(10), 896–903. https://doi.org/10.1016/S2468-1253(25)00133-5

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315044408-14

Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education, 50*(7), 1087–1097. https://doi.org/10.1080/02602938.2025.2503964

Dawson, P. (2021). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge. https://doi.org/10.4324/9780429324178

Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education, 49*(7), 1005–1016. https://doi.org/10.1080/02602938.2024.2386662

Di Lullo, A., Bielik, D., Rong, H., Mavlonova, M., Chun, C., & Subhan, F. (2026). *AI in higher education global survey 2026*. Digital Education Council. https://www.digitaleducationcouncil.com/resource-library-items/ai-in-higher-education-global-survey-2026

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Free Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), Article 6. https://doi.org/10.3390/soc15010006

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Liu, D., & Bridgeman, A. (2023, July 12). *What to do about assessments if we can't out-design or out-run AI?* Teaching@Sydney. https://educational-innovation.sydney.edu.au/teaching@sydney/what-to-do-about-assessments-if-we-cant-out-design-or-out-run-ai/

Lodge, J. M., Bearman, M., Dawson, P., Gniel, H., Harper, R., Liu, D., McLean, J., & Ucnik, L. (2025). *Enacting assessment reform in a time of artificial intelligence*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2025-09/enacting-assessment-reform-in-a-time-of-artificial-intelligence.pdf

Lodge, J. M., de Barba, P., Ainscough, L., Brazil, J. R., Broadbent, J., Ebbert, D., Frankland, S., Gabriel, F., Gašević, D., Hennicke, T., Lim, L.-A., Male, S. A., Mirriahi, N., Oliveira, E. A., Pacitti, H., Raković, M., Russell, J., Taylor-Griffiths, D., & Yang, S. (2026). *Assuring quality learning in a gen AI-integrated future: The role of*

adaptive capabilities. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/assuring-quality-learning-gen-ai-integrated-future-role-adaptive-capabilities

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/guides-resources/resources/corporate-publications/assessment-reform-age-artificial-intelligence

Noddings, N. (2013). *Caring: A relational approach to ethics and moral education* (2nd ed.). University of California Press. https://doi.org/10.1525/9780520957343

Polanyi, M. (1966). *The tacit dimension*. Doubleday.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

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

Tronto, J. C. (1993). *Moral boundaries: A political argument for an ethic of care*. Routledge. https://doi.org/10.4324/9781003070672

van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. *Medical Education, 39*(3), 309–317. https://doi.org/10.1111/j.1365-2929.2005.02094.x

Winstone, N. E., Gravett, K., Bearman, M., Noble, C., Jensen, L. X., Jones, A., & Nicola-Richmond, K. (2026). The care-full craft of feedback in an age of generative AI. *Assessment & Evaluation in Higher Education, 51*(5), 911–927. https://doi.org/10.1080/02602938.2026.2643333

Winstone, N. E., Gravett, K., & Elkington, S. (2026). Black box assessment: Rethinking integrity and learning for a time of generative AI. *Assessment & Evaluation in Higher Education*. Advance online publication. https://doi.org/10.1080/02602938.2026.2661365

Iacono, C. (2026). The apprenticeship layer: What assessment reform cannot secure on its own. 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. 222–229). https://doi.org/10.65106/apubs.2026.3340

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.

© Iacono, C. 2026