

Assessment Learning and Institutional Capacity

An evidence companion to The AI Crisis in Higher Education

Research completed 4 October 2026

This resource accompanies Mark Carrigan's *The AI Crisis in Higher Education: A Strategic Proposal*. It brings together research and public accounts of practice that give substance to the book's agendas of adaptation, integration and relational capacity. Its main finding is that these agendas can reinforce one another when assessment design, disciplinary learning and participation in decisions are connected. The cases offer workable starting points; their educational effects are established unevenly.

The companion is for students, academic staff, learning designers, professional services colleagues and institutional leaders, particularly in humanities and social science settings. It can be read independently of the book. References to its chapters use the printed pages of the supplied edition. Institutional examples come from public sources. This is an analytical resource for public circulation, rather than a statement of any institution's policy or plans. [1]

Executive summary

The book's central proposal has a substantial practical basis. Public cases show universities securing selected evidence of learning, developing purposeful uses of language models, and involving students and staff in the work of redesign. The evidence supports treating these as connected responsibilities. It does not yet establish that one institutional model consistently produces better graduate capabilities than its alternatives.

Adaptation is most convincing when it starts with what a degree claims its graduates can do. Sydney's assessment framework makes this concrete through programme mapping and a combination of secure and open tasks. Interactive orals offer another example: a tourism capstone linked a short conversation to sustained reflection and a specified programme outcome. Both direct attention to demonstrated capability. Security, authorship and assessment validity remain different questions. An authentic or personalised task is not automatically resistant to outsourcing. [5,7,9]

Integration depends on the cognitive work that students undertake. Sydney economics assessments require critique of generated content against disciplinary knowledge and scholarship. Translation research documents students revising their own initial work with AI feedback. Creative education at UAL makes room for critical use, refusal and informed artistic judgement. These examples substantiate the book's account of integration as an educational project, extending beyond learning how to operate a product. [21,22,26,27]

Learning experiments show why design matters. A purpose built university physics tutor improved immediate learning compared with active classroom instruction. A school mathematics experiment found that unrestricted assistance could improve practice performance while weakening subsequent unaided performance; guarded tutoring largely avoided that harm. A university study also reports poorer retention after 45 days with unrestricted ChatGPT support. These results warrant testing specific designs, with attention to retained learning and transfer, rather than a general verdict on AI. [19,20,46]

Relational capacity is visible in projects with a clear route from participation to change. Sydney student partners helped revise finance guidance following pilot feedback. UCC funded student and staff pairs to examine assessments and produce disciplinary resources. UCL and Georgetown cases made assessment evaluation and policy drafting objects of joint inquiry. Such work gives the third agenda a mechanism: people make ambiguities discussable and turn experience into design decisions. Evidence of institutional effects is much thinner than evidence of participation and resource production. [32,33,34,35]

The most defensible strategic approach is selective and revisable. Identify capabilities requiring independent demonstration; design supported encounters with models that preserve meaningful student decisions; resource the relationships through which both can change. Central coordination can improve access and clarity, while disciplines retain responsibility for educational purposes. Evaluation should examine learning, fairness, workload and continued viability. Procurement, attendance at training, declaration forms and rising marks each provide incomplete evidence of success.

Reading the book through the evidence

The book begins from language models becoming routine without settled norms of appropriate use. Chapter 1 connects adoption to the conditions under which people study and work. Chapter 2 examines trust and the tensions between adaptation and integration. Chapter 3 asks whether educational commitments will remain viable under changing commercial conditions. Chapter 4 develops the strategic response, including student agency and the relationships needed for institutional learning. [1, chapters 1 to 4]

Table 1 Connections to the book

Book chapter	Question for the companion	Examples to read alongside it
1 Situation Pages 1 to 11	How do conditions and incentives shape use?	HEPI and Jisc student evidence; finance pilot feedback. [40,41,32]
2 Crisis Pages 12 to 24	What supports assurance and justified trust?	Sydney programme assessment; detection research; disclosure study. [5,12,17]
3 Integration Pages 25 to 38	What capability and dependence could a commitment create?	Learning experiments and specific enterprise evidence. [19,20,29,31,46]
4 Response Pages 39 to 53	How can agency and collective judgement be developed?	Economics, creative practice, partnerships and deliberative forums. [21,26,33,35,39]

The three agendas address different questions. Adaptation asks what justifies an inference about a student's capabilities. Integration asks what students should learn about and through models, including when to work without them. Relational capacity asks how a community acquires the expertise, relationships and authority to revise either response. This distinction prevents a permissions policy, a licence purchase or a professional development event from standing in for all three.

The book calls its proposed stance strategic under integration. Here this means active educational work combined with selective, revisable commitments: securing consequential evidence, developing purposeful capabilities and maintaining the capacity to reconsider a design. It is more demanding than postponing decisions or collecting unrelated pilots. TEQSA's 2023 assessment reform guidance independently joins preparation for a society using models with trustworthy judgements about learning, providing a sector counterpart to that connected agenda. [1, chapter 4; 4]

Four evidence descriptions are used throughout. Evaluated means some systematic assessment of a practice, tool or experience has been conducted. Reported means a public implementation account documents what happened. Advocated identifies a proposal or guidance. Claimed identifies an assertion whose supporting evaluation has not been established, including vendor descriptions of benefits. These describe the basis of a claim, not its methodological strength. A qualitative case can illuminate experience convincingly without estimating causal effects; a poorly controlled evaluation can provide weak evidence of effectiveness.

Confidence is therefore stated in relation to a particular inference. A documented change to an assessment may have good implementation evidence but little evidence of retained learning. A randomised experiment may support a causal conclusion for its specific intervention but have limited disciplinary transferability. The review treats repeated accounts of a single initiative as one case, and distinguishes an institution's advice from a mandatory policy. Practical designs proposed below are the companion's analytical options unless a public case is explicitly identified.

Current UK survey evidence makes these distinctions useful. HEPI's December 2025 survey of 1054 full time undergraduates reports 95% using AI in at least one way and 94% using it to help with assessed work. Only 12% report directly including generated text. The survey, sponsored by Kortext, also records mixed experiences and uneven support. These are self reported activities and perceptions, not estimates of misconduct or independent learning. Jisc's 2025 student evidence similarly emphasises clarity, fair access, participation and continued human contact. [40,41]

Adaptation

Assessing the capability rather than only the submission

The book's adaptation agenda responds to a weakened inference from the quality of submitted work to the capabilities of its producer. Dawson and colleagues' conceptual argument provides a useful foundation: assessment validity concerns whether evidence supports the interpretations and decisions made from it. Catching a breach does not repair an assessment that measures the wrong thing, while eliminating a particular route to cheating does not establish validity. This shifts the design question towards the capability, evidence and decision that a task is meant to connect. [1, chapter 2, pages 12 to 24; 2]

Scarfe and colleagues show the problem in a real assessment system. In five undergraduate psychology modules at Reading, markers did not identify 94% of the AI submissions introduced by researchers, and generated answers generally performed well. The study used GPT 4 and ordinary examination processes, providing strong evidence that those unsupervised tasks could admit convincing model output. It establishes neither how often students cheated nor the vulnerability of every discipline or current assessment. Its value is a demonstrated failure of an inference that otherwise looked plausible. [3]

The resulting practical distinction is between security, authorship and validity. Security concerns the conditions under which assistance can be controlled. Authorship concerns how work was produced. Validity concerns what the evidence warrants about learning. A student can genuinely author a weak essay that fails to demonstrate the intended outcome. A well supervised task can primarily assess rapid recall when the degree claims considered historical interpretation. A polished collaborative product can validly assess collaboration while leaving individual capabilities uncertain. Each problem requires its own design response.

Sydney as a programme response

Sydney's framework separates secure assessment used to assure learning from open assessment that supports learning with available tools. Its public account describes programme mapping, coordinator verification and disciplinary workshops, with learning design support. Open tasks do not permit restrictions on model use within that framework. Secure tasks can allow specified tools when their conditions are controlled. This is a documented institutional architecture, not simply a return to examinations. Public materials do not yet establish its effect on graduate capability or total workload. [5]

The accompanying FAQ clarifies that every unit does not need its own secure task. Assurance can be placed at relevant programme points, including capstones, while other units contribute open learning activities. This is a specific policy design rather than a universal meaning of 'two lanes'. Its absence of an intermediate restricted category also differs from many UK schemes. The implementation lesson is to connect the local task to a degree pathway; copying the lane labels alone would miss that work. [6]

A smaller implemented example makes the approach easier to grasp. Sydney's Graduate Certificate in Educational Studies in Higher Education used four modules in 2024, combining open graded work with interactive oral hurdles in two modules. Bridgeman, Liu and Weeks explicitly distinguish a programme approach from the more intensive programmatic assessment associated with medical education. The latter ordinarily assembles repeated evidence of development over time. Programme mapping can connect existing assessments without importing that entire model. The certificate example establishes an implemented combination, rather than proof of better learning or of feasibility across all flexible degrees. [7]

TEQSA's 2025 guidance proposes three pathways: reform across a degree, assurance within units, and a combination across selected groups of units. It identifies substantial coordination and transition demands, especially in flexible degrees. Its hybrid pathway is useful where complete redesign is not immediately feasible. This is evidence informed guidance, not a trial showing which pathway works best. Its relevance outside Australia lies in the design questions, rather than in transferring Australian regulatory requirements to UK institutions. [8]

For a humanities programme, the essential first task is to name a manageable set of consequential capabilities. These might include interpreting primary material, constructing an argument from scholarship, choosing and explaining a method, or defending a reasoned judgement. The team can then ask where each student demonstrates them, at what level, under what conditions, and with what consequences if the evidence is insufficient. Optional modules, transfers and interdisciplinary routes matter: an assurance point available to some students cannot warrant a claim about every graduate.

Interactive orals with an educational purpose

Yung, Moyle and Yang studied an Australian tourism capstone delivered in 2021. Students maintained reflective journals and completed an online assessment consisting of a five minute presentation and five minute conversation. Analysis covered 90 oral transcripts, 90 journals and three assessors' reflections. The study illuminates reflection and professional identity, with clear alignment to a programme outcome. It has reasonable qualitative credibility for that experience, but no comparison group or test against contemporary model assistance. The conversation's pedagogical purpose is stronger evidence than its title as an AI mitigation. [9]

Earlier business education work by Sotiriadou and colleagues also associates scaffolded interactive orals with authenticity, perceived skill development and integrity. The abstract and Griffith's public implementation account describe preparation and continuing refinement. Because the study predates current language models, it supports the educational lineage of the format rather than establishing present security. This matters to the book's argument: an oral can be a constructive assessment redesign when it elicits a relevant capability and students have opportunities to practise it. [10]

An oral should assess the capability that the degree needs, with published criteria and a moderated approach to questioning. Where rapid verbal performance is incidental to the outcome, anxiety, language proficiency and disability may distort the inference. Preparation, accessible arrangements and equivalent ways of demonstrating the same outcome need design attention. Those arrangements must be considered alongside security rather than treated as exceptions discovered during delivery. A planned oral assessment and a viva used to investigate suspected misconduct also have different purposes and procedures.

Workload should include preparation, scheduling, missed appointments, recording where appropriate, marking, moderation and appeals. Ten minutes of conversation for 300 students requires 50 hours before those other tasks. This arithmetic is an illustration, not a staffing estimate or research finding. A brief oral attached to every essay could therefore consume more capacity than a programme team intended. Selecting a few consequential assurance points can be proportionate, provided their coverage and decision rules are defensible.

Redesign and testing

PANDORA gives a disciplinary example of collective assessment review. At the fully online Spanish institution UNIR, 29 lecturers applied the rubric to 151 translation and interpreting assignments, identifying vulnerabilities and guiding redesign. The project demonstrates a practical way to organise professional judgement across a degree. The field test concerns the rubric and lecturers' analysis; it does not prove that revised tasks resist adversarial model use or improve independent capability. It is useful as a conversation and prioritisation instrument, with that distinction made explicit. [11]

Staging, situated questions, portfolios and reflective statements can make learning more visible. They can also be generated or reconstructed. A useful review therefore asks two questions separately: does the redesign elicit better evidence of the intended learning, and has the relevant outsourcing route actually become harder? An account of process warrants more confidence when it can be connected to observed discussions, material decisions and revisions during teaching. Requiring a longer narrative of process without those connections can add labour while leaving the inference unchanged.

The UCC partnership described under relational capacity offers a documented example of staff and students trying assessments with ChatGPT. For a reproducible local test, the companion proposes recording the task and marking criteria, date, model, access tier, allowed inputs, time available and assistance used in prompting. Outputs should be judged against the actual criteria, preferably without markers knowing their origin. A failure to produce a good answer in one attempt offers little assurance when a student could iterate or supply additional material. Tests must be repeated when task design or relevant tool capabilities change. [35]

Contemporary wearable AI adds a further question. Corbin, Sharpe and Dawson argue that increasingly unobtrusive devices could weaken security based on excluding technology. The retrieved article passages are a conceptual risk argument, rather than an attack study or prevalence estimate. No verified agent or wearable benchmark in this review establishes that all invigilated formats have failed. Security decisions

should be revisable and include accessible technologies in their design; claims of inevitable defeat would go beyond the evidence inspected here. [45]

Detection process records and fair interpretation

Weber Wulff and colleagues evaluated 14 text detection tools in 2023 and found substantial reliability problems, including sensitivity to altered text. Liang and colleagues separately reported bias affecting writers using English as an additional language. These studies justify caution about punitive inference from detector output. They do not supply accuracy estimates for every product in October 2026. Any proposed evidential use requires validation for the relevant population, text and current version, alongside the institution's published procedures. [12,13]

UCL's public guidance states that it does not use GenAI detectors when marking and describes opportunities to clarify concerns with students. Its categories page explains reporting suspected misconduct through a separate process and marking submissions anonymously on merit before any necessary investigation. It also says that use without clear module guidance will not be penalised. This is one documented institutional response; it should not be presented as another university's rule. [14,15]

Briffa's postgraduate International Politics case examines Cadmus as a platform for making writing processes visible. With 19 survey responses before and seven afterwards, it provides situated evidence about transparency, student interpretation and working practices, including work outside the platform. The small, uneven response sets limit comparison. The case illustrates why a process record is not a complete account of intellectual work: reading, planning, assistive support and discussion can happen elsewhere. Claims that such a platform proves authorship exceed what this case establishes. [16]

A prompt fragment, shared chat link or sudden change in style may justify clarification, but each needs contextual interpretation. Evidence that a tool was used is not automatically evidence that a prohibited function was used. Marking should address the published educational criteria; alleged breaches should follow the applicable misconduct procedure. An unsatisfactory explanation of a particular concept may demonstrate a gap in learning without proving how the submitted text was produced. Treating those findings as interchangeable risks both unfairness and weak assessment.

Integration

Agency as a design question

The book's integration agenda encompasses learning with models and learning about their limits and social consequences. Its discussion of Jackson distinguishes activity without models, human directed use, model directed activity with human intervention, and largely delegated activity. 'Shifting left' describes a movement towards greater human control and understanding. Jackson's original paper is a peer reviewed conceptual framework with pedagogical examples, not an empirically validated scale showing that a particular position guarantees learning. It is most useful for asking who undertakes the consequential thinking during an activity. [1, chapter 4, pages 39 to 45; 18]

An initial generated answer can become material for student analysis. Conversely, a formally permitted tutoring interaction can become repeated answer acquisition. The relevant unit of analysis is therefore the pattern of activity across time. Students need to make choices, justify them and demonstrate what they can subsequently do. Reflection can support this work, but a declaration that the model preserved agency does

not itself measure agency. Capability with a tool and capability without it may both be legitimate outcomes; they require distinguishable evidence.

What learning experiments establish

Bastani and colleagues' field experiment involved nearly 1000 secondary school mathematics students in Turkey, comparing no model assistance, GPT Base and a guarded GPT Tutor using teacher supplied hints and solutions. Both assisted designs improved practice performance; the unrestricted design subsequently reduced unaided performance by 17% relative to the comparison group. The guarded tutor largely removed that disadvantage without establishing a general improvement in unaided learning. This is strong evidence for an effect of these designs in that setting, and indirect evidence for university teaching. It concerns short term assessment, not established long term atrophy. [19]

Kestin and colleagues provide a positive university comparison. A randomised crossover study with 194 undergraduate physics students compared a purpose built tutor at home with classroom active learning across two lessons. The tutor condition produced greater immediate learning in less time. It combined model interaction with a designed lesson, pacing and instructional materials, so the result concerns a pedagogical package rather than a general chatbot. The reported outcomes do not establish longer term retention or transfer to humanities reasoning. The retrieved author abstract and publisher passages support this bounded conclusion; full page retrieval was unsuccessful. [20]

Barcaui's published university study recruited 120 undergraduates learning AI concepts and reports lower surprise test performance after 45 days for students assigned unrestricted ChatGPT study support, compared with traditional methods. Reported scores were 57.5% and 68.5%. The accessible publisher abstract reports $t(83)$, which should not be treated as evidence that all 120 contributed to that comparison. Full methods were unavailable in this review, preventing a sufficient check of attrition and intervention equivalence. The result is relevant delayed retention evidence requiring careful replication, rather than a universal estimate of harm. [46]

An accepted 2026 article by Dong reports a positive 14 day retention result from an eight week basketball teaching trial with 240 undergraduates. It compares a structured written GenAI feedback package with conventional oral instructor feedback. Only the accepted abstract was available; the final formatted article was forthcoming. Its motor learning outcome and bundled intervention make it indirect evidence for this companion's intellectual disciplines. It nevertheless cautions against assuming that assisted instruction necessarily weakens unaided retention. [47]

Taken together, these studies substantiate the book's concern with learning design while qualifying any simple opposition between models and independent learning. Their comparisons involve different students, tasks and packages. Averaging them into a single judgement about 'AI in education' would obscure the question an educator needs answered. The appropriate test is whether the particular activity improves the desired capability, whether that effect persists, and whether its cost and conditions are acceptable.

TEQSA's June 2026 adaptive capabilities guidance extends this design question beyond operating today's tools. It identifies digital literacy, distributed cognition, hybrid metacognition and lifelong learning, grounded in disciplinary knowledge. It advocates attention to learning processes, agency and collaborative innovation. This supports the direction of the book's agenda as contemporary sector thinking; the guidance is not an outcome evaluation. Its unnamed faculty examples are illustrative scenarios, and are not counted here as implemented cases. [28]

Economics and law

Sydney economics provides a reported design across two levels. A first year unit with approximately 1000 students per semester required a ChatGPT conversation followed by critique, and later added tutorial practice and stronger links to disciplinary knowledge. A masters capstone with around 200 students allowed model use in literature review work while requiring critical evaluation and comparison with scholarly sources. The public account specifies what students must contribute, making it useful for essay and research teaching. It does not report a controlled test of later independent reasoning or retained learning. [21]

For law, Choi, Monahan and Schwarcz randomly assigned students to realistic legal tasks with or without GPT 4 and blind graded outputs. The published abstract reports substantial time savings and smaller, inconsistent improvements in quality, with uneven benefits across participants. It supports a claim about assisted task performance and productivity, not about learning law or subsequent independent legal judgement. The full article could not be retrieved here, so detailed sample and intervention claims are omitted. [42]

A disciplinary curriculum can make verification part of the work itself. In economics, students might compare a generated causal explanation with a model's assumptions and a source's empirical design. In law, they might check cited authorities, jurisdiction and applicability, and explain the reasons for retaining or rejecting an argument. These are proposed designs, not findings about the cases above. Their educational value depends on students having enough disciplinary knowledge to perform the checks and on assessment actually valuing that reasoning.

Translation and language learning

Xu, Su and Liu studied 29 masters students translating a Chinese text into English and revising with ChatGPT feedback through a purpose built platform. The initial translation was completed without machine translation or online resources, with recorded revision activity, surveys and four interviews. Engagement varied, including different degrees of verification and acceptance of advice. This is a concrete example of an initial unaided attempt followed by model support, and good situated evidence about revision processes. A single session without a comparison group cannot establish durable translation learning. [22]

Translation exposes why a general distinction between presentation and substance is insufficient. Changing tense, register or terminology can alter meaning; translating an argument can perform the capability a language course is meant to develop. In another discipline, translation may instead provide access to material while the assessed capability is historical or sociological judgement. The task must state which of these situations applies. Treating every fluent correction as innocuous, or every language support function as misconduct, would obscure the educational purpose.

Language learning can combine model conversation with independent production, listening, reading and interaction with people. An educator might ask students to compare a model's register choices against a real communicative setting, then produce a new response without that assistance. The proposed evaluation would examine accuracy, appropriateness and transfer to a fresh context. This is a possible design informed by the cases; this review does not establish that a conversational interface can replace sustained human interaction or subject teaching.

Creative practice and critical literacy

UAL's MA Photography and Digital Practice publicly describes assessment through a written positioning statement, research video and portfolio, with peer review and tutor dialogue. Its account asks students to substantiate use, refusal or combinations of the two, while examining technology through critical scholarship. The author explicitly recognises that generative systems could produce the individual artefacts. The educational claim rests on sustained and coherent decisions across practice and process. This is a reported course design, with no comparative outcome evaluation on the page. It gives integration a critical and disciplinary meaning while retaining the problem of assurance. [26]

A London College of Fashion unit offers a complementary example. The six week 'Decolonising in the Age of AI' project combined critical inquiry, workshops, evolving ethical manifestos and creative making, with freedom to use or avoid the software. It addressed labour, environment, bias and colonial relations, culminating in an exhibition. Early student reports of awareness and confidence are useful experience evidence, but the public account provides no sampling or evaluation protocol. The case substantiates teaching about AI as part of creative education without treating adoption as the desired position. [27]

These examples suggest possibilities for literature, history and philosophy as well as studio subjects. Close reading can examine a model's interpretation against passages and scholarly disagreement. A history exercise can contrast a generated account with the provenance and limits of a primary source. Philosophy can ask students to reconstruct and defend an argument in dialogue before evaluating generated objections. In creative writing, students might critique a supplied model text without outsourcing their own composition. These are analytical options, with no claim of evaluated effectiveness attached.

For architecture, music and drama, the same questions concern situated judgement, embodied work and collaboration, although a text model cannot stand in for the whole practice. A design rationale, rehearsal discussion or explanation of a compositional choice can supplement the artefact or performance. An assessment should preserve the capabilities for which the discipline takes responsibility, while specifying which generative tools and functions are relevant. This review found less directly inspected outcome evidence in those subjects; the absence of such evidence here should not be read as a finding that useful practice does not exist.

Permissions and educational purposes

The British University Vietnam AI Assessment Scale pilot describes implementation alongside changes to permissions and teaching. It reports fewer AI related misconduct cases and higher grades and pass rates in comparisons across semesters. Those comparisons combine policy, pedagogy and changing access; permitting more uses can itself reduce recorded breaches. The reported grade increase of 5.9% is therefore not a causal estimate of learning attributable to the scale. The case does demonstrate institutional work to clarify boundaries and normalise discussion, and gives a useful example of a permissions framework accompanying redesign. [23]

UCL and the Open University publicly distinguish restricted or prohibited use, assistance and an integral or required role for models. Their schemes are not identical to Sydney's lanes or to the illustrative categories used here. UQ guidance hosted by TEQSA advises requiring and assessing responsible AI use where it is an explicit learning outcome, while ensuring that optional use does not determine how work is judged. It also separates ordinary acknowledgement from assessment of an AI capability. These are documented approaches and guidance, not evidence that permission categories alone improve learning. [14,24,25]

Table 2 Illustrative categories and cognitive responsibility

Category	Permission description	Design question
A	AI prohibited	Which capability needs independent evidence, and can the conditions support that inference?
B	AI minimal; presentational assistance only	Does the allowed function change meaning or perform a capability being assessed?
C	AI permitted	What substantive decisions must the student make and demonstrate, with or without choosing a model?
D	AI integrated; required in the task	Why is use required, what human judgement is assessed, and what access or equivalent route is needed?

A compulsory model task needs an educational reason for requiring the interaction. If the purpose is to evaluate generated material, an educator supplied output may sometimes allow an equivalent route without a commercial account. If operating and auditing a specific system is itself the outcome, that alternative may not be equivalent. Access, ethical objections and accessible arrangements must therefore be considered against the learning outcome, rather than resolved with a general promise of choice. The UAL cases show one reported way to make critical refusal educationally meaningful. [26,27]

Enterprise provision and revisable commitments

Microsoft's current documentation distinguishes baseline Copilot Chat, experiences inside supported applications, and additional licensed capabilities and organisational grounding. Availability depends on licensing, application and tenant configuration. Consequently 'Copilot' should not be used as a stable intervention description, and Chat inside an application should not automatically be equated with all embedded drafting and workflow functions. Vendor documentation can establish the described product conditions; it cannot establish educational value. A teaching study needs to record what participants could actually do. [29]

The inspected Copilot research illustrates the gap. Hasanein's survey of 760 hospitality and tourism students across nine Egyptian public universities concerns acceptance and use, not retained learning; the abstract does not adequately specify the product configuration. Bano and colleagues' preprint describes 27 interviews following CSIRO's six month M365 trial in 2024. Participants reported useful operational functions and limitations requiring verification. It is workplace experience evidence, rather than a university learning study. Neither justifies a general learning claim for embedded Microsoft 365 automation. [30,31]

For ChatGPT Edu, Gemini and other institutionally provided systems, provision should likewise be distinguished from the lesson in which a feature is used. The core cases here do not provide a comparative evaluation of those deployments. Their possible benefits include more consistent access and governance, but the effect on learning must be assessed in the specific design. The review does not conclude that enterprise provision lacks value; it identifies a mismatch between many deployment claims and the educational outcomes readers need to establish.

Chapter 3's question about dependence remains relevant under several commercial futures. Prices may fall, provision may stabilise, capabilities may improve, or terms and access may change. The companion does

not establish a timetable for an investment correction. It proposes examining whether the educational purpose survives switching product, losing a feature or returning to a route without models. Consumer accounts, enterprise contracts, API access and hosted models have different dependencies. Training, migration, support and redesign belong in the cost of a commitment even where the initial licence is inexpensive. [1, chapter 3, pages 25 to 38]

Reversibility is therefore educational as well as contractual. A team may be able to change supplier yet remain dependent on a category of automated feedback or drafting. Maintaining exportable materials, explicit criteria, alternative activities and staff expertise makes reconsideration more practical. The most useful question is what students and teachers would need to preserve the learning if the particular service disappeared or ceased to be acceptable. This is a proposed stress test of commitments, rather than a forecast of service failure.

Relational capacity

Relationships that change assessment and teaching

The book's third agenda concerns the ability to interpret change together and act on it. Chapter 4 distinguishes networks circulating practice, deliberation examining what is acceptable, and governance through which people influence decisions. It connects these to professional development, embedded learning design and resourcing. The distinction is useful empirically: an event can create contacts without changing a curriculum, and consultation can reveal concerns without giving participants authority. Relational capacity becomes visible when those activities connect to an identifiable decision and a subsequent opportunity to review it. [1, chapter 4, pages 46 to 53]

Sydney's finance capstone offers a particularly clear cycle. A 2024 pilot introduced Cogniti to approximately 700 students, including versions tailored to unit materials. End of semester feedback identified difficulties with prompting, disengagement after poor outputs and uncertainty about integrity boundaries. Staff and student partners then developed more specific guidance and practice for 2025. The public account documents a change informed by student experience, with grant support. It does not establish that the revised approach improved independent finance capabilities. Its value is the connection between a problem, participation and an educational response. [32]

That mechanism is more substantial than asking students whether they like a tool. Learners can identify where instructions clash, where assessment pressure discourages experimentation, or where the intended activity becomes something different in practice. Staff contribute disciplinary and assessment expertise. Learning designers can turn the discussion into teachable sequences and usable resources. A documented revision allows participants to see what their contribution changed and gives the next cohort something concrete to evaluate.

Partnership as joint inquiry

Williams and Ingle's UCL project paired a staff member with a student who had recently completed the relevant medical sciences assessments. They co designed evaluation criteria and examined ChatGPT output, with the student driving much of the inquiry and staff providing mentorship. The published case gives detailed evidence of a partnership process and its insights. A single pair does not represent the full student population, and the project does not estimate a causal effect on learning. It shows how assessment

testing and interpretation can become shared work, with distinct contributions from disciplinary experience and recent student experience. [33]

Georgetown's first year writing case approached policy itself as a learning activity. Students in one section considered restrictive, conditional and permissive positions, drafted a policy, and received responses from another section. Reflections exposed ambiguities and disagreements about writing and acceptable assistance. This is an evaluated reflective case, rather than proof that a university policy reduced misconduct. It substantiates the book's account of deliberation by making disagreement part of inquiry. It also shows why co creation should not be mistaken for a promise that everyone reaches the same position. [34]

UCC's publicly documented project funded student and staff partnerships across four colleges in 2023. Pairs tested existing assessments with ChatGPT, considered redesign and produced disciplinary case studies alongside general guidance. Paying students made participation a resourced activity rather than an assumed contribution. This is strong implementation evidence for a practical connection between adaptation, integration and relationship building. The case does not establish representative participation, sustained institutional influence or improved learning after the redesigns. It offers a concrete organisational method that can be evaluated on those dimensions. [35]

Student partners should not have to speak for all students. Recruitment can deliberately include sceptics, students using assistive technologies, those facing language or access barriers, and people with limited time for extracurricular participation. This is a design recommendation, rather than a claim that the cited projects achieved such representation. Payment and clear responsibilities help make participation feasible; published responses to proposals help distinguish influence from consultation. Staff participation similarly requires time and a route for less confident colleagues to contribute without performing enthusiasm.

Networks deliberation and influence

Melbourne's Student GenAI Innovators programme provides a broader reported infrastructure. Its public record describes the 2024 pilot and 2025 projects, including student voice work with Architecture, Building and Planning, resources developed with the central teaching team and discussion of when to use models. These are observable outputs spanning disciplinary and institutional settings. The record does not provide an estimate of learning gains or demonstrate how much decision authority participants held. It is evidence of participation and resource production, with the next question being how their work shapes continuing decisions. [36]

LSE's British Academy funded project connects a staff and student community of practice with research on programming education, short showcase meetings and an online case repository across social sciences. The inspected page states goals and outputs rather than evaluated outcomes. The format is a plausible way to circulate disciplinary experience without requiring long training courses. Its success would need to be judged by who can participate, what people subsequently change, how cases are critically reviewed, and whether materials remain accessible to those who cannot attend. [37]

Monash's 2026 Challenge Conversation on AI glasses brings specialists into public discussion of assessment, accessibility and learning. The format explicitly aims to think through the problem rather than settle it. The public summary provides a concrete example of deliberative infrastructure around an unsettled issue. No outcome evaluation was available, and the review did not analyse the recording. It is a useful complement to a repository: the question is not simply how a technology can be used, but what its use should mean for educational purposes. [39]

These examples substantiate different parts of relational capacity. A repository circulates experience; a workshop can interrogate it; a committee or programme process can authorise a change. None necessarily supplies the others. A workable connection might involve a published case, discussion with students and staff, a decision by the responsible programme body, and a dated explanation of what will be reviewed. This sequence is proposed here. It avoids assuming that participation in a network by itself creates authority over procurement or curriculum.

Staff judgement resources and reciprocal trust

Corbin and colleagues interviewed 20 teachers responsible for assessment at an Australian university. Their analysis describes competing demands and argues for institutional permission to compromise, diverge and iterate. It provides credible qualitative evidence about situated design difficulties, with a conceptual interpretation rather than a measured effect of a governance intervention. The paper's volume publication is 2026, following online publication in 2025. Its practical contribution is to explain why asking teachers to find a perfect solution can obstruct the provisional judgement that adaptation requires. [43]

Jisc's London South Bank case describes tailored GPTs for marking and feedback in law, project management and radiography, supported by workshops, champions and institutional review. It reports evaluation using grades, surveys and feedback logs, and says this informed a case for further licences. The underlying comparison design and detailed results were not provided on the page. The evidence supports a reported connection between teaching, support and investment decisions, but cannot justify a causal claim about learning or a quantified net workload saving. [38]

This creates a useful question for institutional practice. If a pilot appears to save time, where does that time go? Chapter 4 proposes protecting capacity for development and learning design. This review did not establish a public institutional case demonstrating that automation savings were durably ring fenced for those purposes. Reported faster turnaround should therefore be followed by an account of preparation, checking, maintenance and the work subsequently expected. A support chatbot may help people find guidance, but participation, escalation to a colleague and influence over decisions require additional arrangements. [1, chapter 4, pages 46 to 53]

The book's four trust directions also identify different research questions. Staff trust in students concerns both capability and observance of rules. Students' trust in staff concerns the quality and provenance of teaching, marking and feedback. Trust among students includes the fairness of effort and group contribution. Trust among staff includes consistency and responsible use in teaching. The evidence inspected here is stronger for staff and student interactions than for causal interventions addressing both peer directions. Treating all four as one aggregate 'AI trust' outcome would hide these differences. [1, chapter 2, pages 15 to 16]

Qu and Wang's study of 409 Singaporean undergraduates offers relevant relational evidence. Self reported disclosure was associated especially with perceived peer disclosure and comfort with instructors; qualitative responses described interpersonal risks. The observational mixed methods design supports a situated account of disclosure rather than establishing that a particular trust intervention causes honesty. It suggests that declaration rules work within relationships and expectations. A reciprocal discussion can examine staff use of generated materials and feedback as well as students' use, with different obligations stated clearly. [17]

Justified trust requires people to understand what others remain responsible for. Where staff use a model to prepare materials or support feedback, readers need a clear account of checking and the human decision,

especially when students are asked to disclose their own use. This recommendation does not presume identical disclosure rules for every role. Its purpose is to make educational responsibility legible. Published explanations, opportunities for questions and a route to correction can matter as much as a short declaration.

Where the agendas reinforce and constrain one another

The book argues that adaptation and integration can undercut each other. The cases substantiate observable parts of that account. In Sydney finance, students' uncertainty about integrity boundaries discouraged use during an intended integration pilot. Georgetown's policy inquiry surfaced disagreement about the meaning of assistance. UAL's fashion account describes anxiety about institutional expectations alongside purposeful critical teaching. These are examples of ambiguity affecting experience, rather than controlled estimates of a general institutional conflict. [32,34,27]

Other cases qualify the tension by showing routes through it. Sydney's programme architecture gives independent assurance and open activity different educational roles. UCC makes both possible responses objects of joint testing. The translation study begins with an unaided attempt before model supported revision. These designs demonstrate that adaptation and integration can coexist in an activity or programme. They do not establish that their coexistence is always effective, inexpensive or trusted. [5,35,22]

Three kinds of tension are analytically different. A logical tension arises when the same outcome is simultaneously defined as independent performance and performance achieved through unrestricted delegation. It requires a clearer outcome or separate evidence. A resource tension arises when oral assurance, tool training and participation all demand the same limited staff and student time. It requires prioritisation and resourcing. An implementation tension arises when provision, guidance, interface behaviour and marking criteria communicate inconsistent expectations. It requires concrete explanation and revision. Labelling all three as resistance to change would miss the appropriate response.

Corbin, Dawson and Liu's structural and discursive distinction provides an important qualification. Their conceptual article argues that communicating permissions without changing assessment mechanics can create an illusion of security. That challenges any version of relational capacity consisting mainly of better conversation about unchanged tasks. The abstract and metadata were accessible here, rather than the full paper. The bounded implication is that deliberation needs a route into assessment design; clarity alone cannot control assistance in an unsupervised task. [44]

Central coordination can also support the book's goals. Consistent access, accessible arrangements, common expectations and a manageable assessment calendar can protect fairness. Distributed professional judgement needs that infrastructure. The cases do not justify equating decentralisation with educational quality or treating a central service as necessarily harmful. A productive division of responsibility would set common obligations centrally while allowing disciplines to specify the learning purpose, appropriate tool functions and required evidence. This is the companion's proposed governance principle.

Relational capacity therefore helps most plausibly by making mismatches visible and enabling a responsible change. It is not yet established as a necessary causal mediator of successful reform in every institution. Well designed local teaching can work with modest additional structures, while a large network can fail to change decisions. The evidence supports investing in relationships with an explicit task and

decision route, then evaluating what they accomplish. It leaves the book's larger institutional proposition open to further testing.

Comparing practices

The following tables separate purpose and evidence from implementation demands. Confidence concerns the inference named in the cell. Category fit uses the illustrative A to D scheme above, not the policy of a cited institution. Risks and workload descriptions are analytical judgements unless an evaluated result is identified. Reversibility means the practical ability to change the approach while preserving the educational purpose.

Purpose and evidence

Table 3 Purpose and evidence across major practices

Practice	Purpose and agenda	Evidence and confidence	Sources
Programme mapping	Assure selected degree capabilities. Adaptation and relational capacity.	Reported architecture and Advocated guidance. Good implementation detail; institutional learning effects unestablished.	5, 7, 8
Invigilated written task	Observe independent interpretation or knowledge. Adaptation.	Advocated as one assurance format. Inference depends on task, conditions and coverage; no universal security claim.	2, 8
Digital exam on own device or in a computer room	Combine controlled conditions with digital work. Adaptation.	Advocated security design. No inspected comparative benchmark establishes a particular platform or configuration as secure.	8, 45
Interactive oral	Elicit reasoning in dialogue. Adaptation.	Evaluated qualitative case and earlier survey. Credible situated experience; present security and scalability not demonstrated.	9, 10
Staged work or portfolio	Show developing decisions. Adaptation and integration.	Reported creative course design. Learning visibility plausible; authorship and retained learning require additional evidence.	26
Project with supervision and a selected oral	Interpret substantial independent inquiry. Adaptation.	Analytical option informed by validity and creative capstone design. No general efficacy estimate.	2, 26
Writing process platform	Make parts of production visible. Adaptation.	Evaluated small Cadmus case. Useful experience evidence; incomplete record and limited comparison.	16
Critique generated material	Develop disciplinary evaluation. Integration.	Reported economics design. Clear required contribution; later independent effects unmeasured in the account.	21
AI feedback after an initial attempt	Support revision while retaining a student starting point. Integration.	Evaluated translation process study. Stronger evidence of engagement than of durable learning.	22
Purpose built tutor	Provide structured practice and feedback. Integration.	Evaluated experiments. Positive immediate university result; guarded school design mitigates harm. Limited transfer.	19, 20

Practice	Purpose and agenda	Evidence and confidence	Sources
Permission framework and declaration	Clarify boundaries and acknowledge assistance. Adaptation and integration.	Evaluated before and after pilot; observational disclosure evidence. Policy and relational mechanisms plausible; causal learning claim weak.	17, 23
Critical use and refusal	Develop informed disciplinary judgement. Integration.	Reported creative teaching and Advocated framework. Concrete design; comparative capability effects unavailable.	26, 27, 28
Paid assessment partnership	Turn testing and experience into redesign. All three agendas.	Reported funded project and Evaluated small partnership cases. Good mechanism detail; sustained effects unestablished.	33, 34, 35
Networks with deliberation and a decision route	Circulate, scrutinise and revise practice. Relational capacity.	Reported networks and conversation. Outputs visible; influence and educational effects need evaluation.	36, 37, 39
Enterprise tools and automated support	Provide governed access or assistance. Integration and relational capacity.	Vendor capabilities Claimed; workplace experience Evaluated; feedback pilot evaluation Reported. Student learning inference remains specific to design.	29, 31, 38

Implementation and reversibility

Table 4 Implementation demands for the same practices

Practice	Category fit	Learning equity and disciplinary considerations	Workload and reversibility
Programme mapping	All	Account for electives, transfers and interdisciplinary routes; avoid gaps hidden by a neat map.	High initial coordination. Revisit coverage annually; existing tasks may be removed.
Invigilated written task	A or B	Align time pressure and available resources with the outcome; arrange access and practice.	Scheduling, supervision and moderation. Format can change; curriculum effects take longer.
Digital exam on own device or in a computer room	A or B; C or D under controls	Consider device variation, assistive functions and accessible software. A room alone does not establish security.	Technical support, configuration and contingency planning. Platform dependence can constrain change.
Interactive oral	Usually A or B	Prepare students; distinguish disciplinary reasoning from incidental verbal agility. Adapt arrangements to the outcome.	Individual delivery, scheduling and moderation. Pilot before scaling; easy format revision.
Staged work or portfolio	All	Reward substantive decisions; process narratives may also be generated. Preserve feedback and peer discussion.	Repeated checkpoints can accumulate. Remove unnecessary records; maintain portable formats.
Project with supervision and a selected oral	All	Define independence by method and judgement. Respect research privacy and programme requirements.	Supervision plus selected verification. Avoid duplicating existing milestones; revise component weighting.

Practice	Category fit	Learning equity and disciplinary considerations	Workload and reversibility
Writing process platform	All	Off platform thinking and assistive work complicate interpretation; explain the record and its purpose.	Training, support and data handling. Maintain export and alternative routes.
Critique generated material	C or D; supplied output can suit A	Students need disciplinary foundations and access to sources. Grade verification and reasoning.	Prepare exemplars and criteria. Highly portable if generated material is supplied.
AI feedback after an initial attempt	C or D	Check overconfidence and uncritical uptake. Preserve the student's initial reasoning and a later independent task.	Checking, support and delayed evaluation. Replace the tool while retaining revision pedagogy.
Purpose built tutor	C or D	Guardrails and content need checking; a tutor should supplement suitable human teaching.	Design, testing and maintenance may be substantial. Keep learning materials independent of one interface.
Permission framework and declaration	All	Explain specific functions and ambiguous cases. Disclosure is a situated account, not proof of compliance.	Guidance updates and consistent interpretation. Low technical dependence; ongoing educational labour.
Critical use and refusal	C or D; inquiry can suit A	Grade a justified position, not enthusiasm. Distinguish critical inquiry from operation of a required tool.	Disciplinary preparation and discussion. Strong portability where purposes remain explicit.
Paid assessment partnership	All	Recruit varied experiences and positions; specify influence and pay student labour.	Facilitation, staff time and response to recommendations. Preserve a bounded remit and succession.
Networks with deliberation and a decision route	All	Enable asynchronous access and dissent. Connect examples to actual programme decisions.	Continuing facilitation and learning design. Avoid reliance on a few unpaid champions.
Enterprise tools and automated support	C or D where approved	Provision does not settle assessment permission. Specify feature, data access and human responsibility.	Licence, checking, training and migration costs. Keep exportable materials and an alternative teaching route.

Formats can be combined. A dissertation, for example, can retain a substantial written or creative project while drawing evidence from supervision and a limited oral examination. A programme can assess tool supported performance in one task and independent interpretation in another. The tables should therefore be used to compare the inference and cost of a design, rather than to select one supposedly secure format for every outcome.

Claims that need more evidence

Several repeated claims combine a defensible observation with a much larger conclusion. A generated answer meeting criteria shows task vulnerability; it does not measure the prevalence of student misconduct. Reduced recorded misconduct following new permissions may show that the new rules are workable; it

does not establish more independent learning. Higher marks can reflect better learning, more effective assistance, changed marking or changed tasks. Each explanation needs evidence.

‘Authentic’, ‘personalised’ and ‘reflective’ describe educational features without establishing security. An oral can elicit valuable judgement without validating every programme outcome. A writing history can reveal parts of a process without proving authorship. A declaration can make assistance discussable without guaranteeing compliance. None should bear an inference broader than the evidence it supplies.

Speed, confidence and satisfaction can be worthwhile outcomes. They are nevertheless different from retained understanding. A tool may help an experienced student save time on a routine function while allowing a novice to bypass the very capability they need to develop. Assertions of general skill erosion also require stronger evidence: a short task cannot by itself demonstrate lasting atrophy, and an observed gap can reflect a capability never developed. Chapter 3’s distinction remains important. [1, pages 34 to 35]

The major empirical gap is sustained, disciplinary learning across programmes. The review found few directly inspected studies that track repeated use, transfer to unfamiliar tasks and later independent performance together. Evidence is also limited on the cumulative workload of several reforms, equitable outcomes for different students, and whether procurement decisions remain educationally reversible. This is a gap in the inspected evidence base, not a claim that no relevant research exists.

The institutional question requires more than an adoption dashboard. A stronger evaluation would connect participation and governance decisions to changed teaching, then to student experience and learning. It would also retain disagreement and unsuccessful revisions in the record. A useful case report should explain what was stopped, what became unnecessary and what labour remained. Published success narratives alone make it difficult for another institution to judge viability.

Practical options for programme and faculty discussions

These options apply the companion’s synthesis. They are proposals for readers to adapt, rather than existing or planned actions at the commissioning institution. Timescales are planning ranges, not empirical estimates. Implementation depends on the applicable approval calendar, available expertise and the scale of the programme. Each option needs a named owner and a review decision.

Map a small set of consequential capabilities

Start with one programme or coherent pathway and identify the capabilities for which assurance matters most. Map where every student demonstrates them, whether the evidence concerns performance with or without tools, and how an insufficient demonstration is handled. Prerequisites include usable outcomes, an assessment inventory, student pathway information and authority to alter tasks. A programme leader, academic team and learning designer can do an initial review within a term; significant changes may require the next formal approval cycle.

The workload is concentrated coordination and redesign rather than necessarily more assessment. Count meetings, mapping, preparation and moderation, and identify tasks that can be removed. Review the map before implementation and after the first cohort. Measures include coverage across routes, reliability of judgements, student understanding and total assessment load. Revise if optional pathways bypass assurance, the outcome remains too vague, or the added tasks exceed capacity. Stop expansion while those issues are unresolved.

Pilot a prepared oral or observed task

Use one consequential outcome and a small enough cohort to permit careful preparation and moderation. Publish the criteria, give practice opportunities and agree accessible arrangements before delivery. A conversation should require interpretation, application or explanation that the outcome warrants. Prerequisites include assessor preparation, suitable scheduling, an accommodation process and clarity about whether this is assessment or investigation. Preparation may take one term, with a pilot in the next available delivery cycle.

Record the full labour involved and compare it with the task it replaces. Evaluate student experience, accessibility, consistency between assessors and the additional evidence about capability. Review immediately after delivery before scaling. Revise if questions mainly test unexpected verbal agility, arrangements create inequity, or reliability is weak. Stop treating the task as programme assurance if it cannot support the specified inference, even if participants found the conversation engaging.

Build a disciplinary sequence that returns to independent work

Select a topic where model support has a clear function. Students first attempt a relevant part, use a model for comparison, feedback or questioning, then verify and explain their decisions. Later, assess an unfamiliar task without that support where independence is the intended outcome. Prerequisites include adequate disciplinary teaching, equitable access, clear permissions and a route for students whose access or objections require an alternative. Design within a term and evaluate across at least one delivery cycle.

The workload includes materials, tool checking, tutorial support and evaluation. Collect both assisted performance and later independent evidence, with a comparison where feasible. Ask students about their actual workflow rather than assuming they followed the intended sequence. Review after the immediate activity and delayed task. Modify if students simply adopt generated judgements, feedback discourages independent checking, or the cognitive demands exceed their foundations. Stop the specific design if gains disappear without assistance or the educational purpose cannot survive access problems.

Fund student and staff assessment inquiry

Give a small partnership group a concrete remit: test selected tasks, identify ambiguities and recommend changes. Include participants with different experiences and positions, compensate student labour and provide a realistic staff allocation. Prerequisites are a sponsoring programme body, access to suitable non confidential materials and a commitment to respond publicly to recommendations. A bounded project can run within a term; assessment approval and delivery may take longer.

Evaluate what participants contributed, whose perspectives were absent, which proposals changed a decision and what learners experienced afterwards. Publish an accessible response explaining adoption, modification or rejection. Review both participation and resulting designs after the next delivery. Revise recruitment or remit if the group represents only confident enthusiasts or becomes a source of unpaid support. Stop calling it partnership if participants have no meaningful influence and cannot see how their work is used.

Connect networks to decisions and resource review

Use an existing disciplinary or teaching network where possible. Circulate a short, critically annotated case, discuss its fit and ask a responsible programme body to consider a defined change. Embedded learning design support can help translate proposals into tasks and materials. Prerequisites include an escalation

route, time to participate and access for colleagues unable to attend live meetings. Establishing the process can take a term; continuity needs an annual allocation rather than a one off event.

Measure participation beyond attendance, changes in teaching, the accessibility of resources and whether decisions return to the group for review. Count preparation and maintenance as well as meetings. Review each term and at the annual resource allocation. Revise if the same few people carry the work, examples circulate without scrutiny, or decisions disappear into another committee. Reduce the number of initiatives if their cumulative demands undermine teaching or meaningful participation.

Review an enterprise commitment against its educational purpose

For one consequential tool supported activity, describe the product, functions, data access, configuration and educational outcome. Separate procurement from permission to use a feature in assessment.

Prerequisites include academic and student input alongside IT, accessibility and appropriate procurement expertise. Review before deployment where possible, after a term of use, and before renewal. A new contract or major configuration change may require a longer institutional process.

Evaluate learning and total support costs alongside usage. Test whether the activity still works if a feature changes, availability is reduced or the product is replaced. Specify exportable resources and an alternative teaching route. Modify the commitment where checking and support erase intended benefits or access is unequal. Stop compulsory use of a particular service if its conditions prevent the intended learning or an acceptable route for students, while preserving the capability the curriculum aimed to teach.

Dissertations and substantial independent projects

For a substantial project, state what independence means in relation to the actual method, interpretation and decisions. Agree permissible assistance and acknowledgement early, then use supervision to discuss evidence, methodological choices and emerging conclusions. Selected drafts or project records can support that conversation, but should have a reason beyond accumulating proof of work. An oral component can probe the student's interpretation of a specific decision without claiming to authenticate every line of the project.

At undergraduate level, the emphasis may be on competent inquiry and a defensible argument. Taught masters work may require more demanding methodological judgement. Doctoral work has distinct expectations of original contribution and examination; the small teaching cases in this review cannot determine doctoral standards. Any redesign should follow the relevant programme and examination requirements. Review the combined burden of supervision, records and oral components, including the privacy implications of uploading research material to a service.

Methods scope and access

This is a targeted critical evidence review completed on 4 October 2026. The supplied book was read in full before research. Searches combined its three agendas with assessment validity, programme assurance, two lane approaches, interactive orals, detection, disclosure, retention, transfer, critical literacy, translation, creative practice, student partnership, communities of practice and enterprise Copilot. Citation following from strong papers and institutional pages added related studies and implementation accounts. Both available general web search routes were used, with primary source retrieval for substantive claims.

Priority was given to work from 2023 onwards, retaining relevant earlier assessment research. Sources include peer reviewed articles, a clearly identified preprint, an accepted article abstract, official guidance and public institutional cases. The focus is assessment and learning in higher education, with school and workplace studies marked as indirect. The review includes UK, Australian, European and North American examples, together with relevant Asian and other settings. It is not a systematic database review, and no exhaustive coverage or reproducible screening count is claimed.

Practices were included where the source identified an educational design, a measurable outcome, participant experience or a mechanism connecting participation to change. Launch announcements and vendor descriptions were insufficient evidence of educational effectiveness. Positive findings, null or qualified results and credible arguments against simple integration or prohibition were sought. Public source publication dates were distinguished from the dates of interventions. Several studies concern 2023 tools and cannot be treated as performance estimates for present models.

Full text was accessed for core validity, detection, translation, assessment review and partnership studies and most institutional accounts. Retrieval failed for several publisher pages. For Kestin, Liang, Choi, Hasanein, Barcaui, the structural assessment article and the wearable article, claims were restricted to accessible abstracts or indexed primary passages. The reference notes identify those limitations individually. The accepted Dong abstract is not presented as a fully inspected published study. Funding and author roles are noted where verified; absence of a verified statement is not treated as evidence of no funding or interest.

The bibliographic review corrected an important date distinction: The wicked problem of AI and assessment appeared online in 2025 but has a 2026 volume citation. Jackson's original conceptual article and the published AIAS pilot were located separately from descriptions of their frameworks. The report treats UQ's hosted advice as guidance and TEQSA's constructed examples as illustrations, avoiding invented institutional cases. It does not treat reported feedback or a framework author's implementation narrative as independent confirmation of effectiveness.

Only the book's published argument and verified public cases inform references to identifiable institutions or people. No unpublished commissioning context, faculty plans or personal circumstances are included. The options are general analytical proposals. URLs and live product specifications were checked during the research but can change; institutional readers should verify current local rules before implementation.

References and evidence notes

All sources were accessed or their accessible primary passages inspected by 4 October 2026. Each entry links to the cited source; an open access version is also linked where useful. Evidence notes describe access and the basis of the claims, rather than certifying general effectiveness.

[1] Carrigan, M. (2026). The AI Crisis in Higher Education: A Strategic Proposal. University of Manchester. [Source](#) Book. Supplied edition read in full; public publication and title verified. Page references use the printed pagination of that edition.

[2] Dawson, P., Bearman, M., Dollinger, M. and Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005 to 1016. [Source](#) [Open access version](#)

Peer reviewed conceptual article. Full repository PDF accessed.

[3] Scarfe, P., Watcham, K., Clarke, A. and Roesch, E. (2024). A real-world test of artificial intelligence infiltration of a university examinations system: A ‘Turing Test’ case study. PLOS ONE, 19(6), e0305354. [Source](#)
Evaluated. Full text accessed. Researcher submissions into real assessment; one degree. EPSRC CHAI project support reported.

[4] TEQSA (2023). Assessment reform for the age of artificial intelligence. 23 November. [Source](#)
Advocated. Official sector guidance, not an evaluation or UK regulatory requirement.

[5] University of Sydney (2025). The Sydney Assessment Framework. Teaching at Sydney. [Source](#)
Reported and Advocated. Public implementation account accessed in full.

[6] University of Sydney (n.d.). Frequently asked questions about the two-lane approach to assessment in the age of AI. [Source](#)
Reported and Advocated. Live institutional guidance accessed. Wording is time sensitive.

[7] Bridgeman, A., Liu, D. and Weeks, R. (2024). Program level assessment design and the two-lane approach. Teaching at Sydney, 12 September. [Source](#)
Reported and Advocated. Full public account accessed; includes the Graduate Certificate example and proposed designs.

[8] TEQSA (2025). Enacting assessment reform in a time of artificial intelligence. 24 September; current web PDF accessed in October 2026. [Source](#) [Open access version](#)
Advocated. Full sector guidance accessed. Three reform pathways are proposals, not comparative trial findings.

[9] Yung, R., Moyle, B. and Yang, E. C. L. (2024). Mitigating the Impact of ChatGPT: Cultivating Professional Identity Through Interactive Oral Authentic Assessments. Journal of Hospitality & Tourism Education. Published online 28 March. [Source](#) [Open access version](#)
Evaluated. Full published PDF accessed. Qualitative capstone study of assessment delivered in 2021; funding not established from retrieved material.

[10] Sotiriadou, P., Logan, D., Daly, A. and Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. Studies in Higher Education, 45(11), 2132 to 2148. [Source](#)
Evaluated. Abstract and Griffith implementation account accessed; full article not accessed. Predates current language models.

[11] Brufau Alvira, N., Bannister, P. and Santamaría Urbieto, A. (2025). Validating the PANDORA GenAI Susceptibility Rubric for Higher Education Assessment: A Field Test of All Translation and Interpreting BA Assignments. Higher Education Quarterly, 79(4), e70056. [Source](#)
Evaluated. Full text accessed. UNIR project PP 2023 02 support reported; rubric field test, not independent learning evaluation.

[12] Weber-Wulff, D. and colleagues (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19, article 26. [Source](#)
Evaluated. Full text accessed. Technical evaluation of 14 tools at the time of testing; no estimate of all current products.

[13] Liang, W., Yuksekgonul, M., Mao, Y., Wu, E. and Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7), 100779. [Source](#) [Open access version](#)
Evaluated. Author institutional abstract and indexed primary passages accessed; full publisher page inaccessible. Claims limited to the reported bias finding.

[14] UCL (n.d.). Three categories of GenAI use in assessment. Teaching and Learning. [Source](#)

Reported and Advocated. Full public guidance accessed. Includes permission, acknowledgement and handling suspected misuse.

[15] UCL (n.d.). Engaging with Generative AI in your education and assessment. Assessment Success Guide. [Source](#)

Reported and Advocated. Full student guidance accessed. Live institutional policy context.

[16] Briffa, H. (2026). Making students visible: Transparency and the politics of digital assessment. Politics. Published online. [Source](#)

Evaluated. Full text accessed. Small postgraduate case using Cadmus; surveys and implementation reflection. Funding not established from retrieved material.

[17] Qu, Y. and Wang, J. (2026). To disclose or not to disclose: Peer influence and psychological factors in students' use of generative artificial intelligence. British Journal of Educational Psychology. Published online 14 April. [Source](#)

Evaluated. Full text accessed. Mixed methods observational study; self reported disclosure. Funding not established from retrieved material.

[18] Jackson, J. (2025). Higher order prompting: Applying Bloom's revised taxonomy to the use of large language models in higher education. Studies in Technology Enhanced Learning, 4(1), 1 to 17. [Source](#) [Open access version](#)

Advocated. Full peer reviewed conceptual paper accessed; framework is not a validated scale of learning outcomes.

[19] Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö. and Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. PNAS, 122(26), e2422633122. [Source](#)

Evaluated. Full published text accessed. Turkish school experiment in 2023. Wharton research funding reported; August 2025 correction concerns affiliation.

[20] Kestin, G., Miller, K., Klaes, A., Milbourne, T. and Ponti, G. (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports, 15, 17458. [Source](#) [Open access version](#)

Evaluated. Author abstract and indexed publisher methods and discussion accessed; full page and PDF requests failed. No competing interests declared; funding not established.

[21] Maitra, C., Melatos, M. and Vidal-Robert, J. (2025). Making Peace with AI: ChatGPT-Integrated Assessments in Economics. Teaching at Sydney, 8 May. [Source](#)

Reported. Full practitioner account accessed. Teaching changes in 2023 and 2024; no controlled learning evaluation reported.

[22] Xu, S., Su, Y. and Liu, K. (2025). Investigating student engagement with AI-driven feedback in translation revision: A mixed-methods study. Education and Information Technologies, 30, 16969 to 16995. [Source](#)

Evaluated. Full text accessed. One translation and revision session with 29 masters students; funding not established from retrieved material.

[23] Furze, L., Perkins, M., Roe, J. and MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. Australasian Journal of Educational Technology, 40(4), 38 to 55. [Source](#) [Open access version](#)

Evaluated and Reported. Full published PDF accessed. British University Vietnam institutional before and after account; framework developers among authors.

[24] The Open University (n.d.). Generative AI for students. [Source](#)

Reported and Advocated. Full public student guidance accessed. Three categories distinguish prohibited, permitted and required use.

[25] TEQSA and University of Queensland (n.d.). Principles for criteria and standards in assessment for gen AI use. [Source](#)

Advocated. Full public guidance accessed. Explicitly advice rather than mandatory policy.

[26] Ainsworth, P. (2026). Our approach to AI on MA Photography and Digital Practice at UAL. UAL AI Network, 19 August. [Source](#)

Reported and Advocated. Full practitioner account accessed; posted by Christopher Rowell, with Ainsworth identified as author.

[27] UAL AI Network (2026). Case Study: Decolonising AI in Fashion Education. 29 June. [Source](#)

Reported. Full account of Kate Greenslade and Ami Podrebarac's teaching accessed; early student responses lack an evaluation protocol on the page.

[28] Lodge, J. M. and colleagues (2026). Assuring quality learning in a gen AI-integrated future: The role of adaptive capabilities. TEQSA, 24 June. [Source](#)

Advocated. Full guidance accessed. Illustrative faculty scenarios are not identified institutional evaluations.

[29] Microsoft (n.d.). Overview of Microsoft Copilot Chat. Microsoft Learn. [Source](#)

Claimed for capabilities and service conditions. Full current vendor documentation accessed; provides product specification, not educational effectiveness evidence.

[30] Hasanein, A. M. (2025). Responses to the AI Revolution in Hospitality and Tourism Higher Education: The Perception of Students Towards Accepting and Using Microsoft Copilot. European Journal of Investigation in Health, Psychology and Education, 15(3), 35. [Source](#) [Open access version](#)

Evaluated. Published abstract accessed; full article not accessed. Survey of acceptance and usage intentions; product configuration insufficiently specified in the abstract.

[31] Bano, M., Zowghi, D., Whittle, J., Zhu, L., Reeson, A., Martin, R. and Parsons, J. (2025). A Qualitative Study of User Perception of M365 AI Copilot. arXiv 2503.17661, version 2, 30 March. [Source](#) [Open access version](#)

Evaluated. Full preprint accessed; described as under review. CSIRO workplace study, not university student learning. Authors study their organisation's trial.

[32] Sun, J. (2025). Co-creating Gen-AI guidance: tailored support through student partnership. Teaching at Sydney, 24 April. [Source](#)

Reported, with student feedback. Full public account accessed. Strategic education grant support reported; subsequent learning outcomes not evaluated on the page.

[33] Williams, A. and Ingle, E. (2025). Assessment design through co-creation: Student-staff partnership in evaluating the impact of artificial intelligence. International Journal for Students as Partners, 9(1), 214 to 226. [Source](#) [Open access version](#)

Evaluated and Reported. Full published PDF accessed. One staff and student partnership; qualitative and assessment output analysis. Funding not established from retrieved material.

[34] Debelius, M., Chehak, M., Le, K., Oh, S., Lyons, S., Kim, G., Nutting, H. and Holtschlag, J. (2025). Designing an AI policy: An experiment in co-creation. International Journal for Students as Partners, 9(1), 151 to 160. [Source](#)

Evaluated and Reported. Author uploaded full published text accessed through ResearchGate; journal site blocked retrieval. Two first year writing sections; reflective case study.

[35] Goff, L. (n.d.). Academic Integrity and GenAI: Guiding Responsible Use With Students-as-Partners Collaboration. In Costello, J., Doljanin, Z., McAreavey, N. and Walsh, F. (eds), Teaching Humanities and Social Sciences in the Era of GenerativeAI: Case Studies from around Ireland, chapter 2. UCD College of Arts and Humanities. [Source](#)

Reported. Public chapter accessed. UCC project ran March to December 2023; SATLE funding supported paid student partners. Collection date not established.

[36] University of Melbourne (n.d.). Student GenAI Innovators group. [Source](#)

Reported. Full public record accessed; describes 2024 pilot and 2025 projects. Outputs documented; causal outcomes unavailable on the page.

[37] London School of Economics and Political Science (n.d.). Capacity building on generative AI tools for programming education. [Source](#)

Reported and Advocated. Full project page accessed. British Academy funding March 2025 to June 2026; intended network and repository outputs.

[38] Jisc (2025). AI in Marking and Feedback Pilot: good practice case studies: London South Bank University. 5 November. [Source](#)

Reported evaluation. Full public case account accessed; underlying grade analyses and comparison design not provided on the page.

[39] Sharpe, S., McArthur, V., Seligmann, A. and Fawns, T. (2026). PAAIR challenge conversation: What do AI glasses mean for learning and assessment? Monash Teaching Community, 3 September. [Source](#)

Reported. Public record and summary accessed. Conversation held 27 August; recording not analysed. Deliberation, not an effectiveness evaluation.

[40] Stephenson, R. and Armstrong, C. (2026). Student Generative Artificial Intelligence Survey 2026. HEPI Report 199, 12 March. [Source](#)

Evaluated. Official findings and survey method accessed. Savanta survey of 1054 full time UK undergraduates in December 2025; sponsored by Kortext.

[41] Jisc (2025). Student perceptions of AI 2025. [Source](#)

Evaluated and Reported. Indexed official report passages and official news account accessed; full report page requests failed. Perceptions evidence, not causal learning effects.

[42] Choi, J. H., Monahan, A. B. and Schwarcz, D. (2024). Lawyering in the Age of Artificial Intelligence. Minnesota Law Review, 109(1), 147 onwards. [Source](#) [Open access version](#)

Evaluated. Author institutional abstract accessed; full published PDF request failed. Randomised legal task performance study; claims restricted to abstract.

[43] Corbin, T., Bearman, M., Boud, D. and Dawson, P. (2026). The wicked problem of AI and assessment. Assessment & Evaluation in Higher Education, 51(4), 736 to 752. First published online 3 September 2025. [Source](#) [Open access version](#)

Evaluated and Advocated. Full published text accessed; 20 teacher interviews at one Australian university. Volume date checked against institutional metadata.

[44] Corbin, T., Dawson, P. and 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 to 1097. [Source](#) [Open access version](#)

Advocated. Institutional abstract and metadata accessed; full PDF retrieval failed. Conceptual structural and discursive distinction.

[45] Corbin, T., Sharpe, S. and Dawson, P. (2026). On AI glasses and wearable AI in assessment. *Assessment & Evaluation in Higher Education*. Published online. [Source](#)

Advocated. Indexed publisher passages accessed; full text request failed. Conceptual risk argument, not a demonstrated attack success rate.

[46] Barcaui, A. (2025). ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. *Social Sciences & Humanities Open*, 12, 102287. [Source](#)

Evaluated. Indexed publisher abstract and selected discussion and declarations accessed; full methods page unavailable. No external financial support declared. Retention result requires replication.

[47] Dong, J. (2026). From Generative-AI Feedback to Sport Skill Learning: The Roles of Perceived Competence and Autonomous Motivation in University Physical Education. *Frontiers in Psychology*, 17. Accepted 14 September. [Source](#)

Evaluated, accepted article abstract only. Publisher states final formatted version forthcoming. Motor learning example is indirect for intellectual disciplines; full methods not yet inspected.