

Designing AI-Supported Assessment in Marketing Education: Specified Latitude and the CRAFT Framework

Ying Wu, University of Dundee, United Kingdom
Andrew Ross, University of Dundee, United Kingdom
Malcolm Stewart, University of Stirling, United Kingdom
Fahima Alam, Queen Mary University of London, United Kingdom

The European Conference on Education 2026
Official Conference Proceedings

Abstract

Universities must prepare marketing graduates for practice in which generative artificial intelligence (AI) is routine, while preserving assessment validity and independent disciplinary judgement. Detection-led prohibition is difficult to enforce reliably, and permission without evidential requirements can weaken assessability, yet the literature does not specify the design conditions under which authentic assessment converts permitted AI participation into demonstrable evaluative judgement. This study examines two iterations of an undergraduate marketing module, with a postgraduate module as an embedded contrast, across 191 enrolled and 182 marked students. It combines a documentary audit of review forms, briefs and rubrics with a documentary census of 182 submissions across four assessment sets, analysed abductively against stated criteria. Three findings are advanced. Client authenticity supplied constraints against which generated output could be judged, but the assessment architecture specified no AI participation and carried no AI-related criterion in either iteration. Although generative AI was taught in class and demonstrated for specified creative operations, documented declaration rates were 5.8% and 28.6% across the two undergraduate iterations, and of 182 submissions none evidenced revision or disciplinary justification of generated output. Mean marks in the focal undergraduate module did not rise, but no AI-related criterion existed from which an effect of specification could be evaluated. The case is diagnostic rather than evaluative. The study proposes specified latitude, the task-stage-specific mapping of permitted AI operations to the evidence of judgement they must leave, operationalised through CRAFT as a design response to the documented gap.

Keywords: generative artificial intelligence, authentic assessment, marketing education, evaluative judgement, assessment design

iafor

The International Academic Forum
www.iafor.org

Introduction

Generative artificial intelligence (AI) now performs the operations that marketing assessment has long used as evidence of competence. Systems in general use draft copy, generate campaign concepts, produce brand-styled imagery and propose elements of strategic positioning (Cillo & Rubera, 2025; Grewal et al., 2025b; Guha et al., 2024). The consequence is not primarily a policing problem. It is that the artefacts on which marketing assessment relies, the campaign plan, the concept board, the pitch deck, have lost their former standing as unambiguous evidence of the capabilities they were designed to certify. An assessment regime that continues to infer capability from artefact quality alone is making an inference the technology no longer supports, and the defensibility of such inferences is what assessment validity consists in (Ali & Maroulis, 2026; Kane, 2013).

Institutional responses have converged on three positions, none of which resolves the problem. Detection-led prohibition is difficult to enforce reliably and is discouraged by sector guidance (Miao & Holmes, 2023; Quality Assurance Agency for Higher Education [QAA], 2023; Xia et al., 2024). Permission without evidential requirements preserves professional realism but can weaken assessability, leaving no account of what the student contributed. A third position, appeal to authentic assessment, has become the default recommendation, on the reasoning that tasks situated in real professional contexts resist automated completion. Kofinas et al. (2025) contradict it directly: authentic assessments did not safeguard academic integrity, and markers could not reliably detect AI contribution.

Marketing education is theoretically revealing rather than merely convenient. AI already performs central professional operations in the discipline, campaign artefacts are highly generative and multimodal, and artefact quality is therefore unusually weak evidence of unaided capability (Grewal et al., 2025a; Narang et al., 2025). Client constraints, meanwhile, supply natural standards against which generated material can be judged.

What remains unspecified is the design problem itself. Neither banning AI, nor permitting it without condition, nor appending a generic declaration that AI was used somewhere, tells an assessment designer which AI operations to permit, at which point in the task, and what evidential trace that permission must leave. Recent work establishes that evaluative judgement of AI outputs is the capability now at stake (Bearman et al., 2024) and that the digital must be designed into authentic assessment rather than added to it (Nieminen et al., 2023), but neither specifies the architecture that would produce such judgement or render it markable. This study addresses that specification gap.

The aim is to examine how generative AI can be embedded within authentic marketing assessment so that its use becomes visible, evaluable and educationally productive, while independent disciplinary judgement and assessment validity are preserved. Three questions follow. First, what evidential gaps arise when generative AI is pedagogically introduced and demonstrated but not specified within the assessment architecture? Second, what can cross-iteration and embedded-contrast evidence establish, and not establish, about assessment outcomes under an AI-silent architecture? Third, what design conditions follow for making AI-mediated evaluative judgement visible, attributable and assessable, and what bounds them?

The study examines an undergraduate marketing module delivered across two academic years in a Scottish business school, with a postgraduate module providing an embedded contrast. It draws on the formal assessment architecture, a documentary census of 182 submissions across

four assessment sets, reflective teaching records and institutional assessment records. The case is diagnostic rather than evaluative: it documents an architectural absence rather than testing an implemented intervention. Its contribution is to separate disciplinary authenticity from evidential assessability, to separate pedagogical preparation from assessment architecture, and to propose specified latitude and the CRAFT framework as a design response to the gap the case documents.

Literature Review and Theoretical Development

Generative AI and the Limits of Prohibition

The early scholarly response to generative AI in higher education treated it as a threat to assessment security, and the resulting literature is organised around detection, deterrence and redesign. Xia et al. (2024), reviewing 32 empirical studies, found detection approaches unable to sustain the evidential burden placed on them and recommended that institutions rethink assessment policy rather than invest further in detection. Sector guidance reaches the same conclusion (Miao & Holmes, 2023; QAA, 2023). The open question is therefore not whether prohibition works, which is largely settled, but what a redesign-led response must contain in order to work.

On that question the literature is thinner and, in one respect, misleading. The redesign most often recommended is a move toward authentic, contextualised tasks, on the assumption that situational specificity defeats general-purpose generation. Kofinas et al. (2025) found it wanting, separating two properties the literature conflates. Authenticity may raise the standard generic output must meet, a quality effect, without making provenance visible, an evidential effect.

A second strand specifies permitted use. Perkins et al. (2024) propose a five-level scale from no AI use to full AI use, attaching expectations to permission: at its collaborative levels students evaluate and modify generated output and supply reflective commentary on how it was questioned and edited. The scale is therefore not merely permissive. Its granularity, however, is the task as a whole. It assigns a level to an assessment rather than mapping permitted operations to stages of an authentic task, and leaves open how the resulting process evidence should enter marking criteria and how the judgement it evidences should be made individually attributable. Pilot evidence shows the scale operating at that task level (Furze et al., 2024). The narrower specification problem is the one addressed here.

Authentic Assessment in Marketing Education

Authentic assessment asks students to perform tasks resembling professional practice, judged against criteria, audiences and constraints drawn from the field (Herrington & Oliver, 2000; Villarroel et al., 2018). Ashford-Rowe et al. (2014) specify its critical elements as an ill-structured problem, a complete product and an audience beyond the assessor; Boud and Soler (2016) add that the capability should outlast the course. In marketing this is enacted principally through client-based projects, whose effects on applied skill development are reasonably well evidenced (Cummins & Johnson, 2023; Shanahan et al., 2021), a pattern Vlachopoulos and Makri (2024) confirm across disciplines while noting that implementation quality drives outcomes. Classroom studies report students partnering with industry experts on AI-driven campaign work (Richter et al., 2025) and comparing their output against generated alternatives

(Al-Fattal, 2025). These describe pedagogical designs rather than the criteria against which AI-mediated work is marked.

Generative AI unsettles this literature. Authentic assessment derives much of its warrant from the claim that a real context resists retrieval and simulation; if a generative system can produce a plausible campaign concept for a named client after a short prompt, that warrant weakens. Nieminen et al. (2023) argue that the digital dimension of authentic assessment is routinely under-designed, treated as a delivery medium rather than a constitutive feature of the practice simulated. This does not, however, tell the designer what an AI-inclusive task should require.

Evaluative Judgement Under Generative AI

The most productive theoretical response reframes the capability at stake. Bearman and Ajjawi (2023) argue that because an AI system's reasoning cannot be traced, pedagogy must shift toward judging outputs and processes rather than reconstructing method. Bearman et al. (2024) extend this into evaluative judgement, the capacity to judge the quality of work of self and others (Tai et al., 2018), explaining why a disclosure statement is inadequate: it records that AI was used without evidencing any judgement about what it produced.

Two limitations bound this contribution. It is conceptual, specifying the target capability without the task architecture that would elicit and evidence it. And evaluative judgement of AI output is disciplinary rather than general: judging a generated concept off-brand requires knowledge of the brand, audience and category.

Recent practice-grounded work identifies voice, judgement and taste as human capabilities that remain consequential in AI-mediated marketing work, judgement covering decisions about when generated outputs should be trusted, adapted or rejected (Gonsalves et al., 2026). That advances understanding of what should remain attributable to the learner. It leaves open the question addressed here: how an authentic task can require those judgements to become visible, attributable and markable.

Research Gap and Theoretical Architecture

Table 1 synthesises these strands against the use made of each here. Together they identify a gap that is architectural rather than empirical. Prohibition is unreliable; authenticity constrains quality without securing provenance; permission scales attach expectations at the level of the task; evaluative judgement names the capability without specifying the design that evidences it. What is missing is an account of how permitted AI participation becomes evidence of student judgement that an assessor can attribute and mark.

The theoretical architecture adopted here is deliberately narrow and ordered. Assessment-validity theory establishes why the artefact-to-capability inference matters and what its defensibility requires (Ali & Maroulis, 2026; Kane, 2013). Authentic assessment explains what supplies the disciplinary and professional constraints against which generated output can be judged. Evaluative judgement identifies the human capability that must remain attributable. Specified latitude, developed below, is proposed as the design property connecting them, and reflective practice supplies supporting vocabulary rather than a fourth pillar.

Table 1
Critical Synthesis of Scholarship Informing the Study

Strand	Key sources	Core claim	Limitation for the present problem	Use in this study
Assessment validity and inference	Kane (2013); Ali and Maroulis (2026)	Validity rests on the defensibility of inferences drawn from assessment evidence; generative AI destabilises the artefact-to-capability inference	Establishes the problem without specifying the design response	Supplies the central framing of the paper
Prohibition and detection	Xia et al. (2024); Miao and Holmes (2023); QAA (2023)	Detection is unreliable; redesign is recommended over prohibition	Establishes what fails without specifying what a redesign must contain	Motivates the design problem
Authenticity as an integrity safeguard	Kofinas et al. (2025)	Authentic assessments neither prevented nor revealed AI contribution	Diagnoses failure without separating quality from provenance effects	Motivates the finding on authenticity and specification
Authentic assessment design	Herrington and Oliver (2000); Ashford-Rowe et al. (2014); Villarroel et al. (2018); Boud and Soler (2016); Vlachopoulos and Makri (2024)	Specifies the elements of authenticity and of sustainable capability	Pre-dates generative AI; assumes real context resists simulation	Supplies the client-anchored task component
Client-based projects in marketing	Shanahan et al. (2021); Cummins and Johnson (2023)	Evidences applied skill development; scaling depends on design not enthusiasm	Silent on AI participation within client tasks	Establishes the disciplinary anchor
Permission specification	Perkins et al. (2024); Furze et al. (2024)	Attaches transparency, evaluation and modification expectations to levels of permitted AI use, with pilot implementation evidence	Operates through task-wide permission categories rather than mapping permitted operations to task stages, required evidence, marking criteria or individual attribution	Establishes the narrower specification problem this study addresses
Evaluative judgement and irreducibly human capability	Bearman and Ajjaw (2023); Bearman et al. (2024); Tai et al. (2018); Gonsalves et al. (2026)	Judgement of AI outputs is the capability at stake; practitioner evidence identifies voice, judgement and taste as non-delegable	Identifies what should remain attributable without specifying the architecture that evidences it	Supplies the criteria component and the outcome construct
AI and creative performance	Urban et al. (2024)	AI assistance improved creative output; metacognitive monitoring did not improve	Experimental and decontextualised	Supports the claim that generation improves more readily than judgement
Digital design in authentic assessment	Nieminen et al. (2023)	The digital is routinely under-designed within authentic assessment	Does not specify what an AI-inclusive task should require	Frames the professional-fidelity argument in the discussion
Marketing practice and curricula under AI	Guha et al. (2024); Grewal et al. (2025a, 2025b); Narang et al. (2025); Cillo and Rubera (2025)	AI is embedded in marketing practice and entering marketing curricula	Largely agenda-setting rather than design-specifying	Establishes disciplinary urgency and specificity

Note. Developed by the authors.

Methodology

Research Design and Case

The study is a reflective, documentary-led case study of a bounded instructional setting (Yin, 2018), examined longitudinally. The change between iterations is treated as a cycle of design, observation and adjustment (Schön, 1983) rather than two unrelated snapshots, and its primary evidence consists of artefacts of design rather than informant accounts. It is not design-based research: no intervention was theorised in advance and no testing protocol pre-specified.

The case is BU31013, an undergraduate module in sustainability and digital marketing, delivered in 2024/25 and 2025/26 (115 enrolled; 104 and 108 marked on the two components). BU52039, a taught postgraduate module in brand management and strategy delivered across the same years (76 enrolled; 74 marked on the report component), provides an embedded contrast in which live-client presence varies while its assessment format, weighting and word counts are held constant. In the focal undergraduate module the permitted campaign plan word count rose from 1,500 to 2,000 words between iterations. The full study population is 191 enrolled and 182 marked, and all four assessment sets were censused, giving 182 submissions, one for each marked student. One BU31013 2024/25 record was a duplicate file submitted by a student whose work is already represented in the census, and it was removed so that the census carries a single record per marked student. The two focal iterations hold module, level, teaching team, assessment format and client presence constant while the assessment architecture remained silent on AI in both. Table 2 sets out the case, its design and the evidence available.

Data Sources

Four sources are used. Module assessment review forms, briefs and marking criteria for both iterations document what was formally required and rewarded, and are authoritative for the assessment architecture. All 182 submissions across the four assessment sets were audited for documentary evidence of AI participation, using predefined search terms covering generative tools and disclosure phrasing; 178 were text-searchable, and four non-extractable files are treated conservatively as carrying no declaration. Every match was then read in context, and matches evidencing only a client brand's own AI features, or arising from reference-list URL artefacts, were excluded. AI use was never inferred from writing style, image quality or detection software. Reflective teaching records supply the rationale for design decisions and are treated as artefacts authored by parties with a stake in the account. Institutional assessment records supply mark distributions by component and iteration on the University's 23-point scale.

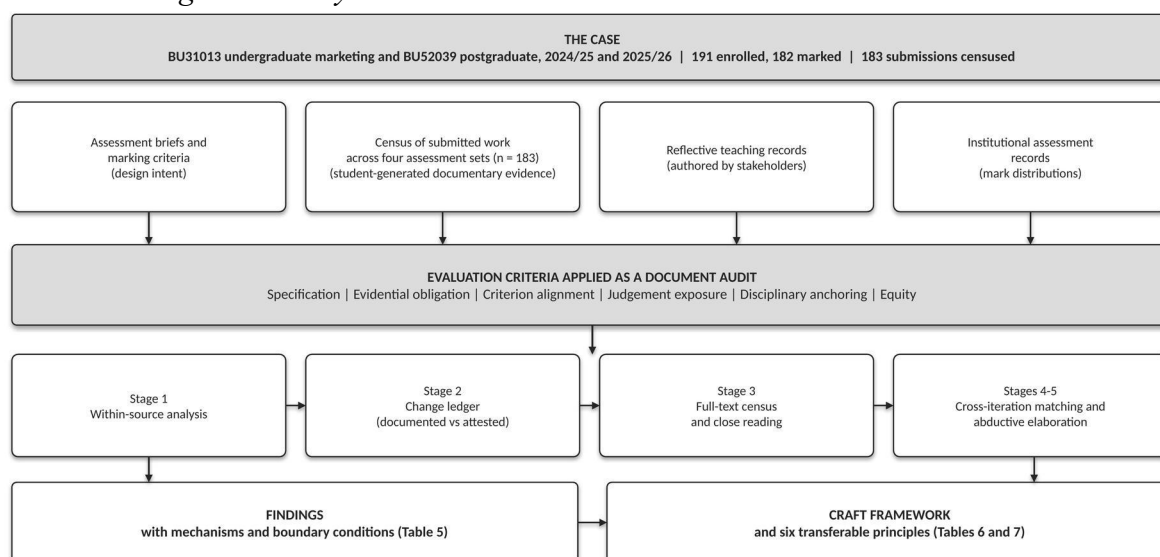
Two limits on this evidence must be stated. Student feedback was available as contextual evidence concerning the module and assessment experience, but did not specifically address generative AI; no claim about students' subjective experience of AI is therefore advanced. Second, the submission audit establishes what students documented, not what they did, and is best read as a measure of the architecture's evidential yield rather than of behaviour.

Table 2*Case Context, Assessment Architecture and Documented Change Across Iterations*

Component	2024/25	2025/26	Available evidence	Analytical relevance
Focal module	BU31013, undergraduate, sustainability and digital marketing	Unchanged	Module assessment review forms	Level, content and teaching team held constant
Live client	Intelligent Growth Solutions, a business-to-business technology company	Campy Growers, a consumer-facing sustainable growing initiative	Review forms; briefs	Client presence constant; client type varies
Assessment structure	Podcast with show notes (30%); SOSTAC campaign plan (70%)	Unchanged	Briefs; review forms	Format and weighting held constant
Permitted word count, campaign plan	1,500 words	2,000 words (plus or minus 10%)	Review forms	A documented change in task size; a confound for cross-iteration mark comparison
AI in the assessment architecture	No reference to generative AI in the review form, brief or marking criteria	Unchanged. No reference to generative AI in the review form, brief or marking criteria	Review forms signed 16 July 2024 and 7 August 2025; briefs; rubrics	The design variable of theoretical interest is absent in both iterations
Marking criteria	Five criteria: client strategy research, analysis of marketing effectiveness, proposed campaign, use of evidence, presentation	Unchanged; no criterion concerns AI	Rubrics embedded in the review forms	No credit attaches to documenting or evaluating AI participation
AI in teaching	Introduced and discussed in lectures and tutorials	Introduced and demonstrated for podcast music and campaign visual mock-ups	Teaching materials; module team record	Pedagogical preparation delivered without architectural specification
Moderation	No moderator assigned; recorded as a new course	Moderated and signed by a named moderator, 7 August 2025	Review forms	A documented change in assurance process between iterations
Embedded contrast module	BU52039, postgraduate: student-selected brand, no live client	BU52039 with a live client; AI appears as a level-of-permitted-use field outside the marking criteria	Review forms; briefs	Varies client presence with format held constant
Enrolment	Undergraduate 56; postgraduate 56	Undergraduate 59; postgraduate 20	Institutional records	Study population of 191 enrolled, 182 marked

Note. Developed by the authors from module assessment review forms, briefs and marking criteria. Client organisations are named only where their involvement is already a matter of public record.

Figure 1
Research Design and Analytical Process



Note. Developed by the authors. Arrows indicate analytical sequence.

Data Analysis

Analysis proceeded abductively in five stages (Miles et al., 2014; Patton, 2015), moving between the sensitising concepts developed above and the case material. Stage one analysed each source within itself. Stage two constructed a change ledger recording each documented alteration between iterations alongside what was held constant, classifying each entry by evidential status. Stage three audited all 182 submissions across the four assessment sets against predefined search terms, with every match subsequently examined in context. Stage four applied cross-iteration pattern matching, using the postgraduate contrast as a comparison in which a different variable moved. Stage five elaborated the framework abductively and bounded it through negative-case analysis.

Interpretation was iterative and reflexive rather than treated as a reliability exercise. The lead author conducted the analysis; co-authors acted as critical interpretive partners, challenging theme boundaries and contribution claims. Because the submission audit was a predefined documentary census rather than an interpretive coding exercise, no inter-rater reliability statistic was calculated or claimed.

Evaluation Criteria

Because assessment design cannot be evaluated without stated criteria, six dimensions were specified in advance of the cross-iteration comparison and applied to each iteration as a document audit rather than a rating of students. Specification asks whether permitted AI operations are named in the brief at the level of the task stage. Evidential obligation asks whether permitted use carries a requirement to produce a trace. Criterion alignment asks whether marking criteria name critical evaluation of AI output with band descriptors. Judgement exposure asks whether the design contains a point at which the student must select, reject or defend without AI mediation. Disciplinary anchoring asks whether the task supplies constraints against which output can be judged wrong. Equity asks whether the design assumes uniform prior AI access or capability, an assumption that survey and interview evidence on perceived risk, trust and tech-savviness gives reason to doubt (Oc et al., 2025); it is treated

throughout as a cross-cutting condition on the other five dimensions rather than as a constitutive component of specified latitude.

Rigour, Reflexivity and Ethics

Credibility, dependability and confirmability are addressed in the terms set out by Lincoln and Guba (1985). Credibility rests on triangulation across documentary, submission-level and institutional sources, and on reporting the postgraduate contrast, which does not support the study's principal design claim. Dependability rests on the change ledger, on predefined and reportable search terms, and on a stated exclusion rule applied to every match. Confirmability is limited by the authors' position as insider researchers who designed, taught and marked the assessments analysed. Ethical approval was granted by the School of Business Research Ethics Committee, University of Dundee (UoD-SoB-STAFF-2025-16). Submitted work was first audited for the presence or absence of explicit documentary evidence of AI participation; matched passages were then examined in context to characterise the depth of disclosure. No student names, identifiers or marks were linked to submissions, no student-authored text is reproduced, and results are reported only in aggregate.

Findings

Authenticity Without Specification

Both iterations of the focal case were anchored in a live external client, and in both the client relationship performed the work the authentic assessment literature predicts: it supplied brand constraints, an audience beyond the marker and a real recipient of the output. The assessment architecture, however, specified nothing about AI. The module assessment review forms signed in July 2024 and August 2025 contain no reference to generative AI, and neither the briefs nor the marking criteria name a permitted AI operation. The 2025/26 campaign plan rubric assesses research into the client's strategy, analysis of marketing effectiveness, the proposed campaign, use of evidence, and presentation. None of its five criteria concerns AI. In the postgraduate module, AI appears only as a level-of-permitted-use field referring students to institutional guidance, positioned outside the marking criteria entirely.

Authenticity and specification operate on different objects. Authenticity acts on output quality, raising the standard generic material must meet before it is credible to a client. Specification acts on provenance, determining whether the origin of that material is recoverable. A task can therefore be highly authentic and wholly unspecified, constraining what students can plausibly submit without establishing what they contributed. This offers one mechanism consistent with Kofinas et al.'s (2025) result: authenticity was not structured to do evidential work. The boundary condition is that the separation is observable only where a task is genuinely ill-structured, since in narrow tasks quality and provenance constraints are harder to distinguish.

Pedagogical Introduction Without Evidential Obligation

Generative AI was not absent from the module. It was introduced and demonstrated in lectures and tutorials, and students were shown its use for defined creative operations, including music for the podcast and visual mock-ups for the campaign. What did not follow it was any architectural requirement: no brief clause obliged students to submit generated artefacts, document prompts or account for revisions, and no criterion attached credit to doing so. Table 3 reports the result across all four assessment sets. In the demonstrated iteration, 16 of 56

submissions (28.6%) declared AI use and 10 (17.9%) named a tool. In the preceding iteration of the same module, where AI was introduced and discussed but not demonstrated for named operations, the rate was 3 of 52 (5.8%). Descriptively, the declaration rate was roughly five times higher in 2025/26, although this contrast cannot be attributed to the demonstration itself because cohort, client context and the wider AI environment also changed. Seven in ten submissions in that iteration still carried no documentary evidence from which AI participation could be established either way.

The observed pattern is consistent with a mechanism in which the absence of an evidential obligation reduces documentary visibility. A student who did not use AI and one who used it without documenting it are evidentially indistinguishable, leaving the marker no basis for the inference. Teaching students to use generative AI critically is a pedagogical act; making that criticality visible to an assessor is an architectural one, and the first does not deliver the second. Classroom interventions can measurably shift students' confidence in generated output (Beninger et al., 2025), yet such gains remain invisible to assessment unless the architecture obliges a trace. Two boundary conditions apply. Absence of a declaration evidences no documented trace, not absence of use. The keyword method also cannot detect disclosure phrased in unanticipated terms.

Table 3

Census of Documented AI Participation Across Four Assessment Sets

Assessment set	Condition at the time of submission	n	Declarations, n (%)	Highest depth of declaration reached
BU31013 2024/25	AI introduced and discussed; architecture silent	52	3 (5.8)	Declaration only
BU31013 2025/26	AI demonstrated for defined operations; architecture silent	56	16 (28.6)	Claimed evaluation, one submission
BU52039 2024/25	AI not taught; no live client; architecture silent	54	0 (0.0)	None
BU52039 2025/26	Level-of-permitted-use field in the brief, outside the marking criteria	20	3 (15.0)	Purpose
All four sets	No AI-related marking criterion in any set; one set included a task-level permitted-use field outside the criteria	182	22 (12.1)	No submission reached transformation or disciplinary justification
Excluded or non-evidentiary cases	AI as client-brand subject matter (48); reference-list URL artefact only (6); text not extractable (4)	58	–	Do not evidence the student's own use

Note. Developed by the authors. Search terms were AI-generated, generative AI, AI-assisted, artificial intelligence, AI use, AI statement, AI declaration, AI was used, used AI and AI tool, together with named tools including ChatGPT, Copilot, Midjourney, Firefly, Gemini, DeepSeek, Perplexity, Claude and Grammarly. Every match was read in context and excluded unless it evidenced the student's own use; the largest excluded category is discussion of a client brand's own AI features, which accounts for 48 submissions and, in the postgraduate 2024/25 set alone, for 35 of 54. AI use was never inferred from writing style, image quality or detection software. Because no brief or criterion required disclosure, all recorded declarations were voluntary. Four submissions could not be text-extracted and are counted conservatively as carrying none.

Disclosure Depth and the Absence of Attainment Movement

All 22 declarations across the four sets were then read in context and classified by depth. Nineteen stated a purpose and went no further; two acknowledged use without stating one. One evaluated the output, and did so by asserting that a check had been completed rather than reporting what it found. None explained how generated material had been revised or rejected,

and none justified such a decision by reference to the client, the audience, the brand or any marketing criterion. The postgraduate 2025/26 brief is the sharpest test available here, because it carried a level-of-permitted-AI-use field outside the marking criteria. No submission reproduced that field or any statement derived from it. Task-level permission without an attached evidential obligation therefore left no recoverable trace in submitted work. Table 4 reports marks on the University's 23-point scale. Means moved slightly downward between iterations, the podcast from 17.98 to 17.22 ($d = -0.31$) and the campaign plan from 16.80 to 16.48 ($d = -0.09$). The postgraduate contrast, where client presence rather than AI treatment varied, shows the larger movement, the individual report rising from 17.18 to 18.85 ($d = 0.53$).

These figures function as a negative boundary on what can be claimed rather than as a test of the framework. No AI-related criterion existed in either iteration, so the marks cannot evaluate an architecture that was never implemented. Three further changes preclude cross-iteration inference. The client changed from a business-to-business technology company to a consumer-facing organisation, which the companion analysis identifies as affecting how readily students applied taught concepts (Wu et al., 2025). The permitted word count for the campaign plan rose from 1,500 to 2,000 words. Generative tool capability also advanced materially between iterations.

Table 4

Assessment Records by Iteration and Component (23-point scale)

Module and iteration	Condition	Component	n	M	SD	Change in M	d
BU31013 2024/25	AI-silent architecture; B2B client; 1,500 words	Podcast (30%)	50	17.98	2.08	-	-
BU31013 2025/26	AI-silent architecture; consumer client; 2,000 words	Podcast (30%)	54	17.22	2.77	-0.76	-0.31
BU31013 2024/25	AI-silent architecture; B2B client; 1,500 words	Campaign plan (70%)	52	16.80	3.63	-	-
BU31013 2025/26	AI-silent architecture; consumer client; 2,000 words	Campaign plan (70%)	56	16.48	3.65	-0.32	-0.09
BU52039 2024/25	No live client	Group component (30%)	56	19.58	2.16	-	-
BU52039 2025/26	Live client	Group component (30%)	16	20.00	1.22	+0.42	0.21
BU52039 2024/25	No live client	Individual report (70%)	54	17.18	3.53	-	-
BU52039 2025/26	Live client	Individual report (70%)	20	18.85	1.67	+1.67	0.53

Note. M = mean; SD = standard deviation; d = Cohen's d on pooled standard deviations. Marks are on the University of Dundee 23-point scale; percentage conversion is not meaningful on this scale and is not reported. Effect sizes are descriptive, not tests. Cohorts are intact groups differing in size and composition, and no causal inference is drawn. Source: institutional assessment records. Neither undergraduate iteration contained an AI-related criterion; the condition column records what did change.

Table 5*Findings, Evidence, Mechanisms and Boundary Conditions*

Finding	Empirical basis	Proposed mechanism	Status	Boundary condition
Authenticity supplies disciplinary constraints without specifying AI participation	Both iterations client-anchored; neither review form, brief nor rubric refers to generative AI	Authenticity acts on output quality; specification acts on provenance. The two are separable properties of a task	Observed	Separable only where the task is genuinely ill-structured
Pedagogical introduction does not itself create an evidential obligation	AI introduced and demonstrated for podcast music and campaign visuals; no brief clause or criterion requires a trace	Teaching criticality is a pedagogical act; making it visible to an assessor is an architectural one	Observed	Teaching record rests partly on the module team's account
Voluntary traces yield limited and uneven assessment evidence	Census of 182 submissions across four sets: declaration rates of 5.8%, 28.6%, 0% and 15.0%; 160 (87.9%) carry no declaration	Where evidence is neither required nor rewarded, the assessment yields limited documentary traces	Observed	Absence of a trace is not absence of use; extent of actual AI participation is unknown
Declarations evidence use but not evaluative judgement	Of 22 declarations, 19 state a purpose only, 2 acknowledge use only, 1 evaluates the output, 0 explain revision, 0 justify it on disciplinary grounds	A disclosure regime without an evidential obligation produces traces that record use without evidencing judgement about it	Observed; full-text classification of every matched passage in all four sets	Classification rests on what students chose to write; a declaration may understate actual use
Attainment data cannot evaluate an architecture that was not implemented	Podcast $d = -0.31$; campaign plan $d = -0.09$; postgraduate report $d = 0.53$ where client presence varied	No AI-related criterion existed in either iteration, so no criterion-linked attainment change was predicted	Negative boundary on claims	Client, permitted word count, tool capability and cohort composition all changed

Note. Developed by the authors. Effect sizes are descriptive, not tests. Status distinguishes what the case observed from what remains to be verified.

Discussion

Theoretical Interpretation

The findings address the first question diagnostically: where AI is introduced and demonstrated but not specified, AI-mediated contribution is largely unrecoverable from submitted work. The second is answered as a boundary rather than an effect: outcomes changed little across the undergraduate iterations, but with no AI-related criterion in either year they cannot evaluate specified latitude. The third is addressed prospectively through CRAFT, which states the conditions under which permitted AI participation could be rendered visible, attributable and assessable.

The theoretical significance lies in two separations. First, authenticity supplies disciplinary, client and audience constraints against which generated output can be judged, but those constraints do not tell an assessor which part of the production process belongs to the student. This refines authentic-assessment theory rather than rejecting it, and explains why Kofinas et al. (2025) could find authentic assessments simultaneously demanding and unauditable. Second, pedagogical preparation and evidential architecture are distinct: students were taught

and shown AI use, yet fewer than three in ten left any trace, under an architecture in which nothing required one.

This changes what authenticity means under AI conditions. An authentic marketing task that excludes generative AI is an increasingly incomplete representation of professional practice, extending Nieminen et al.'s (2023) argument from under-designed digital media to professional fidelity. Authenticity and integrity are not in tension as the prohibition literature assumes; they diverge only when authenticity is asked to do evidential work it was never structured to perform. Evaluative judgement also does not follow from AI access: judgement requires something against which output can be judged wrong, and the client's brand, audience and category constraints supply it.

The CRAFT Framework

The framework proposed here is organised around a single construct. Specified latitude is the task-stage-specific mapping of permitted AI operations to the evidence of the student's own evaluative judgement that those operations must leave behind. So defined it is not a synonym for clearer instructions, nor for a task-level permission category: instructions and permission levels govern what a student may do, whereas specified latitude governs, operation by operation, what the exercise of that permission must make visible and how that visibility enters marking. CRAFT is therefore advanced as a design response to a documented architectural absence, not as a description of an implemented intervention, and its components are stated so that they can be tested and challenged.

CRAFT names five components through which specified latitude is realised. Client-anchored task supplies the disciplinary constraints against which generated output can be judged inadequate; the component is a disciplinary constraint anchor, instantiated here as a live marketing client, with patients, legal fact patterns, engineering specifications, datasets and laboratory conditions as analogous anchors elsewhere. Rule-specified latitude names permitted AI operations by task stage rather than by blanket permission. Attributable trace obliges the exercise of that latitude to leave submitted evidence, such as the generated artefacts alongside the account of their revision. Forced judgement point locates an unmediated moment at which the student selects, rejects and defends. Tuned criteria make critical evaluation of AI output an assessed dimension with band descriptors. The components are interdependent, and the framework's substantive claim concerns their interaction: each fails in a statable way without the others. Table 6 sets out each component with the evidence that would challenge its role.

Three propositions follow, stated so that CRAFT can be falsified rather than merely adopted. First, disciplinary authenticity increases the constraints against which generated output can be judged, but does not itself increase the observability of AI-mediated contribution. Second, permitted AI participation becomes assessable only where task-stage-specific latitude is coupled with an attributable evidential trace and with aligned marking criteria. Third, an individually attributable judgement point strengthens assurance only where it is anchored in disciplinary constraints and connected to documented process evidence. The first is consistent with what this case observed; the second and third are prospective. Equity bounds all three rather than sitting within any one of them: trace requirements, unequal prior AI proficiency and individually performed defence impose unequal burdens, so each component carries an accessibility and support condition (Oc et al., 2025).

The Exposure Point as a Proposed Assurance Mechanism

The design pairs a written output carrying strategic rationale with a live, student-performed client-facing pitch. Its assurance value is proposed rather than demonstrated: records establish that a pitch occurred, not that individuals defended particular decisions, and group formats complicate attribution. The proposed mechanism is that a defensible account of a creative decision is difficult to hold without having made it, whereas a polished artefact is not.

Locating assurance at such a point rather than in the artefact would align the design with the argument that pedagogy must judge outputs and processes rather than reconstruct method (Bearman & Ajjawi, 2023). Disclosure alone cannot substitute: a disclosure statement is itself an artefact subject to the same generation problem as the work it describes.

Four failure conditions follow. Latitude specified without an attributable trace restores clarity about rules while leaving provenance unrecoverable, the limitation of permission scales used alone. A trace obliged without tuned criteria is submitted but not rewarded, and students economise on it. Criteria tuned without a client anchor leave markers with no disciplinary standard against which evaluation can itself be assessed. A forced judgement point without the other four becomes an integrity check, since there is no documented development to defend.

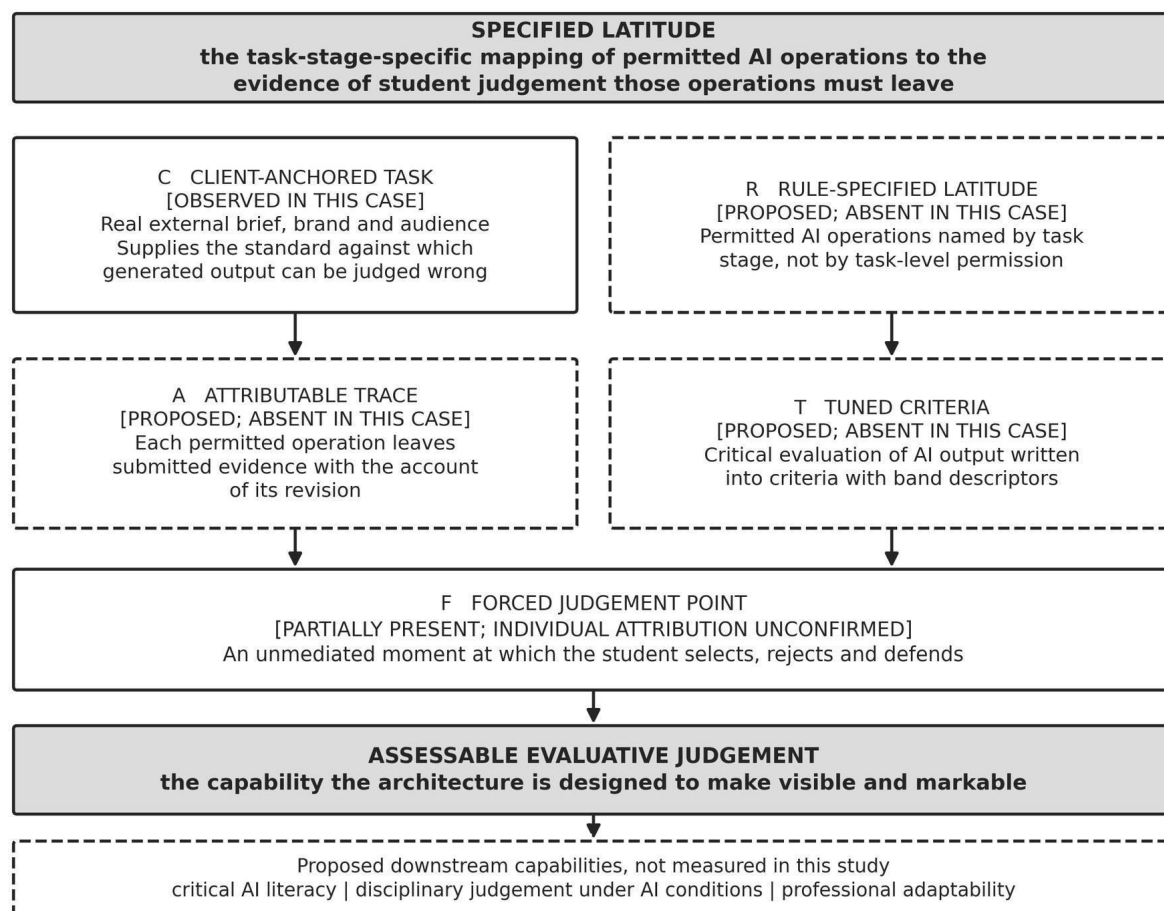
Table 6*CRAFT Components, Status in This Case, and Disconfirming Conditions*

Component	Design action	Status in this case	Empirical basis or absence	Evidence that would challenge this role
C: Client-anchored task	Set the brief with a real external organisation, its constraints and its audience	Observed	Documented in both focal undergraduate iterations; present in BU52039 2025/26 but absent in BU52039 2024/25	Simulated clients with equivalent constraints producing the same evaluative demand
R: Rule-specified latitude	Name permitted AI operations by task stage rather than granting task-level permission	Absent; theoretically proposed	No review form, brief or rubric refers to generative AI in either iteration; where a permitted-use field did exist, in the postgraduate 2025/26 brief, no submission reproduced it	Unspecified tasks yielding equally recoverable accounts of student contribution
A: Attributable trace	Oblige each permitted operation to leave submitted evidence alongside the account of its revision	Absent as a requirement; theoretically proposed	No brief clause requires generated artefacts, prompts or revision accounts; 160 of 182 submissions (87.9%) carry no explicit AI-use declaration	Trace requirements fail if assessors cannot distinguish authentic process evidence from AI-generated accounts of that process
F: Forced judgement point	Locate an unmediated, individually attributable moment at which the student selects, rejects and defends	Partially present; individual attribution unconfirmed	A live client-facing pitch occurred in both iterations, but records do not establish that individuals defended particular decisions	Written reflection alone producing equivalent assurance
T: Tuned criteria	Write critical evaluation of AI output into marking criteria with descriptors at each band	Absent; theoretically proposed	The five campaign plan criteria concern client research, analysis, the proposed campaign, evidence and presentation; none concerns AI	Documented traces attracting equivalent student effort without criterion weighting

Note. Developed by the authors. The case diagnoses the architectural absence; CRAFT proposes the remedy. Only the client anchor was present in the observed iterations, so the framework is advanced for prospective testing rather than as an evaluated intervention.

Figure 2

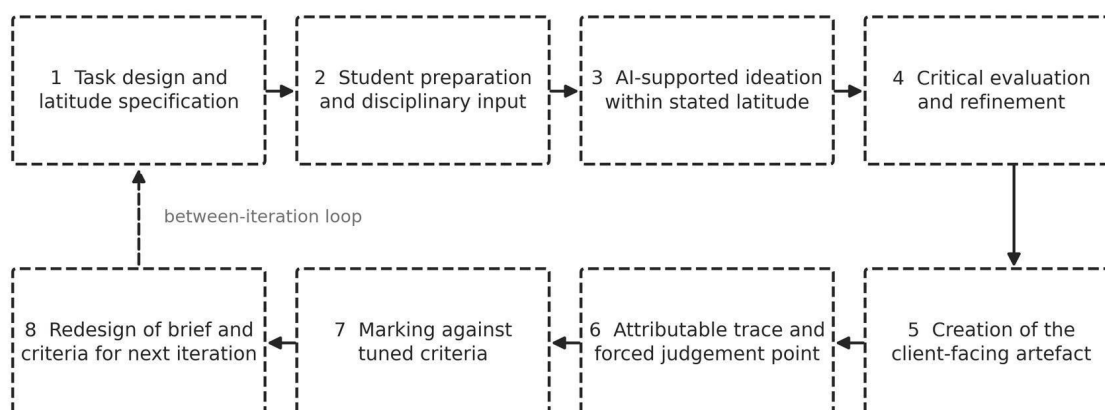
Specified Latitude and the CRAFT Framework: Status of Each Component in This Case



Note. Developed by the authors. Components are interdependent rather than sequential; failure conditions are specified in Table 6.

Figure 3

Proposed AI-Supported Assessment Redesign Lifecycle



Note. Developed by the authors. Dashed outlines indicate that the lifecycles is the prospective application of CRAFT derived from the diagnostic case. It does not describe an intervention implemented across the two cohorts analysed here.

Transferable Design Principles

Six principles follow. Table 7 states each with its design action, rationale, intended outcome, the risk it addresses and its boundary condition; the binding constraints are task structure, marker capability, criteria written at band level and staff time.

Table 7

Transferable Principles for Designing AI-Supported Assessment

Principle	Design action	Theoretical rationale	Intended outcome	Risk addressed and boundary
1. Begin from a disciplinary problem that supplies constraints	Anchor the task in an external organisation with a brand, an audience and a category	Authentic assessment supplies the standard against which output can be judged inadequate (Villarroel et al., 2018)	Generated output becomes judgeable rather than merely plausible	Addresses generic AI-default output. Fails where the task admits a narrow range of acceptable answers
2. Specify latitude by task stage, not by blanket permission	State which AI operations are permitted at ideation, at production and at refinement	Extends permission scales (Perkins et al., 2024) from level to operation	Ambiguity is removed without prohibition	Addresses undisclosed use arising from unstated scope. Requires briefs to be rewritten, not annotated
3. Attach an evidential obligation to every permission granted	Require generated artefacts to be submitted alongside the account of their revision	Repairs the assessment inference that artefact quality no longer supports	Provenance becomes a submitted object	Addresses unrecoverable authorship. Requires marker capability to judge the account
4. Mark the documented distance, not the polish of the artefact	Write critical evaluation of AI output into criteria with descriptors at each band	Operationalises evaluative judgement of AI outputs (Bearman et al., 2024)	The target capability becomes markable and is signalled to students	Addresses reward for undisclosed generation. Presupposes criteria written at band level
5. Locate an individually attributable exposure point	Pair the documented written output with a live, student-performed defence	Assurance shifts from product originality to accountable judgement (Bearman and Ajjawi, 2023)	Students must hold the reasoning they submit	Addresses outsourced authorship. Staff time cost scales with cohort size; group formats complicate attribution
6. Treat specification as an assessability intervention	Justify redesign by what it makes evidenceable, not by expected marks	Follows from the diagnostic boundary established by the assessment records: attainment cannot evaluate an AI-assessment architecture that is absent from the criteria	Realistic expectations and defensible evaluation of the redesign	Addresses overclaiming. Most likely to be resisted, since it offers no immediate metric improvement

Note. Developed by the authors from the findings reported above. Principles are stated with boundary conditions; transferability depends on those conditions holding in the adopting context.

Contributions and Implications

Theoretically, the study makes three connected claims. It distinguishes disciplinary authenticity from evidential assessability: authentic client constraints supply standards against which generated output is judged, without making the student's AI-mediated contribution visible. It distinguishes pedagogical preparation from assessment architecture: AI was taught and demonstrated, yet without evidential obligations little documentary visibility followed. And it develops specified latitude, the task-stage-specific connection between permitted AI operations and the evidence of judgement they must leave, operationalised through CRAFT and stated as

three falsifiable propositions. The contribution is one of theory refinement: specified latitude extends authentic-assessment and evaluative-judgement accounts by specifying the evidential conditions under which AI-mediated work can support defensible inferences about student capability.

Empirically, the study contributes a four-set census of 182 submissions in which pedagogical introduction of generative AI coexisted with its absence from the assessment architecture. Its most useful result is negative: across two module codes and two years, no submission evidenced revision or disciplinary justification of generated output, and the iteration carrying an explicit permitted-use field left no trace of it.

Methodologically, the study shows that iterative assessment design can be analysed through a change ledger distinguishing documented from attested change, combined with a documentary census of submitted work against predefined terms. Pedagogically, it supplies six principles and a framework whose disconfirming conditions are stated.

For policy, three implications follow. Task-level permission guidance is insufficient where validity depends on making judgement attributable; guidance should let module teams map permission to discipline-specific evidential requirements. Integrity procedures premised on artefact originality address a property inspection cannot reliably establish, and are better reoriented toward whether a student can account for documented development. Moderation needs criteria naming critical evaluation of AI output, and markers equipped to apply them.

Limitations and Future Research

Four limitations bound interpretation. The audit establishes documented AI participation, not actual use; declarations were classified from what students wrote, and a submission may understate its own AI use, as one submission containing an apparent unedited generator artefact suggests. Four submissions in a format the extraction pipeline could not open are counted conservatively as carrying no declaration. Second, no AI-specific student feedback exists, so no claim about students' experience of AI is advanced. Third, the authors designed, taught and marked the assessments analysed. Fourth, because no element of the proposed architecture beyond the client anchor was implemented, the case can diagnose the gap but cannot evaluate the remedy; a prospective implementation in which the remaining components are built into the assessment and tested is the study this one indicates.

Four lines of work follow: the student-voice evidence this study lacks, through purpose-designed reflective instruments; multi-institutional implementation testing the three propositions by teams holding no stake in the design; validated measures of evaluative judgement, since attainment on existing criteria is the wrong instrument; and whether specified latitude advantages students with prior AI fluency.

Conclusion

Generative AI has weakened the inference from artefact to capability on which marketing assessment depended, and the defensibility of that inference is what assessment validity consists in (Ali & Maroulis, 2026; Kane, 2013). This study examined two iterations of a client-anchored undergraduate marketing module, with a postgraduate contrast, and found that client authenticity supplied constraints against which generated output could be judged while the assessment architecture specified no AI participation and carried no AI-related criterion in

either year. Generative AI was nonetheless taught and demonstrated for defined creative operations, yet fewer than three in ten submitted campaign plans contained an explicit AI-use trace. Mean marks in the focal undergraduate module did not rise, but because no AI criterion existed they cannot evaluate an architecture that was never implemented. The paper proposes specified latitude, the task-stage-specific connection between permitted AI operations and the evidence of human judgement those operations must leave, operationalised through CRAFT. The case identifies a potentially transferable design problem requiring testing beyond this setting.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, OpenAI's ChatGPT (GPT-5.3) was used only to support language refinement and improve the clarity, concision and readability of text drafted by the authors, including the presentation of the research findings. The tool was not used to generate, analyse or interpret the study data, or to determine the findings or conclusions. All AI-assisted revisions were subsequently reviewed, verified and edited by the authors, who take full responsibility for the final content of the manuscript.

References

- Al-Fattal, A. (2025). You do it, AI does it, you compare and reflect: Exploring reflective learning with generative AI in principles of marketing. *Journal of Marketing Education*. Advance online publication. <https://doi.org/10.1177/02734753251356691>
- Ali, R., & Maroulis, J. (2026). Assessment validity in the age of generative artificial intelligence: A critical review. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2026.2683046>
- Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205–222. <https://doi.org/10.1080/02602938.2013.819566>
- Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, 54(5), 1160–1173. <https://doi.org/10.1111/bjet.13337>
- 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>
- Beninger, S., Reppel, A., Stanton, J., & Watson, F. (2025). Facilitating generative AI literacy in the face of evolving technology: Interventions in marketing classrooms. *Journal of Marketing Education*, 47(2), 112–125. <https://doi.org/10.1177/02734753251316569>
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413. <https://doi.org/10.1080/02602938.2015.1018133>
- Cillo, P., & Rubera, G. (2025). Generative AI in innovation and marketing processes: A roadmap of research opportunities. *Journal of the Academy of Marketing Science*, 53(3), 684–701. <https://doi.org/10.1007/s11747-024-01044-7>
- Cummins, S., & Johnson, J. S. (2023). The impact of live cases on student skill development in marketing courses. *Journal of Marketing Education*, 45(1), 55–69. <https://doi.org/10.1177/02734753211034553>
- Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, 40(4), 38–55. <https://doi.org/10.14742/ajet.9434>
- Gonsalves, C., Van Vuren, S., & Gill-Simmen, L. (2026). Assessing the irreducibly human in marketing education: A voice, judgement and taste framework for the generative AI era. *Journal of Marketing Education*. Advance online publication. <https://doi.org/10.1177/02734753261454375>
- Grewal, D., Guha, A., Saturnino, C. B., & Becker, M. (2025a). The future of marketing and marketing education. *Journal of Marketing Education*, 47(1), 61–77. <https://doi.org/10.1177/02734753241269838>

- Grewal, D., Saturnino, C. B., Davenport, T., & Guha, A. (2025b). How generative AI is shaping the future of marketing. *Journal of the Academy of Marketing Science*, 53, 702–722. <https://doi.org/10.1007/s11747-024-01064-3>
- Guha, A., Grewal, D., & Atlas, S. M. (2024). Generative AI and marketing education: What the future holds. *Journal of Marketing Education*, 46(1), 6–17. <https://doi.org/10.1177/02734753231215436>
- Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development*, 48(3), 23–48. <https://doi.org/10.1007/BF02319856>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kofinas, A. K., Tsay, C. H.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56(6), 2522–2549. <https://doi.org/10.1111/bjet.13585>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.
- Narang, U., Sachdev, V., & Liu, R. (2025). When AI wears many hats: The role of generative artificial intelligence in marketing education. *Journal of Public Policy & Marketing*, 44(3), 473–489. <https://doi.org/10.1177/07439156251328237>
- Nieminen, J. H., Bearman, M., & Ajjawi, R. (2023). Designing the digital in authentic assessment: Is it fit for purpose? *Assessment & Evaluation in Higher Education*, 48(4), 529–543. <https://doi.org/10.1080/02602938.2022.2089627>
- Oc, Y., Gonsalves, C., & Quamina, L. T. (2025). Generative AI in higher education assessments: Examining risk and tech-savviness on student's adoption. *Journal of Marketing Education*, 47(2), 138–155. <https://doi.org/10.1177/02734753241302459>
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage.
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The artificial intelligence assessment scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6). <https://doi.org/10.53761/q3azde36>
- Quality Assurance Agency for Higher Education. (2023). *Reconsidering assessment for the ChatGPT era: QAA advice on developing sustainable assessment strategies*. QAA.

- Richter, S., Giroux, M., Piven, I., Sima, H., & Dodd, P. (2025). A constructivist approach to integrating AI in marketing education: Bridging theory and practice. *Journal of Marketing Education*, 47(2), 94–111. <https://doi.org/10.1177/02734753241288876>
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Shanahan, D. E., Palmer, L. H., & Salas, J. (2021). Achieving scaled and sustained use of client-based projects in business school marketing education: A proposed suprastructure. *Journal of Marketing Education*, 43(1), 59–74. <https://doi.org/10.1177/0273475319881179>
- 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>
- Urban, M., Dechterenko, F., Lukavsky, J., Hrabalova, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, 215, Article 105031. <https://doi.org/10.1016/j.compedu.2024.105031>
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840–854. <https://doi.org/10.1080/02602938.2017.1412396>
- Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83, Article 101425. <https://doi.org/10.1016/j.stueduc.2024.101425>
- Wu, Y., Ross, A., & Stewart, M. (2025). AI-enabled industry-collaborative assessment (AI-ICAM): Advancing pedagogy and university-industry collaboration. In *The European Conference on Education 2025: Official conference proceedings* (pp. 467–487). IAFOR. <https://doi.org/10.22492/issn.2188-1162.2025.38>
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21, Article 40. <https://doi.org/10.1186/s41239-024-00468-z>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.

Contact email: ywu002@dundee.ac.uk