Thayer-inspired teaching routines as active learning in higher business education

Paulo Botelho Pires [1], Helena Costa Oliveira [1*], Isabel Maldonado [1], and José Duarte Santos [1]

[1] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal
Email: paulo.pires@iscap.ipp.pt (P.B.P.); helena@iscap.ipp.pt (H.C.O.); imaldonado@iscap.ipp.pt (I.M.) and jdsantos@iscap.ipp.pt (J.D.S.)

[*]Corresponding author

***Abstract* — This paper reports an embedded case study of a Thayer-inspired teaching routine (M1–M2–M3) implemented across four business curricular units (covering global competitiveness, strategic marketing, international marketing, and B2B marketing) during the 2025–2026 academic year. The routine operationalises (M1) a 10–20 minutes conceptual framing with explicit learning outcomes; (M2) small-group (≤3 students) applied problem work using permitted digital tools (including AI under disclosure and verification rules); and (M3) an unsupported oral defence (no slides, notes, or devices), with mandatory individual contribution and structured questioning. Drawing on evidence on active learning effectiveness and constructive alignment, the M1–M2–M3 routine is conceptualised as integrity-by-design: students may use tools to learn and build artefacts (M2), while their individual understanding and reasoning are authenticated through a technology-free defence (M3). Across the four embedded units of analysis, findings report a trajectory from initial student discomfort and insufficient preparation to progressively stronger conceptual accuracy and argumentation. The paper documents the positive learning outcomes achieved and the structural challenges encountered (including anxiety, overload, and unequal academic capital) and discusses their implications for assessment design in AI-normalised environments.**

***Keywords* — Thayer method; active learning; constructive alignment; authentic assessment; oral defence; academic integrity; generative AI; assessment equity**

I. Introduction

Generative AI has amplified long-standing concerns around authorship, originality, and the authenticity of assessment evidence in higher education. In response, educators face a dual design problem: enabling students to use contemporary tools responsibly while ensuring that assessment evidence still reflects individual learning.

One promising response to this challenge lies in the redesign of teaching and assessment practices according to integrity-by-design principles, in which learning activities may incorporate digital tools while assessment mechanisms verify individual reasoning and ownership of knowledge. Within this context, the Thayer method, which emphasises student preparation, active engagement, and demonstration of understanding through questioning and discussion, offers a potentially valuable pedagogical framework for contemporary higher education environments.

This paper examines the implementation of a Thayer-inspired teaching routine (M1–M2–M3) designed to reconcile active learning with integrity-by-design assessment. The routine consists of three sequential phases: (M1) a brief conceptual framing with explicit learning outcomes; (M2) small-group applied work in which students may use digital tools, including generative AI, under explicit disclosure and verification rules; and (M3) an unsupported oral defence in which students must individually demonstrate their reasoning without the aid of slides, notes, or digital devices.

The study investigates how these routine functions work when implemented across multiple curricular contexts. Specifically, the intervention was applied during the 2025/2026 academic year across four business curricular units (global competitiveness, strategic marketing, international marketing, and B2B marketing) within a Portuguese public higher education institution. The aim is to explore how the M1–M2–M3 routine operates as a mechanism for combining active learning, authentic assessment, and academic integrity in an AI-normalised educational environment.

The study adopts an embedded case study design, treating the four curricular units as subcases within a single instructional intervention. Evidence is triangulated from assessment records, rubric-based oral defence scores, learning artefacts produced during group work, instructor observation notes, and student perception data collected through questionnaires. This design allows for within-unit analysis and cross-unit comparison to identify recurring patterns of learning outcomes, student adaptation processes, and structural challenges associated with the routine.

The findings reveal a clear trajectory of adaptation over time. Initial implementation generated significant student discomfort, particularly due to the high-accountability nature of unsupported oral defence and the requirement for autonomous preparation. However, as students became accustomed to the M1–M2–M3 cycle, improvements were observed in conceptual accuracy, argumentation quality, and the ability to link theoretical frameworks to applied decisions. At the same time, the study documents several structural tensions (including anxiety, workload pressures, and unequal academic preparation) that shape how such accountability-centred designs function in practice.

The intervention is conceived as integrity-by-design: M2 supports authentic tool-mediated practice, while M3 secures evidence of student ownership through real-time reasoning and explanation. Implementation, however, proved neither frictionless nor uniform.

The paper contributes with a replicable pedagogical routine that integrates active learning with integrity-by-design assessment in AI-permissive environments. It

https://doi.org/10.34630/cip.vi.7677

advances a critical discussion of structural conditions (such as academic capital, workload distribution, and assessment design) that influence the effectiveness and equity of accountability-centred pedagogical models. This paper documents the achievements and the structural tensions encountered, contributing a nuanced account of how Thayer-inspired accountability design functions in practice.

The remainder of the paper is structured as follows. Section 2 presents the literature review; Section 3 presents the research methodology and evidence sources. Section 4 reports the results and Section 5 discusses the implications of the findings for pedagogy and assessment design in AI-normalised higher education. Finally, section 6 presents the conclusions, limitations, and suggestions for future research.

II. Literature Review

This section reviews literature on the Thayer method, active learning approaches, and constructive alignment, with particular attention to their implications for assessment in contexts where artificial intelligence tools are increasingly used.

The Thayer method emphasises student responsibility for preparation, with classroom time centred on demonstration and intensive questioning rather than extended lecturing [1, 2]. Multi-course adaptations in higher education share features such as scheduled preparation, minimal lecturing, maximised active work, and frequent evaluation [3]. This logic aligns with evidence that active learning outperforms lecture-centric instruction in achievement and retention [4, 5]. However, the method's effectiveness depends critically on the quality of implementation, student readiness, and contextual supports [6].

Constructive alignment [7] proposes that intended learning outcomes, teaching-learning activities, and assessment tasks should cohere so that assessment directly evidences the targeted outcomes. Assessment-for-learning emphasises feedback loops as levers for learning gains [8]. Feedback supports self-regulation when it clarifies standards, enables self-evaluation, and promotes action [9]. In M1–M2–M3, alignment is operationalised as: explicit criteria (M1) → deliberate practice with feedback (M2) → authentic demonstration (M3).

Authentic assessment evaluates the application of knowledge, judgement, and communication in meaningful contexts [10, 11]. Oral defence formats can serve as authentic demonstrations of reasoning and as validation mechanisms for ownership when written outputs may be tool-assisted [12]. However, systematic reviews identify reliability and equity risks, including performance anxiety, language effects, and assessor bias, that require rubric transparency, calibration, and scaffolding [12]. In AI-normalised contexts, integrity-by-design combines tool-enabled practice with robust demonstration requirements [13, 14, 15].

A critical reading of Thayer-inspired approaches must acknowledge the structural tensions inherent in transferring full preparation responsibility to students. Drawing on [16] zone of proximal development, designs that demand independent first contact with new material before instructor mediation risk requiring students to operate beyond their current cognitive reach, generating anxiety rather than curiosity [17]. The emphasis on textual preparation also presupposes academic literacy capital that is unequally distributed across student populations [18], potentially amplifying rather than equalising existing disparities. Moreover, graded participation requirements risk producing instrumental rather than genuine engagement, substituting performative compliance for authentic intellectual risk-taking.

III. Method

A. Purpose and Research Design

The study adopts a single-case design with four embedded units of analysis, enabling within-unit description and cross-unit synthesis. This design supports analytic generalisation through pattern matching and triangulation rather than statistical generalisation [19]. Each curricular unit is analysed as a subcase and then compared to identify replicated mechanisms and boundary conditions. Within this framework, the following research questions (RQs) guide the empirical analysis:

RQ1. How can a Thayer-inspired M1–M2–M3 routine be implemented with constructive alignment across four business curricular units?

RQ2. What learning evidence and student perceptions (including challenges and adaptations) are observed within and across units under this routine?

RQ3. How does combining tool-permitted applied work (M2) with an unsupported oral defence (M3) shape perceived authenticity, integrity, and equity?

The case is situated in a Portuguese public higher education institution (School of Business Studies), academic year 2025–2026. The four business curricular units are treated as embedded units of analysis (see Table 1), spanning undergraduate and postgraduate levels. Table 1 describes the four units and their assessment structures.

Table 1. Embedded business units of analysis and assessment overview.

Business Curricular Unit	Cycle / ECTS	Thematic Focus	Key Assessment Evidence
Business Unit A (Global Competitiveness)	MSc / 6 ECTS	Globalisation, competitiveness, institutional and ethical issues	M3 oral defence; group project; participation
Business Unit B (Strategic Marketing)	MSc / 6 ECTS	Strategic analysis and marketing strategy formulation	M3 oral defence; group strategic plan; participation
Business Unit C (International Marketing)	BSc / 6 ECTS	Market selection, entry modes, global marketing mix	M3 oral defence; group project; individual test; participation
Business Unit D (B2B Marketing)	BSc / 5 ECTS	Organisational buying, relational value, key accounts, B2B digital channels	M3 oral defence; group work; participation; exam

All units share the same M1–M2–M3 session routine and an assessment approach that combines applied group work with an unsupported oral defence. Routine M1 (10–20 minutes) consists of a brief exposition of essential concepts,

explicit learning outcomes, and success criteria aligned to the M3 rubric. Typical misconceptions are addressed proactively. Routine M2 (≤3 students per group) applied tasks (case studies, decision matrices, planning, and diagnostic exercises) with free but accountable use of digital tools (LMS, collaborative suites, analytics, visualisation, simulations, and digital bibliography). Generative AI may be used as a learning aid provided that use is disclosed, outputs are verified, and final decisions are justified with independent reasoning. Routine M3 (unsupported oral defence) consists of oral presentation and Q&A without slides, notes, or devices. Each student must contribute individually and respond to directed questioning that probes conceptual accuracy, application, and argumentation. Formative feedback, including a brief peer review, is provided at the end of each session.

B. Data Analysis

Qualitative material — open-ended questionnaire responses, instructor session-observation notes, and analytical annotations on the M2 learning artefacts — was analysed using reflexive thematic analysis following the six-phase procedure of Braun and Clarke [21]. The corpus was first read repeatedly for familiarisation, with analytic memos recording initial impressions. Codes were then generated inductively, staying close to participants' own wording, and collated into candidate themes through iterative comparison within each unit and across the four units. Candidate themes were reviewed against both the coded extracts and the full dataset to verify internal coherence and external distinctiveness, and were subsequently refined, defined, and named in relation to the three research questions. To strengthen dependability, coding was performed independently by two members of the research team and reconciled through discussion; disagreements were resolved by returning to the source extracts rather than by majority decision, and an audit trail of coding decisions and theme revisions was retained. Finally, the qualitative themes were integrated with the quantitative strand in a joint display [22], in which thematic findings were juxtaposed with the descriptive rubric and assessment patterns for each unit, allowing convergence and divergence between perceived and demonstrated learning to be examined. Quantitative data (final and component grades and M3 rubric dimension scores) were summarised descriptively at the unit level and synthesised across units; consistent with the embedded single-case design, these summaries are interpreted as analytic, non-causal patterns rather than as statistical effects [19]. Table 2 details the evidence matrix.

Table 2. Evidence sources and planned analyses.

Evidence Source	Construct Focus	Instrument/ Measure	Timing	Planned Analysis
Assessment records	Achievement; progression	Final and component grades	End of term	Descriptive patterns by unit; cross-unit comparison (non-causal)
M3 rubric scores	Reasoning; communication; ownership	Rubric with criteria and anchors	Sessional / end-of-term	Dimension-level profiles; inter-rater checks if feasible; cross-unit synthesis
Learning artefacts (M2)	Decision quality; evidence use	Artefact scoring + content analysis	Throughout term	Coding of decision logic; exemplar analysis; alignment to LOs
Student perceptions	Perceived learning; fairness; integrity climate	Survey (Likert + open text)	Mid/end of term	Thematic analysis; joint display with rubric outcomes
LMS traces (optional)	Engagement; pacing proxies	Access/submission timestamps	Continuous	Exploratory associations with outcomes (privacy-limited)

Student perceptions were collected through an end-of-term questionnaire administered online via the institutional learning management system during the final weeks of the semester. The instrument combined closed items rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree), grouped into thematic blocks – perceived learning and conceptual understanding, perceived fairness and integrity climate, and affective experience (anxiety, workload, and confidence) – and open-ended prompts asking students to describe what most helped and hindered their learning across the M1–M2–M3 routine, their experience of the unsupported M3 oral defence, and any suggestions for improvement. Completion was voluntary, and responses were anonymous. Across the four units, 47 of 149 enrolled students completed the questionnaire. The open-ended items produced free-text comments, which constituted the primary corpus for the thematic analysis reported in Section 4. The full questionnaire, the coding frame, and the anonymised response set are available from the authors upon request.

C. Ethics, Data Protection, and AI Disclosure

The study was conducted under institutional norms for educational evaluation and research ethics. Student data were minimised and processed in line with the General Data Protection Regulation [20]. AI tool use was permitted in M2 under explicit disclosure and verification expectations; M3 provided a technology-free setting for confirming individual understanding and reducing the risk of misattributed authorship.

IV. RESULTS

Results are reported across two dimensions: a trajectory of learning gains over the term, and a systematic account of the structural challenges encountered. This dual structure avoids confirmation bias by ensuring that difficulties are documented with the same rigour as achievements.

Implementation across all four units revealed a consistent initial phase of student resistance and distress that must be acknowledged before reporting eventual outcomes. Two patterns dominated this early period:

- **Challenge 1:** Anxiety under direct accountability: Students across units expressed significant discomfort with the no-support constraint of M3.

The absence of slides, notes, or devices, combined with direct instructor questioning, created acute performance stress, particularly at the undergraduate level. Students accustomed to lecture formats reported feeling exposed and unprepared in a way they described as qualitatively different from written assessments.

- **Challenge 2:** Insufficient autonomous preparation: In early sessions, students were consistently underprepared for M1 and M2, treating theoretical content as optional or deferrable. The design assumption that students would arrive having engaged substantively with preparation materials was not initially met. This produced weak M2 outputs and thin M3 responses, amplifying student anxiety and creating a negative feedback loop.

These early dynamics reflect the paradox of accountability-centred designs: the conditions that make oral defence an effective integrity mechanism are precisely those that generate the highest affective load when students are not yet acculturated to the routine.

Beyond the initial adaptation phase, several structural challenges persisted across units to varying degrees. Table 3 documents these with their observed manifestations and the mitigations applied.

Table 3. Structural challenges, observed manifestations, and mitigations applied across the four business units.

Challenge	Observed Manifestation	Mitigation Applied
Anxiety and exposure stress (M3)	Students distressed by no-support constraint; discomfort with direct instructor questioning without slides or notes	Low-stakes rehearsal sessions; calibration exemplars shared in advance; rubric made fully transparent from Week 1
Insufficient prior preparation	Students arrived at sessions unprepared, treating theoretical content as optional	Preparation guides issued weekly; preparation quality incorporated into participation component; M1 criteria made explicit
Surface-level participation	Graded participation produced formulaic, non-generative contributions lacking genuine reasoning	Participation rubric shifted from frequency to quality of reasoning; peer-review prompts embedded in M2
Overload and shallow self-study	Heavy timetable loads prevented adequate autonomous preparation; superficial reading became the norm	Task scoping reduced; M2 tasks designed to reward depth over breadth; feedforward used to scaffold next session
Unequal academic capital	Students with weaker prior academic literacy struggled disproportionately with text-heavy preparation	Supplementary access materials provided; M1 re-anchored key concepts before M2 tasks began

Of the five challenges documented, overload and unequal academic capital proved most persistent. Students carrying heavy timetable loads reported that adequate preparation for all sessions was structurally impossible, generating surface engagement rather than the deep processing the design assumed. Students with weaker prior academic literacy, particularly in reading dense theoretical texts, were doubly disadvantaged: their preparation was weaker, and their oral performance in M3 reflected this compounding.

Despite the challenges documented above, implementation logs and observation notes across all four units recorded a consistent trajectory of improvement from mid-term onwards. As the M1–M2–M3 cycle became routine, students adapted their preparation habits, and the quality of M3 defences improved measurably.

Rubric-based evaluations of M3 showed progressively stronger conceptual accuracy, tighter argument structure, and more explicit linkage between theory and case evidence. Instructors noted improvements in justification quality (the ability to articulate why a recommendation follows from a framework) and in responsiveness to probing questions. Where written artefacts in M2 displayed stylistic homogeneity consistent with AI-assisted drafting, the oral defence reliably differentiated levels of understanding.

End-of-term perception data reflected the dual character of students' experience. On the positive side, students identified M2 as a space for genuine experimentation and peer learning, and M3 as a consolidation catalyst that, once normalised, improved their confidence in applying concepts to novel cases. Respondents described M3 as improving perceived fairness in an AI-available context precisely because grades depended on individual demonstration.

These patterns were echoed in students' own words. Several respondents valued M2 as a space for authentic, accountable experimentation; as one postgraduate student commented: “Knowing I would have to defend the work without notes made me actually check the AI's answer instead of just pasting it”. The consolidating function of M3 recurred across units: “The oral defence was the most stressful part, but it was also where I finally saw how much I understood”. The perceived-fairness theme in an AI-available context was likewise explicit: “It felt fairer than a written exam, because the mark depended on what I could explain to myself”. The countervailing concerns were expressed with equal force, on performance anxiety, “I knew the content, but I froze under the direct questioning”, on the perceived arbitrariness of participation grading, “I never understood exactly what counted as good participation”, and on competing workload, ”With four units demanding this kind of preparation every week, something always had to give”. These extracts are

reproduced verbatim from the open-ended responses; they are illustrative rather than exhaustive and were selected to represent the range of positions expressed within and across the four units.

Persistent concerns clustered around three themes: (i) performance anxiety, particularly for students with weaker oral fluency; (ii) the perceived arbitrariness of participation grading; and (iii) the tension between preparation demands and competing academic commitments. Students whose discomfort with direct questioning persisted throughout the term reported a sense of structural disadvantage that rubric transparency and rehearsal sessions only partially mitigated.

No major academic integrity incidents were recorded during the implementation period. Students reported that the requirement to defend decisions without aids reduced incentives for uncritical tool use and encouraged verification of AI-generated content during M2. From the instructor's perspective, the routine shifted integrity management from detection towards pedagogy: the assessment anchor (individual oral reasoning) remained resistant to tool-mediated misrepresentation.

V. Discussion

The findings support a qualified account of M1–M2–M3 as integrity-by-design active learning: the routine produced genuine learning gains, but did so through a trajectory that included significant initial friction and persistent structural tensions. This qualification matters for theoretical interpretation and practical adoption.

Where the design succeeded, it did so through three aligned mechanisms. First, M1 operationalised constructive alignment [7] by making intended outcomes and performance criteria explicit and recurring, reducing the ambiguity that generates surface engagement. Second, M2 instantiated active learning through small-group problem work [4, 5], providing the repeated deliberate practice that the learning sciences associate with durable knowledge acquisition. Third, M3 constituted an authentic assessment [10, 11] that closed the feedback loop [8, 9] and made reasoning visible to the instructor and to students themselves.

The integrity-by-design logic also proved effective in practice: the combination of tool-permitted M2 and tool-free M3 shifted student behaviour towards verification and source triangulation during applied work, consistent with UNESCO guidance on human-centred AI integration [15].

The challenges documented in Section 4 are not incidental failures of implementation; they reflect structural features of accountability-centred designs that require explicit theorisation. Three tensions are particularly significant:

- **The autonomy paradox.** The method assumes students can perform meaningful independent first contact with new material; that is, that they can operate productively before instructor mediation. From a Vygotskian perspective [16], this inverts the developmental logic of scaffolded learning: students are placed in the zone of proximal development without the guided support that makes that zone productive. When this gap is too large, the result is not productive struggle but affective flooding, as anxiety closes down rather than opens up thinking.
- **The academic capital asymmetry.** Text-heavy preparation demands presuppose academic literacy capital distributed unequally across student populations [18]. Students with weaker prior access to academic reading cultures are doubly penalised: disadvantaged in preparation and consequently in oral performance. Far from functioning as an equalising mechanism, the design can amplify pre-existing inequalities if scaffolding is insufficient.
- **The participation instrumentalisation risk.** Grading participation introduces the risk that students optimise for performative contribution rather than genuine intellectual engagement. The authentic curiosity and productive vulnerability necessary for deep learning, including the willingness to say 'I don't understand', are suppressed in contexts where participation is evaluated for quality rather than rewarded for honesty.

These tensions take on a wider significance when situated within the broader transformation currently reshaping higher education, in which two forces are converging. The first is the continuing diversification and massification of the student body, which widens the distribution of prior academic capital within any single cohort [18]. The second is the normalisation of generative AI, which simultaneously lowers the cost of producing fluent written artefacts and raises the premium on demonstrable individual understanding. These forces interact: recent work cautions that AI tools may widen rather than narrow attainment gaps, because students with stronger prior literacy and metacognitive resources are better placed to use them critically, whereas less-prepared students risk substituting tool output for understanding [13, 14]. A growing critical literature similarly warns against techno-solutionist adoption that frames AI as a neutral efficiency gain rather than as a force that interacts with existing inequalities and institutional arrangements [23, 24]. Read against this backdrop, the academic-capital asymmetry observed here is not a peculiarity of the Thayer-inspired routine but a localised instance of a system-level problem: as cohorts diversify and AI becomes ambient, assessments that authenticate individual reasoning, such as M3, become more necessary for integrity, yet also more exposed to equity risk unless the preparation phase is deliberately scaffolded for the full range of academic capital present. The contribution of M1–M2–M3 is therefore not only a replicable routine but a concrete

demonstration that, in AI-era assessment, integrity-by-design and equity-by-design must be pursued together if such designs are to widen rather than restrict participation.

These tensions do not render the M1–M2–M3 routine unworkable; they define the conditions under which it can work well. The evidence suggests three necessary conditions. First, the instructor must adopt a hyperactive rather than reactive role: M1 must genuinely reduce ambiguity and anxiety rather than merely announcing that content was assigned. Second, preparation materials must be designed for accessibility across the range of academic capital present in the cohort, not calibrated to the median student. Third, participation assessment must explicitly reward reasoning quality and intellectual honesty, not punish visible confusion.

When these conditions are met, the design appears capable of producing the shift documented from mid-term onwards: from anxious compliance to genuine engagement with the preparation–application–demonstration cycle.

VI. Conclusion

This paper set out to examine how a Thayer-inspired M1–M2–M3 routine can combine active learning with integrity-by-design assessment across four business curricular units in an AI-normalised environment. The evidence supports a qualified but clear conclusion. The routine, brief conceptual framing (M1), tool-permitted small-group application (M2), and an unsupported individual oral defence (M3), produced a measurable trajectory from initial anxiety and under preparation towards stronger conceptual accuracy, tighter argumentation, and more explicit linkage of theory to applied decisions by mid-term. Critically, where M2 artefacts showed signs of AI-assisted homogeneity, the technology-free M3 defence reliably differentiated levels of individual understanding, and no major integrity incidents were recorded over the implementation period. In relation to the study's research questions, the routine proved implementable with constructive alignment across heterogeneous units (RQ1), generated learning evidence and student perceptions that were positive overall but unevenly distributed (RQ2), and reshaped perceived authenticity, integrity, and fairness in an AI-available context by anchoring grades in individual demonstration (RQ3).

At the same time, the study's central contribution is its candid account of the structural conditions that determine whether such a design succeeds. Three tensions – the autonomy paradox, academic-capital asymmetry, and the instrumentalisation of graded participation – recurred across units and were only partially mitigated. These are not incidental implementation failures but defining features of accountability-centred assessment: the very mechanisms that make the oral defence robust against tool-mediated misrepresentation are those that impose the heaviest affective and preparatory load on the least-resourced students. The routine works well, in short, only under specific conditions – a proactive rather than reactive instructor role in M1, preparation materials designed for the full range of academic capital in the cohort, and participation criteria that reward reasoning and intellectual honesty rather than performative compliance.

The broader implication for AI-normalised higher education is that integrity-by-design approaches are most defensible when they are simultaneously demanding and equitable: the accountability anchor of oral defence must be paired with sufficient scaffolding to prevent the design from functioning as a mechanism of exclusion rather than a mechanism of learning.

Several limitations bound the claims advanced here. This is an embedded single-case study conducted in one Portuguese public business school, across four curricular units, in a single academic year; its aim is analytic generalisation through pattern matching and triangulation, not statistical generalisation to other institutions, disciplines, or national systems [19]. The number of units and participants is small, and the design includes no control or comparison condition: the learning gains reported therefore, cannot be attributed causally to the M1–M2–M3 routine rather than to maturation, instructor effects, or cohort-specific factors, and no claim is made that the routine outperforms alternative assessment formats. Perception data are self-reported and subject to social-desirability and non-response bias. The qualitative findings reflect the interpretive judgement of a research team who were also the instructors; this dual role affords contextual insight but introduces potential bias that the audit trail and integration with rubric evidence only partially offset. Rubric scoring was likewise conducted by the instructors, with inter-rater agreement checked only where feasible. Accordingly, what travels beyond this case is the routine's design logic and the catalogue of structural tensions and mitigations, not effect sizes or generalisable performance differences; the patterns reported should be treated as hypotheses for testing in larger, multi-institutional, and ideally comparative or longitudinal studies.

Future research should examine longer-term transfer effects, multi-instructor replication, and the specific relationship between academic capital, preparation quality, and oral defence performance across diverse student populations. In addition, a larger-scale quantitative study, supported by a validated student questionnaire, is needed to probe more deeply the difficulties identified here and to test targeted interventions for improving student engagement and equalising preparation.

References

[1] Shell, A. E. (2002). The Thayer method of instruction at the United States Military Academy: A modest history and a modern personal account. PRIMUS, 12(1), 27–38. https://doi.org/10.1080/10511970208984015

[2] Connors, E. (2000). The Thayer method: Student active learning with positive results. Journal of Mathematics and Science: Collaborative Explorations, 4(1), 101–117. https://doi.org/10.25891/P6SW-GR81

[3] O'Halloran, K. P., Tangirala, S., Sun, F., Anagho, L. E., Agbegha, G., Runck, C., Roth, D., & Erickson, A. H. (2020). An experiential report on the Thayer method of teaching across college-level chemistry, biology, math, and physics courses. Georgia Journal of Science, 78(2), Article 4.

[4] Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

[5] Prince, M. (2004). Does active learning work? A review of the research. Journal of Engineering Education, 93(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

[6] Bonwell, C. C., & Eison, J. A. (1991). Active learning: Creating excitement in the classroom (ERIC Digest; ED336049). ERIC.

[7] Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32(3), 347–364. https://doi.org/10.1007/BF00138871

[8] Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. https://doi.org/10.1080/0969595980050102

[9] Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199–218. https://doi.org/10.1080/03075070600572090

[10] Wiggins, G. (1990). The case for authentic assessment. Practical Assessment, Research, and Evaluation, 2(1), Article 2. https://doi.org/10.7275/ffb1-mm19

[11] Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. Educational Technology Research and Development, 52(3), 67–86. https://doi.org/10.1007/BF02504676

[12] Nallaya, S., Gentili, S., Weeks, S., & Baldock, K. (2024). The validity, reliability, academic integrity and integration of oral assessments in higher education: A systematic review. Issues in Educational Research, 34(2), 629–646.

[13] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

[14] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[15] Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/EWZM9535

[16] Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

[17] Krashen, S. D. (1982). Principles and practice in second language acquisition. Pergamon.

[18] Bourdieu, P., & Passeron, J.-C. (1977). Reproduction in education, society and culture. SAGE.

[19] Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE.

[20] European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union. http://data.europa.eu/eli/reg/2016/679/oj

[21] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

[22] Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs — principles and practices. Health Services Research, 48(6 Pt 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117

[23] Selwyn, N. (2022). The future of AI and education: Some cautionary notes. European Journal of Education, 57(4), 620–631. https://doi.org/10.1111/ejed.12532

[24] Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. Learning, Media and Technology, 45(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995