Student Profiles and Perspectives on Being Assessed on the Use of Generative AI for Graded Coursework

Manuel Oliveira[1], Carlos Zednik[1], Bert Sadowski[1], Gunter Bombaerts[1], and Rianne Conijn[1]

[1]Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, The Netherlands

Last updated: March 30, 2026

Corresponding author: Manuel Oliveira (m.j.barbosa.de.oliveira@tue.nl), Human Technology Interaction, Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Den Dolech, Eindhoven, 5200MB, The Netherlands.

Abstract

As generative AI (GenAI) tools become embedded in higher education, understanding who adopts them for graded work, and why, has important implications for pedagogy and assessment design. This study examined AI adoption among graduate students ($N = 45$) completing argumentative essays at a technical university, in a context where AI use was explicitly permitted under the condition that the interactions were documented and graded on process-oriented criteria. We investigated three individual-level predictors (writing self-efficacy, need for cognition, AI literacy) using survey data and AI use records. Exploratory quantitative analyses showed that only AI literacy was able to differentiate between adopters ($n = 28$, 62.2%) and non-adopters ($n = 17$), which suggests how this factor may be more relevant to consider in adoption decisions compared to individual cognitive dispositions or self-reported confidence in task-related competencies. A qualitative thematic analysis of open-ended responses identified five themes for reasons to adopt AI in this process-based assessment context: strategic efficiency, skill compensation, performative compliance, intellectual agency, and ambivalence. A parallel thematic analysis of students' learning experience reflections ($n = 26$) identified six themes: AI as a research and source discovery tool, cognitive scaffolding for clarity and structure, efficiency and reduced cognitive load, concerns about AI reliability and hallucination, preference for autonomous learning, and preservation of intellectual ownership and critical agency. Together, findings suggest that domain-specific AI competence plays a primary role in adoption decisions than confidence in writing ability or individual information-processing tendencies, thereby offering an initial empirical window into a largely unexamined assessment context: being *assessed on* one's use of GenAI.

Keywords: Generative AI; process-based assessment; AI literacy; writing self-efficacy; need for cognition; AI adoption; qualitative analysis

Contents

1 Introduction

The rapid diffusion of generative AI (GenAI) tools into higher education has prompted institutions to rethink how student learning is assessed. Students now encounter assessments in which they are expected to engage with, manage, and document their use of AI tools as part of a graded task. Yet while the literature on student reactions to assessment innovation has grown substantially (Afrifa-Yamoah et al., 2024; Struyven et al., 2003), empirical evidence on how students navigate being *assessed specifically on their use of GenAI* remains scarce. This is a consequential gap: unlike generic digital tools, GenAI can produce fluent academic prose on demand, potentially substituting for or supplementing cognitive work that is central to essay-based assessment. How students decide whether to adopt such tools, and what individual characteristics shape that decision, is not yet well understood.

The present study examines AI adoption patterns among graduate students completing individual graded ethics essays at Eindhoven University of Technology (TU/e). Students were permitted to use AI tools and required to document their use; the essays were submitted for a final grade. Drawing on survey data and verified AI use records, we investigate three individual-level predictors: **writing self-efficacy** (WSE; Sun and Wang, 2020), **need for cognition** (NFC; Cacioppo and Petty, 1982), and **AI literacy** (AL; B. Wang et al., 2022). These constructs represent, respectively, students' beliefs in their own writing capabilities, their dispositional orientation toward effortful cognitive work, and their practical competence with AI technologies. To the best of our knowledge, the role of these constructs as predictors of AI adoption has not been examined in the specific context where students are assessed on their process of GenAI use for graded writing assignments. For a subset of students, prompt-log data and open-ended reflective responses extend the analysis to the character of AI use and students' own accounts of their experience.

Three research questions guide the study:

RQ1. Do writing self-efficacy, need for cognition, and AI literacy predict whether a student adopts AI for their essay?

RQ2. What reasons do students give for their decision to use or not use AI?

RQ3. How do students reflect on their learning experiences when completing the essay with or without AI assistance?

RQ1 is addressed quantitatively; RQ2 and RQ3 through reflexive thematic analysis of open-text survey responses (Braun & Clarke, 2006).

2 Background

2.1 Generative AI in Graded Assessment Contexts

Generative AI tools capable of producing fluent academic text have disrupted assumptions about what student writing products represent. Higher education institutions have diversified their responses (ranging from outright bans to explicit integration of AI use as part of assessed tasks) (Afrifa-Yamoah et al., 2024). Contexts in which students are expected to use, report, and justify AI involvement in a graded assignment constitute a distinctly novel form of assessment innovation. Rather than assessing writing ability in isolation, such tasks require students to exercise judgment about *when*, *how*, and *to what degree* to delegate cognitive work to an automated system.

Research on student reactions to assessment innovation more broadly shows that individual characteristics moderate whether students engage productively with novel formats. Perceived relevance, self-regulation capacity, and prior experience with the assessment type all shape how students approach and perform in innovative assessment designs (Sadeghi & Abolfazli Khonbi, 2015; Struyven et al., 2003). Assessment environments perceived as fair and congruent with students' sense of competence tend to foster engagement, while those that challenge self-efficacy or introduce unfamiliar demands can increase anxiety (Folwell et al., 2025; Li et al., 2024). Despite this growing body of evidence, the specific predictors of AI adoption decisions within graded assessment tasks have not been systematically examined.

2.2 Writing Self-Efficacy and Academic Writing Behavior

Self-efficacy, defined here as one's belief in one's capacity to execute behaviors needed to produce specific outcomes, is a well-established predictor of academic engagement and performance (Bandura, 1997). In the domain of writing, self-efficacy encompasses beliefs about one's ability to generate ideas, organise text, apply grammar, revise effectively, and self-regulate the writing process. Sun and Wang (2020) developed and validated the Questionnaire for English Writing Self-Efficacy (QEWSE), demonstrating that domain-specific self-efficacy predicts writing performance across five content areas in university populations. Emerging evidence in GenAI-integrated educational contexts confirms that self-efficacy plays a mechanistically relevant role. A structural equation model with 346 university students in a GenAI-supported entrepreneurship education context demonstrated a substantial positive effect of GenAI-integrated pedagogy on entrepreneurial self-efficacy ($\beta = 0.52$, $f^2 = 0.44$, $p < .001$), which in turn partially mediated the programme's positive effect on students' entrepreneurial intentions (indirect effect $\beta = 0.19$, $p < .001$) (Xie & Wang, 2025). A parallel chain-mediation study in design education similarly found that generative AI integration positively predicted domain-specific self-efficacy ($\beta = 0.53$, $p < .001$), which in turn shaped students' creative outcomes (Hwang & Wu, 2025). Earlier mediation work shows that student–GenAI interaction can influence learning achievement via self-efficacy as an intermediate mechanism (L. Wang & Luo, 2023).

Assessment research confirms that self-efficacy shapes responses to novel task demands: students with higher domain confidence tend to engage more actively and persist through difficulty, while lower self-efficacy is linked to avoidance and disengagement (Afrifa-Yamoah et al., 2024; Folwell et al., 2025). In writing-specific contexts involving GenAI, the relationship between WSE and engagement is empirically nuanced at the dimension level. Examining writing self-efficacy across ideation, construction, and self-regulation dimensions in a GenAI-feedback context, Zhang et al. (2025) found that GenAI feedback significantly affected efficacy beliefs in ideation and construction—but not self-regulation—and that all three dimensions correlated with students' writing engagement under AI-assisted conditions. These dimension-specific effects suggest that the facets of WSE most sensitive to GenAI may be those concerned with generating and organising content rather than managing the writing process overall.

In the context of AI-assisted writing, WSE may operate in competing directions depending on students' existing confidence level. High-WSE students may regard AI as redundant to their own capabilities and engage with it selectively to verify or extend their work; conversely, students with lower writing confidence may turn to AI as a resource for compensating for perceived skill deficits. This latter dynamic carries risks: TAM-based analyses of ChatGPT adoption for writing have noted that low-efficacy students are more prone to over-reliance on AI outputs, potentially at the cost of skill development (Fauzi et al., 2025). To the best of our knowledge, the role of writing self-efficacy as a *predictor* (rather than an outcome) of AI adoption behavior remains understudied, especially in the novel context of graded writing tasks where students are assessed on their process of GenAI use.

2.3 Need for Cognition and Engagement with Learning Tasks

Need for cognition (NFC) describes a stable individual disposition to seek out and enjoy effortful thinking (Cacioppo & Petty, 1982). In educational settings, higher NFC is associated with deeper processing strategies, greater engagement with challenging material, and more deliberate information-seeking behavior. In the context of AI-assisted academic writing, NFC is theoretically relevant in at least two ways. First, students high in NFC may resist delegating intellectually engaging work (such as argumentation and critical analysis), preferring to retain cognitive authorship. Experimental evidence supports the plausibility of this mechanism: across four pre-registered studies (N = 3,562), collaboration with GenAI on cognitively demanding text-generation tasks significantly undermined workers' intrinsic motivation in subsequent solo work (medium effect, $d = 0.44$–0.51), as GenAI displaced the intellectually engaging elements that make such tasks inherently rewarding (Wu et al., 2025). Students high in NFC—who derive particular satisfaction from effortful thinking—may therefore anticipate and avoid this motivational cost, preferring to maintain direct cognitive engagement with intellectually challenging writing tasks. Second, high-NFC students might adopt AI *strategically* as an intellectual interlocutor rather than as a text generator, using dialogue with the system as a form of extended thinking. Despite this theoretical relevance, to the best of our knowledge, NFC has not been examined as a predictor of AI adoption decisions in graded assessment contexts, and particularly not in the novel setting where students are assessed on their process of GenAI use.

2.4 AI Literacy as a Predictor of Technology Adoption

AI literacy encompasses the competencies needed to understand, use, evaluate, and behave ethically toward AI systems (B. Wang et al., 2022). The AI Literacy Scale (AILS; B. Wang et al., 2022) captures four dimensions: *Awareness* (identifying AI in practical contexts), *Use* (applying AI tools proficiently), *Evaluation* (critically assessing AI outputs), and *Ethics* (recognising responsibilities and risks of AI use). Since the widespread emergence of large-language-model tools such as ChatGPT, AI literacy has gained particular urgency as an educational competency: students are now expected to engage with systems capable of producing fluent academic text, yet evidence suggests that competence levels—and confidence in using AI—vary considerably across student populations. Studies in post-secondary settings indicate that students who feel competent in their use of GenAI tools deploy them more broadly and perceive them as effective learning resources, while first-year university students with limited AI competencies tend to engage with these tools in shallow, instrumental ways without adequately reflecting on their limitations or ethical implications (Hua & Cunningham, 2025; Miller, 2024).

Technology adoption research consistently finds that perceived competence—encompassing self-assessed ability and subjective ease of use—predicts voluntary adoption of new tools (Davis, 1989). Empirical work in GenAI-specific educational contexts supports this link directly. Studies applying the Technology Acceptance Model (TAM) in higher education settings have found that students' self-efficacy regarding AI tools significantly shapes their attitudes toward those tools, which in turn predicts behavioral intention and actual adoption (Sukirman et al., 2024). A TAM-based investigation of ChatGPT adoption for academic writing among undergraduate students similarly identified digital literacy and learner autonomy as essential competencies for meaningful AI integration: students with higher competence engaged with AI proactively and purposefully, while students with lower AI competency struggled to move beyond superficial usage (Fauzi et al., 2025). Together, these findings suggest that AI literacy constitutes a practically meaningful enabler—or barrier—of adoption, distinct from general technology attitudes or task-specific confidence such as writing self-efficacy.

Applied to the specific context of graded academic writing, students with higher AI literacy may face a lower practical barrier to adoption, be better positioned to use GenAI purposefully, and be more capable of critically evaluating AI-generated outputs—precisely the competencies that are explicitly assessed when GenAI use is itself graded. Despite the clear theoretical and emerging empirical basis, to the best of our knowledge, AI literacy as measured by a validated scale has not been examined as a predictor of adoption behavior specifically in graded assessment contexts, and is notably absent from work on the context where students are themselves evaluated on their process of GenAI use.

2.5 Summary and Research Questions

The foregoing review identifies a clear empirical gap: while evidence on student reactions to assessment innovation is accumulating (Afrifa-Yamoah et al., 2024; Struyven et al., 2003), the specific context of AI-assisted graded assessment has not been examined through the lens of individual psychological predictors. Writing self-efficacy, need for cognition, and AI literacy are theoretically motivated predictors of whether and how students engage with GenAI in graded essay tasks, yet none of these relationships has been systematically studied. The present study addresses this gap in a naturalistic assessment setting where AI use was both permitted and documented. Quantitative analyses test the predictive relationships (RQ1); qualitative analyses of students' own explanations and reflections (RQ2–RQ3) provide a complementary interpretive lens on the mechanisms underlying the adoption patterns observed.

3 Method

3.1 Participants

Participants were $N = 45$ graduate students enrolled in two courses at Eindhoven University of Technology (TU/e) in the academic year 2024–2025: Data Science Ethics (DSE; $n = 27$) and Philosophy and Ethics of AI (PEAI; $n = 18$). Both courses required students to write an individual essay as a graded assignment; essay topics addressed the societal and ethical dimensions of AI technologies.

Survey design and timing. Data collection differed across courses in structure, timing, and incentive. Table 1 summarises all three dimensions and the participant counts at each wave.

In DSE, students were invited at two survey moments. *Moment 1* (M1) was administered before essay submission (pre-essay). *Moment 2* (M2) was administered after essay submission but before course grades were posted: students had completed the writing task at this point but did not yet know their outcome. DSE participation was incentivised at €5 per completed survey. In PEAI, all data collection took place at a single pre-essay moment (M1); no post-essay survey was administered for this cohort. PEAI participation was entirely voluntary with no monetary reward.

Table 1: Survey administration by course and wave.

Course	Wave	Timing relative to essay	Incentive	Raw n	Retained[a]
DSE	M1	Before submission (pre-essay)	€5/survey	17	9
DSE	M2	After submission; *before grades posted*	€5/survey	38	27
PEAI	M1	Before submission (pre-essay)	None	32	22
Total				**87**	**58**

Note. DSE = Data Science Ethics; PEAI = Philosophy and Ethics of AI. [a]Retained after removing non-consenting, incomplete, and ineligible responses; M1 waves additionally exclude students who did not plan to write an individual essay.

DSE students completing both waves. Three DSE students completed surveys at both M1 and M2. Although repeated measurements were initially intended to support pre–post comparisons, the number of students with data at both waves ($n = 3$) was too small to conduct such analyses in a meaningful way. We therefore retained only the M2 data for those three participants, consistent with the approach applied to all other DSE M2 respondents, and discarded their M1 responses. This decision pools the three overlapping cases with the larger M2 cohort, allowing a more uniform basis for cross-student comparisons: the great majority of the DSE sample answered at the same post-submission, pre-grade moment.

Sample derivation. The 58 consent-eligible rows (Table 1) were subjected to a four-indicator response quality screen (Section 3.5); 11 rows were flagged and excluded, leaving 47 quality-clean rows. Deduplication to one row per student (M2 retained over M1 for the three DSE overlap cases) yielded the final analytical sample of $N = 45$ students (DSE: $n = 27$; PEAI: $n = 18$).

Table 2 shows the sample composition by course and AI adoption status.

Of the 45 students, 28 (62.2%) verified that they used AI in their individual essay (AI adopters) and 17 did not (non-adopters). AI adoption was substantially higher in DSE (23 of 27 students, 85%) than in PEAI (5 of 18 students, 28%).

All students in the analytical sample provided informed consent and consented to anonymised data storage. The qualitative data (RQ2–RQ3) are restricted to DSE students who completed M2, as the open-ended items on AI use were included only in that survey.

Table 2: Analytical sample by course and AI adoption status.

Course	Non-Adopters	AI Adopters	Total
Data Science Ethics (DSE)	4	23	27
Philosophy & Ethics of AI (PEAI)	13	5	18
Total	**17**	**28**	**45**

3.2 Course Context and AI Use Policy

Both courses addressed the ethical and societal dimensions of data science and artificial intelligence technologies. Core learning objectives included the ability to construct well-reasoned, evidence-based arguments in written form, to identify and evaluate ethical and philosophical arguments, to apply ethical theories to contemporary technological contexts, and to critically integrate scholarly literature. The graded argumentative essay served as the primary assessment of these competencies.

Use of generative AI tools (e.g., ChatGPT) for essay writing was explicitly permitted but optional. Students who chose to use AI were required to submit complete interaction logs (sequences of prompts and AI outputs) alongside their essays. To ensure equitable access to AI-assisted writing strategies, all courses included at least one lecture on argumentative writing techniques and basic prompt engineering. This training aimed to mitigate potential disparities in GenAI proficiency among students.

Students who used AI were evaluated on two dimensions: (1) traditional essay quality, assessed using rubrics focused on ethical analysis, argumentative structure, critical thinking, integration of course concepts, and writing clarity; and (2) AI interaction quality, assessed using a rubric evaluating three criteria (Oliveira et al., 2025): *AI for Writing* (technical mastery of prompt formatting and strategic use of AI as a writing aid), *AI for Argumentation* (critical engagement with AI-generated content and use of AI to improve argumentative structure), and *AI for Course Content* (demonstration of course material understanding through content-focused prompts). Both scores contributed to the final course grade. Students who did not use AI were evaluated solely on traditional essay quality. All data collection followed informed consent procedures and received ethical approval from the Ethical Review Board of Eindhoven University of Technology.

3.3 Measures

Writing Self-Efficacy (WSE). The Questionnaire for English Writing Self-Efficacy (QEWSE; Sun and Wang, 2020) comprises 27 items rated on a 7-point scale (1 = *low self-efficacy* to 7 = *high self-efficacy*). Items are competency statements of the form *"I can . . ."* and cover five content areas:

- *Ideation*: generating, expressing, and organising ideas for writing.

- *Organisation*: structuring sentences and paragraphs into coherent compositions.
- *Grammar*: producing grammatically accurate sentences, verb tenses, and spelling.
- *Revision*: reviewing and improving drafted text.
- *Self-regulation*: managing attention, time, persistence, and emotional state during writing.

All items are positively worded; higher scores indicate stronger perceived writing capability. Internal consistency in the present sample: $\alpha = 0.91$. Exact item wordings are listed in Appendix A.

Need for Cognition (NFC). The six-item NFC short form (NFC-6; Coelho et al., 2020) was administered on a 5-point scale (1 = *extremely uncharacteristic* to 5 = *extremely characteristic*). The scale captures the *tendency to engage in and enjoy effortful cognitive activity*, operationalised as enjoyment of complex problems, preference for intellectual challenges, and intrinsic motivation for cognitively demanding tasks. Two items are reverse-worded (items 3 and 4, e.g. *"Thinking is not my idea of fun"*) and were recoded prior to scoring. Internal consistency: $\alpha = 0.56$. Full item wordings are listed in Appendix A.

AI Literacy (AL). Twelve items from the AI Literacy Scale (AILS; B. Wang et al., 2022) assessed four dimensions, each measured with three items on a 7-point scale (1 = *strongly disagree* to 7 = *strongly agree*). Three items were reverse-worded and recoded. Dimensions are:

- *Awareness* (3 items): the ability to *identify and comprehend* AI technology in practical applications.
- *Use* (3 items): the ability to *apply and exploit* AI tools to accomplish tasks proficiently.
- *Evaluation* (3 items): the ability to *analyse, select, and critically assess* AI applications and their outputs.
- *Ethics* (3 items): awareness of the *responsibilities and risks* associated with AI use.

Overall $\alpha = 0.72$; subscale α: Awareness = 0.46, Use = 0.65, Evaluation = 0.57, Ethics = 0.56. Full item wordings are listed in Appendix A.

AI adoption. Verified via course records as a binary variable (0 = non-adopter, 1 = adopter).

Open-text responses. Two post-essay items (DSE, Moment 2) captured reasons for the AI adoption decision and reflections on the learning experience. These items form the basis of the qualitative analyses (Section 5).

3.4 Procedure

Students completed an online survey (Qualtrics) before their essay (Moment 1, M1; both courses) and, for DSE students, a follow-up survey after essay submission (Moment 2, M2). The M1 survey included the WSE, NFC, and AL scales. The M2 survey included post-task AI use confirmation and the two open-text items. All three composite scores used in the analyses were drawn from M1 responses.

It is important to note that participation was voluntary and survey completion was not linked to the graded task itself, which is typical of observational studies in naturalistic classroom settings. As a result, very few individual students completed *both* M1 and M2: most students who responded did so at one moment only (primarily M2 for DSE, primarily M1 for PEAI). The design therefore does not support within-person comparisons of attitudes before and after essay submission; the two moments reflect largely distinct subsamples from the same course populations, not a repeated-measures panel. Where a student did complete both moments, the later response was retained in the analytical dataset to preserve the most task-proximal psychometric information.

3.5 Data Quality Screening

Survey responses were screened for careless responding using four item-level indicators (identical response strings, near-zero within-person variance, very low person-total correlation, and implausibly fast completion time) and a multivariate outlier check on the three composite scores. Cases flagged on any single indicator were excluded from the main analyses (11 cases removed; see Section 3.1) and retained in a separate dataset for sensitivity comparisons.

3.6 Analytic Strategy

All analyses used $\alpha = .05$ (two-tailed). Sensitivity power analyses establishing the minimum detectable effect sizes for the actual sample sizes are reported in the subsection that follows.

For **RQ1**, we first compared AI adopters and non-adopters on WSE, NFC, and AL using independent-samples t-tests with Cohen's d as the effect size. We then modelled AI adoption as a function of each psychometric predictor separately using binary logistic regression with standardised predictors (to allow OR comparison), followed by a joint model. Exploratory item-level analyses examined the predictive value of each WSE (27 items), NFC (6 items), and AL (12 items) item individually. All item-level results are reported with a Bonferroni-corrected threshold ($p^* = 0.0011$) and treated as hypothesis-generating only.

3.7 Statistical Power Considerations

Sensitivity power analyses were conducted to establish the minimum effect sizes detectable at 80% power given the actual sample sizes ($n_{\text{adopters}} = 28$, $n_{\text{non-adopters}} = 17$, $\alpha = .05$, two-tailed). For independent-samples comparisons, the minimum detectable Cohen's d is $d_{\text{min}} = 0.88$. For the binary prediction model, the minimum detectable OR is approximately $\text{OR}_{\text{min}} \approx 4.93$ (Demidenko approximation). The study is adequately powered only for medium-to-large effects; all results should therefore be interpreted as exploratory.

4 Quantitative Analyses

4.1 Descriptive Overview

Table 3 presents descriptive statistics by AI adoption status for the three psychometric scales, two perceived-clarity items, and within-course z-scores for essay grade and AI interaction score. All results should be interpreted as exploratory given the modest sample size ($N = 45$).

Table 3: Descriptive statistics by AI adoption status.

Measure	AI Adopters ($n = 28$)	Non-Adopters ($n = 17$)	Full sample (N
Psychometric scales			
Writing Self-Efficacy (WSE)	5.52 (0.62)	5.56 (0.71)	5.54 (0.64
Need for Cognition (NFC)	3.70 (0.43)	3.53 (0.50)	3.64 (0.46
AI Literacy (AL)	5.40 (0.55)	4.99 (0.61)	5.24 (0.60
Course context perceptions (1 = not clear, 7 = perfectly clear)			
Clarity: AI use instructions	6.04 (1.14)	5.65 (1.06)	5.89 (1.11
Clarity: AI interaction grading	5.39 (1.42)	5.35 (1.54)	5.38 (1.45
Grade outcomes (within-course z-scores)			
Essay score	-0.06 (1.04)	0.16 (0.86)	-0.00 (0.99
AI interaction score[a]	-0.00 (0.98)	—	-0.00 (0.98

Note. Scale ranges: WSE 1–7; NFC 1–5; AL 1–7; clarity items 1–7. α: WSE = 0.91, NFC = 0.56, AI
[a]AI interaction score available for AI adopters only; z-score computed within course.

4.2 RQ1a: AI Adopters vs. Non-Adopters

Table 4 presents Welch t-tests comparing AI adopters and non-adopters on all three psychometric measures and the four AL subscales. Group score distributions are displayed in Figure 1.

Table 4: Group comparisons: AI adopters ($n = 28$) vs. non-adopters ($n = 17$). Welch's t-tests.

Measure	Adopters M (SD)	Non-Adopters M (SD)	t	df	p	d
Writing Self-Efficacy	5.52 (0.62)	5.56 (0.71)	-0.20	30.3	0.840	-0.06
Need for Cognition	3.70 (0.43)	3.53 (0.50)	1.18	29.7	0.248	0.38
AI Literacy (overall)	5.40 (0.55)	4.99 (0.61)	2.30	31.1	0.029	0.72
Awareness	5.70 (0.74)	5.22 (0.88)	1.90	29.6	0.067	0.61
Use	5.88 (0.74)	5.41 (1.00)	1.67	26.7	0.107	0.55
Evaluation	5.27 (0.83)	5.24 (0.70)	0.17	38.3	0.869	0.05
Ethics	4.74 (0.92)	4.08 (1.08)	2.09	29.6	0.045	0.67

Note. Cohen's d computed with pooled SD. Minimum detectable $d = 0.88$ at 80% power. Two-tailed tests, $\alpha = .05$.

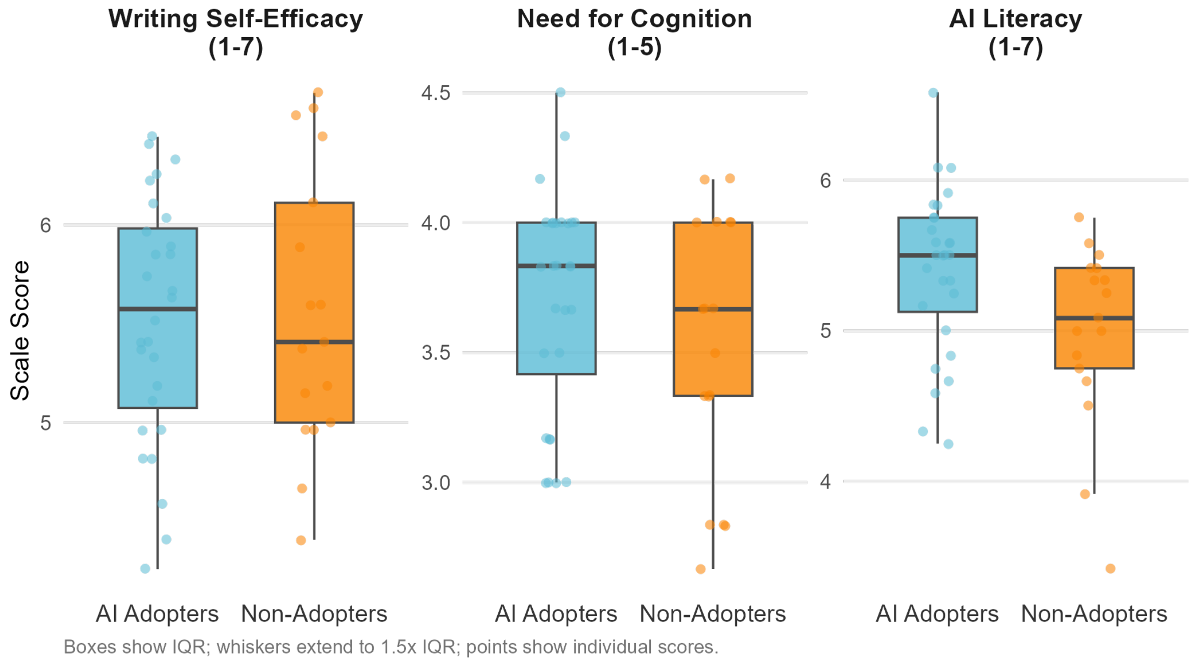

Figure 1: Writing self-efficacy, need for cognition, and AI literacy scores by AI adoption group. Boxes show IQR; whiskers extend to 1.5× IQR; points show individual scores.

4.3 RQ1b: Predicting AI Adoption

Predictors were standardised ($M = 0$, $SD = 1$) prior to analysis so that coefficients reflect the change in log-odds per one-standard-deviation increase. Table 5 reports univariable and joint model results.

Table 5: Binary logistic regression predicting AI adoption (1 = adopter). Standardised predictors.

Predictor	B	SE	z	p	OR	95% CI
Univariable models						
Writing Self-Efficacy	-0.07	0.31	-0.22	0.829	0.94	[0.51, 1.72]
Need for Cognition	0.39	0.32	1.22	0.223	1.47	[0.79, 2.74]
AI Literacy	0.78	0.37	2.11	0.035	2.18	[1.06, 4.48]
Joint model						
Writing Self-Efficacy	-0.04	0.38	-0.11	0.909	0.96	[0.46, 2.00]
Need for Cognition	0.15	0.40	0.37	0.711	1.16	[0.53, 2.55]
AI Literacy	0.71	0.41	1.75	0.081	2.04	[0.92, 4.52]

Note. OR = odds ratio. 95% CI based on $\pm 1.96 \times SE$. Minimum detectable OR $\approx$ 4.93 at 80% power.

4.4 Grade Outcomes: Adoption-Group Differences

To examine whether AI adopters and non-adopters differed in essay performance, we compared within-course essay z-scores using a two-tailed Mann–Whitney U test. AI interaction scores were available for adopters only and are therefore not compared across groups. Table 6 presents the results.

Table 6: Non-parametric comparison of essay grade z-scores by AI adoption status.

Measure	Adopters M (SD)	Non-Adopters M (SD)	W	p	r
Essay score (z)	-0.06 (1.04)	0.16 (0.86)	156.0	$= 0.963$	0.01

Note. Mann–Whitney U test (two-tailed). AI interaction score is available for AI adopters only

4.5 Performance Prediction Analyses

Two exploratory linear regressions examined whether WSE, NFC, and AL predicted grade outcomes. The first model (Table 7) regressed within-course essay z-score on the three standardised predictors for all students with available grades. The second model (Table 8) regressed within-course AI interaction z-score on the same predictors, restricted to AI adopters. Both analyses are underpowered and results should be treated as hypothesis-generating only.

Table 7: Linear regression predicting essay grade (z-score). Standardised predictors.

Predictor	$\hat{\beta}$	SE	t	p	95% CI
Writing Self-Efficacy	-0.05	0.20	-0.26	0.797	[-0.45, 0.35]
Need for Cognition	0.22	0.22	1.02	0.314	[-0.22, 0.66]
AI Literacy	-0.00	0.20	-0.02	0.983	[-0.41, 0.40]

Note. $n = 39$. $R^2 = 0.042$; $F(3, 35) = 0.52$, $p = 0.673$.

Table 8: Linear regression predicting AI interaction score (z-score). AI adopters only. Standardised predictors.

Predictor	$\hat{\beta}$	SE	t	p	95% CI
Writing Self-Efficacy	-0.20	0.21	-0.97	0.343	[-0.63, 0.23]
Need for Cognition	0.15	0.31	0.50	0.623	[-0.48, 0.79]
AI Literacy	0.23	0.29	0.79	0.435	[-0.37, 0.82]

Note. AI adopters only ($n = 28$). $R^2 = 0.144$; $F(3, 24) = 1.35$, $p = 0.283$.

As a robustness check, Table 9 replicates the essay grade model using raw scores rather than within-course z-scores; the pattern of results is consistent.

Table 9: Linear regression predicting essay grade (raw score). Robustness check.

Predictor	$\hat{\beta}$	SE	t	p	95% CI
Writing Self-Efficacy	-1.87	3.03	-0.62	0.541	[-8.03, 4.29]
Need for Cognition	4.08	3.30	1.23	0.226	[-2.64, 10.79]
AI Literacy	1.94	3.19	0.61	0.547	[-4.54, 8.42]
AI Adoption (0/1)	-4.32	5.85	-0.74	0.466	[-16.22, 7.58]

Note. $R^2 = 0.103$; $F(4, 34) = 0.98$, $p = 0.432$.

4.6 Exploratory Item-Level Analyses

Each WSE item ($k = 27$), NFC item ($k = 6$), and AL item ($k = 12$) was examined as an individual predictor of AI adoption via logistic regression. All tests are exploratory. Bonferroni-corrected significance threshold: $\alpha^* = .05/45 = 0.0011$.

Figure 2 presents a forest plot of OR estimates. These item-level analyses are exploratory and heavily underpowered given the sample size; they should be treated as hypothesis-generating only.

5 Qualitative Method

5.1 Data

Open-text survey responses were collected at M2 from DSE students only (see Section 3.3 for survey timing details). Two items were analysed:

1. **Reasons for AI adoption decision** (`reasons_ai_use`): *"Can you briefly provide some reason(s) why you decided to use (or not use) AI? Please describe 1 to 3 reasons."*

2. **Learning experience reflections** (`reflect_ai_use`): *"Please reflect on your learning experience completing the assignment with AI (if you used it) or without AI (if you did not use it)."*

Responses were manually linked to verified AI use status from course records to ensure accurate group labelling (AI user / non-user) in the analysis.

5.2 Analytic Framework

Analyses followed the **reflexive thematic analysis** (RTA) framework of Braun and Clarke (2006, 2019). Unlike template or codebook approaches, RTA treats themes as researcher-constructed interpretations of patterned meaning across a corpus, not as pre-existing categories to be counted. The six phases are:

1. **Familiarisation.** Repeated, immersive reading of all responses to develop depth of engagement with the data and begin noting preliminary ideas (before any systematic coding begins).

2. **Initial coding.** Systematic generation of concise labels (*codes*) that capture meaningful semantic features of the data relevant to the research question. Each code is anchored to a specific excerpt.

3. **Searching for themes.** Sorting codes into potential thematic groupings by identifying patterns of shared meaning that recur across responses.

4. **Reviewing themes.** Checking candidate themes against both the coded extracts and the full corpus to ensure internal coherence (within themes) and external distinctiveness (between themes); merging, splitting, or discarding as needed.

5. **Defining and naming themes.** Articulating the *essence* of each theme (what it captures, why it matters) and settling on a name that reflects its central concept, not merely its surface content.

6. **Producing the report.** Writing a narrative account of the themes, supported by verbatim illustrative extracts, and situating findings in relation to the research questions.

5.3 Implementation

The two open-text items were analysed through different implementation pathways: the reasons corpus (RQ2) via researcher-led analysis, and the reflections corpus (RQ3) via LLM-assisted analysis. Both pathways followed the six-phase RTA structure above.

5.3.1 RQ2 — Reasons for AI Adoption (Researcher-Led)

All six phases were conducted by the first author. To support systematic management of coding and theme development (rather than relying on unstructured spreadsheets or word-processor notes), a purpose-built Streamlit web application (*Thematica*) was used throughout. The application provided a structured interface for familiarisation, per-response code assignment, cross-response code review, and iterative theme construction and definition, without delegating interpretive decisions to automated processes. Coding and theme labels, descriptions, and all representative extracts reported in Section 6.1 represent the researcher's own interpretive judgements, grounded directly in the corpus.

5.3.2 RQ3 — Learning Experience Reflections (LLM-Assisted)

Given the exploratory goals and the size of the reflections corpus, phases 2–5 were implemented with large language model (LLM) assistance using the Claude API (Anthropic, model `claude-haiku-4-5-20251001`). LLM-assisted qualitative coding has been used in recent educational research as an efficiency-oriented tool, provided the researcher retains interpretive authority and documents the analytical chain transparently (cf. Bijker et al., 2024). Phases 1 and 6 were conducted by the researcher.

Each API call followed a two-component structure: a *system prompt* defining the model's role, analytical framework, and output format, and a *user prompt* supplying the data for that call. All prompts are reproduced verbatim in Appendix B. The design logic of each phase was as follows.

Phase 2 (Initial coding): Each response was submitted in a separate API call. The system prompt established the model as a qualitative researcher within Braun & Clarke's framework, embedded an item-specific research context paragraph anchoring the task to the study's RQs, and required 2–5 semantic codes per response (concise descriptive phrases of 3–8 words). Instructions explicitly directed the model to stay grounded in the participant's own words while distinguishing surface content from latent significance. The user prompt supplied the participant's group label (AI user or non-user) alongside the response text. Output was constrained to a JSON array of strings to enable programmatic parsing and audit. Temperature was set to 0.2, favouring close fidelity to the data over generative variation.

Phase 3 (Candidate theme generation): All Phase 2 codes were assembled into a single structured block (one line per participant, with group label and codes separated by pipe characters) and submitted in one call. The system prompt required 3–6 candidate themes, each returned as a JSON object with five fields: theme name, central concept, description, supporting codes, and separate characterisations of how AI users and non-users manifested the theme. Requiring structured JSON output enforced consistent theme representation and produced output directly importable into the researcher's review workbooks. Temperature was set to 0.4, allowing slightly more generative synthesis across the full code set.

Phase 4 (Theme review): The candidate theme JSON and the full raw corpus (participant ID, group label, and response text for all participants) were submitted together in one call. The system prompt specified four explicit review criteria (data support, internal coherence, inter-theme distinctiveness, and collective coverage) and authorised the model to merge, split, remove, or rename themes. A `review_notes` field was required per theme to document what changed and why, creating a traceable rationale for every structural decision. Temperature was 0.3.

Phase 5 (Final definition): The reviewed themes and full raw corpus were submitted together. The system prompt required a publication-ready definition per theme (3–4 sentences), a one-to-two-sentence analytical insight, a coverage note, and 2–3 illustrative quotes. The instruction to reproduce "exact verbatim text" and not to paraphrase was stated explicitly in the prompt. Participant IDs were included in the corpus block to enable traceable attribution. Temperature was 0.3.

Requesting JSON-structured output across all phases was a deliberate design choice that constrained the model's response space, made structural deviations detectable, and generated a complete audit trail of intermediate analytical states in structured workbooks. All outputs were inspected by the researcher before acceptance. No theme was retained without clear grounding in the corpus.

6 Qualitative Analyses

6.1 RQ2 — Reasons for AI Adoption or Non-Adoption

The reasons corpus comprised **28 responses** (23 AI users, 5 non-users) from DSE students. Analysis yielded five themes (three capturing adoption motivations among AI users and two capturing non-adoption reasoning) summarised in Table 10.

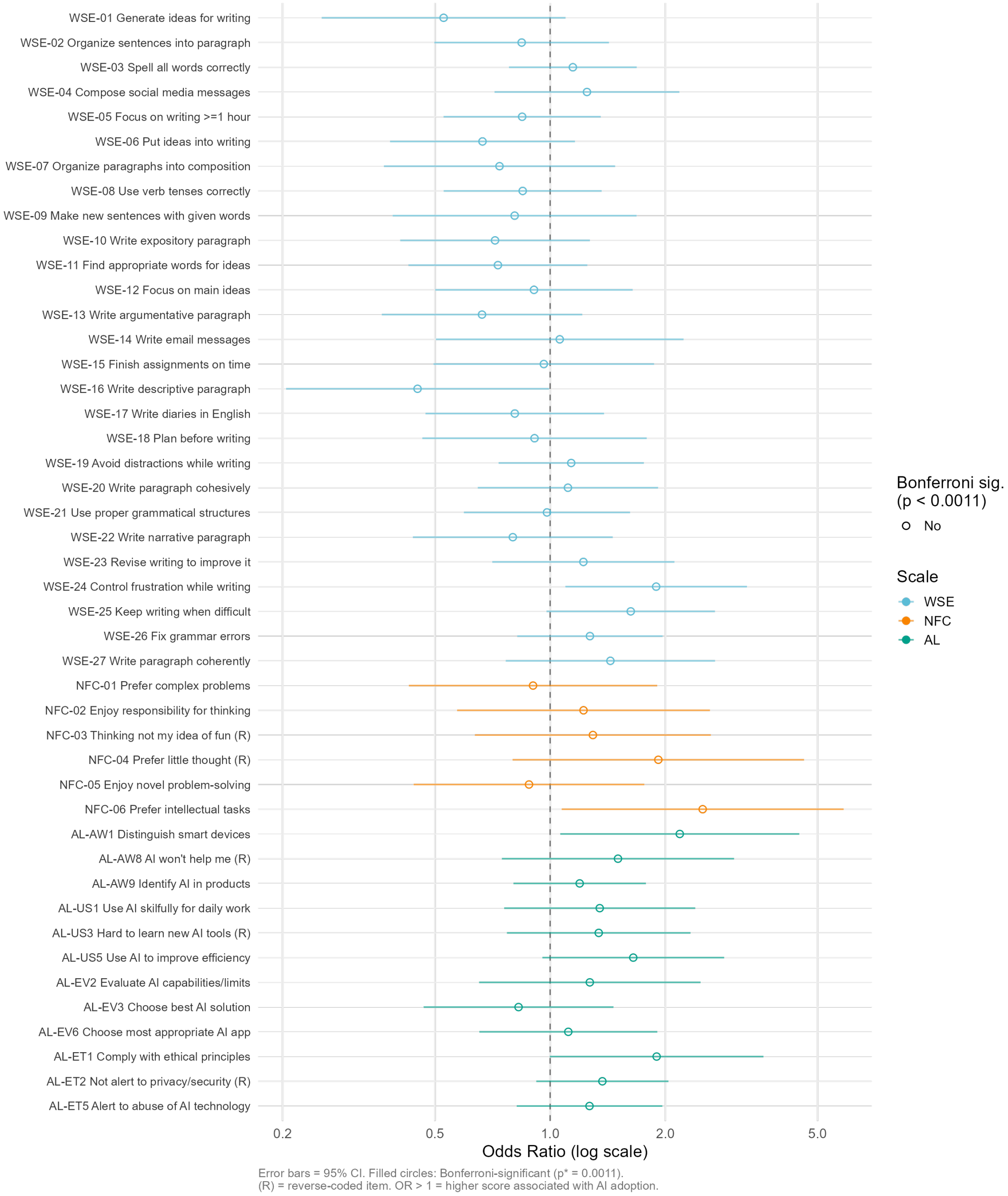

Figure 2: Forest plot of item-level odds ratios (OR) predicting AI adoption. Filled circles indicate Bonferroni-significant results ($p^* = 0.0011$). Error bars show 95% CIs. OR > 1 indicates association with higher adoption probability. (R) = reverse-coded item.

Table 10: Summary of themes — reasons for AI adoption/non-adoption ($N = 28$ responses).

#	Theme	Group	Codes (n)
1	Task efficiency and cognitive offloading	AI Users	9
2	Compensation for skill deficits and output augmentation	AI Users	10
3	Performative compliance	AI Users	4
4	Risk aversion and preservation of intellectual agency	Non-users	5
5	Ambivalence about AI use	Non-users	3

Theme 1: Task Efficiency and Cognitive Offloading

Central concept: Streamlining the assignment workflow by offloading effort-intensive sub-tasks to AI.
This theme draws on the following analytical codes: *AI for efficiency*; *AI as facilitator of task initiation*; *AI to increase focus on meaningful tasks*; *AI for information retrieval*; *AI for summarisation*; *AI to reduce tool complexity*; *AI as a tool to avoid effort*; *probing skill offloading*.

This theme captures the most prevalent motivation for AI adoption across the corpus: reducing the time and cognitive friction associated with the procedural, low-creativity components of academic writing. Students framed AI as a tool that handles the *scaffolding* work (proofreading, reference formatting, information retrieval, structural organisation) so that cognitive resources can be directed toward reasoning and argument construction. Several responses explicitly contrasted AI-assisted efficiency against traditional tool chains (browser searches, manual revision). A related strand reflects a weaker form of the same motivation: using AI to get *started* when facing a blank page or mental block.

The efficiency rationale maps directly onto the perceived usefulness construct in Davis (1989)'s Technology Acceptance Model: AI adoption is driven by expectations that the tool will improve task performance, and time savings on scaffolding work plausibly constitute a meaningful performance gain in an essay context. Whether such efficiency gains reflect genuine strategic delegation or rationalised effort-avoidance cannot be resolved from self-report data alone; however, quotes describing AI as preserving focus for "creative and logical thinking" suggest intentional cognitive delegation rather than wholesale disengagement. This distinction matters theoretically: Struyven et al. (2003) noted that students' surface versus deep engagement with novel assessment formats often mirrors the perceived learning value they assign to the task. Efficiency-oriented adoption is consistent with deep engagement when cognitive resources freed from procedural work are actively reinvested in higher-order reasoning. The boundary with Theme 2 (skill compensation) is worth noting: both involve AI reducing cognitive burden, but efficiency-oriented users appear to have a defined target for the freed resources, while compensatory users may not.

Representative quotes:

> "It helps me output my ideas faster by saving me time on checking grammar, making paragraphs more coherent, correcting spelling etc."
> (AI User) *Illustrates: AI for efficiency*

> "Using AI really kicked off the writing journey, using it for bouncing ideas off of it and finding relevant articles really boosts the speed of doing a writing assignment."

(AI User) *Illustrates: AI as facilitator of task initiation; AI for efficiency*

"Saves time in the mundane parts (restructure, grammatical correctness, finding related articles and posts for research). Allows me to focus mainly on the creative and logical thinking, without having to worry about putting everything together..."
(AI User) *Illustrates: AI to increase focus on meaningful tasks*

Theme 2: Compensation for Skill Deficits and Output Augmentation

Central concept: Using AI to compensate for self-perceived gaps in writing ability, language proficiency, or argumentation, and to raise baseline output quality.
This theme draws on the following analytical codes: *writing improvement — specific*; *writing improvement — general*; *AI language support for non-natives*; *AI as compensation for known personal flaws*; *explicit AI for augmentation*; *AI as personal tutor*; *AI for structuring output*; *AI for dialectics*; *AI as brainstorming tool.*

Where Theme 1 is primarily about *speed*, this theme is about *quality remediation*. Students reported using AI because they identified specific limitations in their own academic writing (structural organisation, non-native language production, argumentation depth, or the ability to express ideas coherently in English). The theme also includes an *augmentation* strand, in which AI is used not merely to fix deficiencies but to extend or enrich the student's own ideas. A more cognitively engaged variant involves using AI as a dialectical partner (submitting one's own arguments and requesting counterarguments).

Bandura (1997)'s model of self-efficacy predicts that students with lower domain confidence are more likely to adopt compensatory strategies: using AI as a substitute for perceived capability deficits follows the same logic as avoidance-oriented adaptations to challenging task demands. The augmentative strand within this theme is theoretically distinct: using AI as a dialectical partner to extend one's own ideas implies a higher level of autonomous engagement and a more active role in shaping the output. The interaction with AI literacy is relevant here (B. Wang et al., 2022): students with stronger competence and evaluation dimensions are better positioned to critically assess and selectively integrate AI suggestions, making augmentative use more likely than wholesale acceptance of AI-generated content. This implies that the compensatory and augmentative sub-patterns within this theme may be differentially distributed across the AI literacy spectrum, even though the data do not allow this to be tested directly.

Representative quotes:

"I decided to use AI to give a better structure to my report, as I'm not as good as on my mother tongue to express myself in a concise and structured way."

(AI User) *Illustrates: AI language support for non-natives; Writing Improvement — Specific*

"I used AI to help me extend my ideas I came up with. With help of providing reasons why my chosen solution would be effective approach in addressing ethical issue."
(AI User) *Illustrates: Explicit AI for augmentation*

"I also liked to challenge myself with a different mindset after I already gave my idea to the AI bot to see how it affects the main point."
(AI User) *Illustrates: AI for dialectics*

Theme 3: Performative Compliance

Central concept: Strategic engineering of AI use to meet perceived assessment expectations or to maximise the appearance of learning evidence aligned with grading criteria.
This theme draws on the following analytical codes: *perceived implicit expectation to use AI*; *AI to verify compliance*; *assessment context reason*; *signalling ethical use.*

This theme is the smallest in frequency but analytically distinct from the other two AI-user themes. Rather than being motivated by intrinsic learning or task-completion goals, these students appeared to use AI instrumentally in response to the assessment context itself (because AI use was *encouraged and graded*, because they sought to *signal ethical use* to the instructor, or because they used AI to explicitly verify whether their output met the stated rubric). This suggests a surface-level, extrinsically motivated pattern of adoption that may not generalise beyond graded assessment contexts.

Struyven et al. (2003) documented that students' responses to novel assessment formats are substantially shaped by their reading of grading signals and instructor expectations, with surface-oriented adaptation a common response when students are uncertain about what constitutes genuinely valued engagement. Performative compliance in this context is a direct instantiation of that dynamic: AI use motivated by the graded and encouraged nature of the task rather than by intrinsic perceived utility. Afrifa-Yamoah et al. (2024) offer a complementary lens through goal orientation theory: performance-oriented students are more likely to engage instrumentally with assessment tasks (adopting AI because it is graded and encouraged) whereas mastery-oriented students are more likely to evaluate adoption in terms of its learning value. Performative compliance thus maps onto a performance goal orientation, and the presence of both orientations within the adopter group has direct consequences for interpreting the RQ1 finding. If a subset of adopters engaged with AI for extrinsic rather than competence-driven reasons, the observed AI literacy difference between adopters and non-adopters partly reflects selection (competent, purposeful adopters) rather than a clean predictor relationship. Performative compliers may have varied AI literacy without this being detectable in the aggregate analysis, which treats all adopters as a homogeneous group.

Representative quotes:

> "The use of AI was encouraged and graded, so I did my best to utilize them well, to make the essay better."
> (AI User) *Illustrates: Perceived implicit expectation to use AI*

> "To review my paper according to the rubric."
> (AI User) *Illustrates: AI to verify compliance*

Theme 4: Risk Aversion and Preservation of Intellectual Agency

Central concept: Refusing AI use to protect grade security, intellectual ownership, or a sense of personal control over the final product.
This theme draws on the following analytical codes: *avoid AI use to feel agency over content generation*; *graded AI use perceived as risky*; *not using due to high confidence in own ability*; *AI underperformance experience.*

Non-adopters in this theme were not indifferent to AI but actively *chose against it.* Two distinct but co-occurring rationales appear: (a) **grade risk** (a perception that delegating writing to AI in an individually graded, argumentative assignment is a gamble, and that AI outputs may not meet performance standards reliably); and (b) **agency** (a principled desire for the assignment to reflect the student's own intellectual effort and reasoning, independent of AI assistance).

The confidence-based non-adoption rationale is consistent with Bandura (1997)'s account of how high self-efficacy produces mastery-oriented behavior: students confident in their writing capabilities may perceive AI assistance as unnecessary and potentially disruptive to a product that reflects their own competence. Autonomy functions here not as indifference but as an active psychological need: Folwell et al. (2025) found that assessment formats which provide choice and self-direction support students' autonomy orientation, and this theme suggests that the choice to *not* use AI can itself be an expression of that orientation when the student's existing competence makes external scaffolding feel redundant. The quantitative data show no significant WSE group difference between adopters and non-adopters, which is consistent with this theme co-existing with the compensatory adoption pattern in Theme 2: high-WSE students who did not adopt and lower-WSE students who did adopt would both contribute to a null aggregate WSE difference. Rather than disconfirming WSE's relevance, the null finding may reflect these two competing pathways operating simultaneously and cancelling at the group level.

Representative quotes:

> "I did not use AI because I want the assignment to reflect my own capabilities and not those of AI agents."
> (AI Non-user) *Illustrates: Avoid AI use to feel agency over content generation*

> "I did not want to risk getting a low grade because the AI did not do the assignment correctly...I was confident in my own skills that my grade would've been enough to pass the course."
> (AI Non-user) *Illustrates: Graded AI use perceived as risky; Not using due to high confidence in own ability*

> "I decided not to use AI because I thought that it would be easier to understand arguments if I myself come up with them rather than AI giving me solutions."
> (AI Non-user) *Illustrates: Avoid AI use to feel agency over content generation*

Theme 5: Ambivalence About AI Use

Central concept: Acknowledging the instrumental value of AI for writing tasks while nonetheless choosing not to use it.
This theme draws on the following analytical codes: *recognizes value writing improvement*; *recognizes value brainstorming*.

This theme captures a cognitively interesting stance: students who articulate *reasons to use AI* (grammar correction, idea generation, source finding) but are categorised as non-adopters. Rather than expressing outright rejection, these responses convey ambivalence: awareness of what AI could do, without the motivational tipping-point required to adopt it. The theme is small (2 analytical codes, 7 segments). Its defining feature is the *absence of a strong reason against* use, whereas Theme 4 features an *active reason against*.

The decoupling of AI awareness from AI use in this theme is consistent with B. Wang et al. (2022)'s observation that the four AILS dimensions (Awareness, Use, Evaluation, Ethics) do not necessarily develop in tandem: a student can recognise AI's potential utility without possessing the operational competence to deploy it effectively. A further moderating factor is the quality of preparation and prior experience with the assessment format (Struyven et al., 2003): students who have not previously encountered a process-based AI writing task may be uncertain not only about how to use AI, but about what appropriate use even looks like in this context, compounding the gap between awareness and action. Sadeghi and Abolfazli Khonbi (2015) found that this assessment format uncertainty amplifies inaction under ambivalence, suggesting that for these students the barrier is simultaneously one of practical AI competence and of assessment literacy. The contrast with Theme 4 is consequential: principled non-adopters in Theme 4 articulate a specific reason for non-adoption (grade risk, authorship, confidence); ambivalent non-adopters describe conditions for potential use without reaching a motivational threshold. This distinction suggests different intervention logics: for Theme 4 non-adopters, addressing concerns about grading fairness or AI reliability may be relevant; for ambivalent non-adopters, co-developing shared examples of what process-appropriate AI use looks like may lower the threshold more effectively than competence training alone.

Representative quote:

> "For grammar correction / for reference style / for new ideas (sometimes)."
> (AI Non-user) *Illustrates: Recognizes value writing improvement; Recognizes value brainstorming. (Note: this response lists potential uses of AI yet comes from a non-adopter — the listing itself constitutes the expression of ambivalence.)*

6.2 RQ3 — Learning Experience Reflections

Note: Themes for this section were derived via LLM-assisted analysis (phases 2–5) as described in Section 5.3, with researcher review and verification at phases 1 and 6. Results are presented in the same format as Section 6.1.

The reflections corpus comprised **26 responses** from DSE students. Analysis yielded six themes, summarised in Table 11.

Table 11: Summary of themes (learning experience reflections, $n = 26$ responses).

#	Theme	Group	Participants (n)
1	AI as a research and source discovery tool	Mixed	6
2	AI as a cognitive scaffold for clarity and structure	AI users	6
3	AI-enabled efficiency and reduced cognitive load	AI users	6
4	Concerns about AI reliability and hallucination	AI users	2
5	Preference for autonomous learning without AI	Mixed	4
6	Preservation of intellectual ownership and critical agency	AI users	3

Theme 1: AI as a Research and Source Discovery Tool

Central concept: AI accelerates the location, filtering, and synthesis of academic sources while preserving active intellectual engagement with content evaluation.
This theme draws on the following analytical codes: *research efficiency and source discovery*; *bounded AI use preserving learning*; *source discovery and synthesis support*.

This theme captures participants' use of AI primarily for efficient source discovery and research support: locating relevant articles, papers, and materials faster than traditional search methods would allow. Critically, this use case preserved active learning because participants remained responsible for evaluating source quality, relevance, and credibility, and for synthesising findings into their own arguments. The theme reflects a bounded, instrumental application of AI that enhanced research efficiency without displacing the intellectual work of interpretation and critical analysis.

This theme reveals a pedagogically sound adoption pattern where AI functions as a research accelerator rather than a content generator, allowing students to spend less time on search logistics and more time on higher-order synthesis and evaluation. It demonstrates that AI can enhance learning outcomes when its role is clearly circumscribed to information retrieval.

Frequency note: Supported by 6 participants (P05, P07, P10, P11, P13, P23), spanning both AI users and non-users.

Representative quotes:

> "I learned to use AI as a help for generating relevant sources to gain broader understanding of my chosen moral problem in the assignment. This was more time efficient than to use Google search engine."
> (P05, AI User)
> *Illustrates: Research efficiency and source discovery*

> "I learned a lot from the assignment as I only used AI for research purposes and it did not generate any text for me."
> (P07, AI User)
> *Illustrates: Bounded AI use preserving learning*

> "It was super helpful for finding and summarising related articles."
> (P23, AI User)
> *Illustrates: Source discovery and synthesis support*

Theme 2: AI as a Cognitive Scaffold for Clarity and Structure

Central concept: AI helps students clarify complex concepts and organise arguments while they retain responsibility for generating and evaluating ideas.
This theme draws on the following analytical codes: *scaffolding for clarity and critical thinking*; *AI as thinking partner for clarity*; *bounded scaffolding preserving intellectual ownership.*

This theme describes participants' use of AI to scaffold conceptual understanding and structural organisation of their work. Rather than generating content, AI functioned as a thinking partner: helping students clarify confusing material, organise ideas into coherent structures, and gain perspective on their own reasoning. Participants explicitly noted that this scaffolding enhanced their understanding of both the assignment rubric and the ethical concepts themselves, while they maintained intellectual ownership of their arguments and conclusions. The theme reflects a collaborative relationship where AI provides organisational and clarificatory support without substituting for critical thinking.

This theme illustrates how AI can function as a metacognitive tool that deepens learning by making thinking visible and organised, rather than replacing it. Scaffolding uses of AI may enhance rather than diminish learning outcomes when students remain actively engaged in idea generation and evaluation.

Frequency note: Supported by 6 participants (P04, P06, P18, P20, P23, P26), all AI users.

Representative quotes:

> "I completed the individual report using AI to help clarify complex concepts and guide my structure. The process helped me better understand the material while still allowing me to think critically and express my own ideas."
> (P06, AI User)
> *Illustrates: Scaffolding for clarity and critical thinking*

> "AI was able to help me clear my mind quickly. It was able to complement my lack of logic or creative parts to give me suggestions."
> (P18, AI User)
> *Illustrates: AI as thinking partner for clarity*

> "Sure, AI helped in structuring my paragraph but it did not do the assignment for me, it merely assisted."
> (P26, AI User)
> *Illustrates: Bounded scaffolding preserving intellectual ownership*

Theme 3: AI-Enabled Efficiency and Reduced Cognitive Load

Central concept: AI accelerates routine task completion and reduces cognitive burden, freeing mental resources for higher-order learning and engagement.
This theme draws on the following analytical codes: *efficiency in research process*; *reduced effort on revision and comparison tasks*; *efficiency enabling expanded learning.*

This theme captures participants' experience of AI as a time-saving and effort-reducing tool that streamlined routine processes such as searching, comparing sources, revising drafts, and formatting. Participants reported that this efficiency gain allowed them to focus cognitive resources on more substantive intellectual work (understanding concepts, developing arguments, and engaging deeply with material). The theme reflects a positive framing of efficiency gains as enabling rather than replacing learning, with participants emphasising that reduced effort on routine tasks translated into greater capacity for meaningful engagement with content.

Efficiency gains from AI adoption are not inherently detrimental to learning; they can enhance learning by reallocating cognitive effort from procedural to conceptual work. However, the learning benefit depends on how students redirect the freed cognitive capacity.

Frequency note: Supported by 6 participants (P05, P10, P11, P19, P21, P25), all AI users.

Representative quotes:

> "It helped me speed up the process of searching for specific articles/research papers, which accurately covered the topics I was trying to find coverage on."
> (P11, AI User)
> *Illustrates: Efficiency in research process*

> "I enjoyed having the flexibility to quickly compare topics and chose what I want to write about. And I also appreciated that I did not have to struggle too long when I wanted to change things in my drafts."
> (P21, AI User)
> *Illustrates: Reduced effort on revision and comparison tasks*

> "It helped me learn extra stuff without effort."
> (P19, AI User)
> *Illustrates: Efficiency enabling expanded learning*

Theme 4: Concerns About AI Reliability and Hallucination

Central concept: AI outputs require vigilant verification because the tool can generate plausible but false or irrelevant information.
This theme draws on the following analytical codes: *need for critical vigilance with AI outputs*; *AI hallucination in source discovery.*

This theme reflects participants' awareness of AI's limitations, specifically its tendency to generate hallucinations (false citations, non-existent papers, or information that appears credible but is inaccurate or off-topic). Participants who encountered this issue emphasised the necessity of critical verification and the risk that uncritical reliance on AI could lead them astray. The theme reveals a cautionary stance toward AI adoption, grounded in concrete experience of the tool's unreliability in specific contexts (e.g., finding domain-specific examples or obscure academic sources).

This theme highlights a critical limitation of AI as a research tool: while it can accelerate discovery, it cannot be trusted without human oversight. Effective AI use in academic contexts requires sustained critical vigilance and verification practices, adding a hidden cognitive cost that may offset efficiency gains.

Frequency note: Supported by 2 participants (P15, P17), both AI users. This is a lower-prevalence theme but analytically distinct and substantive, representing a minority but important cautionary perspective.

Representative quotes:

> "AI can definitely help a ton if you give it specific instructions. It has great capabilities, however it can easily lead you down a wrong path if you're not vigilant."
> (P15, AI User)
> *Illustrates: Need for critical vigilance with AI outputs*

> "It was really bad at finding examples for the specific case that I wanted to cover, because it was hallucinating papers which didn't exist, or were not associated with my topic."
> (P17, AI User)
> *Illustrates: AI hallucination in source discovery*

Theme 5: Preference for Autonomous Learning Without AI

Central concept: Some students deliberately avoid AI to preserve independent skill development, personal agency, and their own working methods.
This theme draws on the following analytical codes: *selective AI use to preserve autonomous learning*; *preference for autonomous method and control*; *valuing independent writing skill development.*

This theme captures participants who made a deliberate choice to complete work without AI assistance, citing pedagogical or personal reasons for this preference. These participants valued the learning that comes from working through material independently, developing their own writing and research skills, and maintaining control over their working processes and intellectual approach. The theme reflects a philosophical stance that autonomous, unaided work produces deeper learning and stronger skill development than AI-assisted approaches, and that personal ownership of method and process matters for learning.

This theme reveals that AI adoption is not universally preferred and that some students prioritise skill development and autonomy over efficiency. Individual differences in learning philosophy and goals shape adoption decisions in ways that aggregate adoption rates cannot capture.

Frequency note: Supported by 4 participants (P03, P08, P12, P16), spanning both AI users (who used it selectively) and non-users.

Representative quotes:

> "I would say I only use AI if I do not understand something or if I needed some skillful argumentation like in my final individual report where neither was the case here."
> (P03, AI User)
> *Illustrates: Selective AI use to preserve autonomous learning*

> "I think completing the assignment without AI was much better than if I were to have used it. I liked to do things in my own way, and I think that AI creates a different system, different to how I would structure my work."
> (P16, Non-AI User)
> *Illustrates: Preference for autonomous method and control*

> "I think not using AI helped me learn writing, but maybe using AI would be a good option to get insight into the issue outside of my own perspective."
> (P08, AI User)
> *Illustrates: Valuing independent writing skill development*

Theme 6: Preservation of Intellectual Ownership and Critical Agency

Central concept: Using AI does not diminish intellectual ownership; students retain responsibility for ideas, arguments, and critical evaluation regardless of tool use.
This theme draws on the following analytical codes: *maintaining intellectual ownership despite AI use*; *critical thinking and idea ownership preserved*; *learning and content ownership maintained*.

This theme reflects participants' explicit assertion that using AI did not compromise their intellectual integrity or critical agency. Participants emphasised that they remained the authors and originators of their ideas and arguments, with AI functioning as an assistant to execution rather than a generator of intellectual content. The theme reveals active negotiation of the boundary between legitimate assistance and problematic outsourcing, with participants carefully distinguishing between AI's role in helping them organise or clarify their own thinking (rather than generating ideas on their behalf). This theme addresses a key concern in academic integrity discourse by demonstrating that students can use AI while maintaining genuine intellectual ownership.

Intellectual ownership is not automatically lost through AI use; rather, it depends on how the tool is deployed and what boundaries students maintain. Explicit reflection on the distinction between assistance and outsourcing may be crucial for preserving learning value and academic integrity in AI-assisted work.

Frequency note: Supported by 3 participants (P04, P06, P26), all AI users.

Representative quotes:

> “Even though I used AI, it did not discredit the fact that the work was originally mine — the ideas were of my own. Sure, AI helped in structuring my paragraph but it did not do the assignment for me, it merely assisted.”
> (P26, AI User)
> *Illustrates: Maintaining intellectual ownership despite AI use*

> “The process helped me better understand the material while still allowing me to think critically and express my own ideas.”
> (P06, AI User)
> *Illustrates: Critical thinking and idea ownership preserved*

> “Through this way I still learned the content, just in a different way because I didn’t have to search for it in books or papers.”
> (P04, AI User)
> *Illustrates: Learning and content ownership maintained*

7 Discussion

7.1 Overview

This study examined whether writing self-efficacy (WSE), need for cognition (NFC), and AI literacy (AL) predict AI adoption and prompt diversity among graduate students completing individually graded ethics essays in a context where AI use was explicitly permitted and required to be documented. Data combined psychometric survey responses, verified AI use records, prompt-log annotations, and open-ended reflective responses. The approach was deliberately exploratory: the sample is small, the assessment context is novel, and the constructs have not previously been studied together in this setting. Findings should be read accordingly.

7.2 Summary of Findings

AI literacy was the only psychometric variable that meaningfully differentiated AI adopters from non-adopters, in both group comparisons and logistic regression; WSE and NFC showed no reliable group differences. Among AI users, none of the three predictors significantly predicted prompt diversity, though this analysis was severely underpowered ($n = 23$). Qualitatively, five adoption-related patterns emerged from DSE students' open-ended responses: *strategic efficiency* (AI as a scaffolding and time-saving tool), *skill compensation and augmentation* (AI as a substitute or extender for perceived writing or language gaps), *performative compliance* (adoption driven by the graded and explicitly encouraged nature of AI use), *intellectual agency* (principled non-adoption to preserve authorship or manage grade risk), and *ambivalence* (awareness of AI value without a sufficient motivational tipping-point to adopt). The learning experience reflections (RQ3) yielded six themes across three broad clusters: instrumental AI use (source discovery, cognitive scaffolding, and efficiency), critical caution about AI output reliability (hallucination awareness and verification concerns), and metacognitive regulation of intellectual agency (deliberate autonomous work and explicit ownership assertion alongside AI use).

7.3 The Role of AI Literacy

The finding that AI literacy (but not WSE or NFC) distinguished adopters from non-adopters is consistent with technology adoption theory. Davis (1989) Technology Acceptance Model identifies perceived ease of use and perceived usefulness as the primary drivers of voluntary technology adoption; both are closely related to the competence and confidence dimensions captured by the AILS (B. Wang et al., 2022). Students who understand how AI tools work, can operate them effectively, and can evaluate their outputs are better positioned to convert that knowledge into action. The qualitative data reinforce this reading: adoption was described in terms of practical capability ("it helps me output my ideas faster", "I can use it to check whether my output met the rubric"), suggesting that knowing *how* to use AI effectively was a prerequisite for adoption in this task context.

The null results for WSE and NFC are worth noting but should not be over-interpreted given sample size. Theoretically, WSE could operate in competing directions (students with high writing confidence may see AI as redundant, while students with lower confidence may rely on it for compensation), producing the null group difference observed (Bandura, 1997; Folwell et al., 2025). The qualitative RQ2 Theme 2 (skill compensation) is consistent with the low-WSE adoption pathway, and RQ2 Theme 4 (intellectual agency) with the high-WSE non-adoption pathway, suggesting both dynamics may be present without cancelling in a detectable aggregate effect. NFC's theoretical relevance to this context (students high in NFC resisting cognitive delegation) is also plausible (Cacioppo & Petty, 1982), but the low internal consistency of the NFC-6 in this sample ($\alpha = 0.56$) limits precision, and no reliable signal emerged.

7.4 Qualitative Patterns in Context

The qualitative findings add interpretive texture that the quantitative data alone cannot provide. The prevalence of *strategic efficiency* (RQ2 Theme 1: task efficiency and cognitive offloading) reasoning—offloading scaffolding tasks such as grammar, structure, and references to preserve cognitive resources for argumentation—mirrors the efficiency logic identified in studies on technology adoption in academic work (Davis, 1989) and is consistent with how AI tools have been described in broader higher education contexts (Bijker et al., 2024). This pattern does not imply passive or uncritical use; several students described iterative, prompt-refining interactions more consistent with augmentation than delegation.

Skill compensation and augmentation (RQ2 Theme 2) reflects a heterogeneous theme. Compensatory use (filling gaps in language or structural ability) and augmentative use (extending already-developed ideas through AI dialogue) represent meaningfully different relationships with the tool (the former more dependent, the latter more agentic), though both are coded under this theme because the boundary between them in student accounts was often unclear.

Performative compliance (RQ2 Theme 3) is perhaps the most contextually distinctive pattern: adoption motivated not by perceived learning value or personal competence, but by the fact that AI use was graded and encouraged. Struyven et al. (2003) noted that students' responses to novel assessment formats are shaped by their interpretation of instructor expectations; this theme suggests that a graded AI-integration context can produce surface-level, extrinsically motivated adoption that may not reflect genuine engagement with the tool's capabilities. This has implications for interpreting aggregate

AI adoption rates in such contexts.

Intellectual agency (RQ2 Theme 4: risk aversion and preservation of intellectual agency)—active non-adoption to preserve authorship or manage the risk that AI output would underperform one's own work—suggests that some students maintain a principled boundary around AI assistance even when it is permitted. Notably, the risk-management rationale included an implicit competence signal: confidence in one's own abilities was cited as a reason not to risk relying on AI. This echoes Folwell et al. (2025)'s finding that assessment confidence mediates engagement with novel task formats.

Ambivalence (RQ2 Theme 5) captures students who recognised potential AI value but did not adopt. This group is theoretically interesting because it decouples awareness from uptake: consistent with B. Wang et al. (2022)'s model, AI literacy may be insufficient at a moderate level to translate perceived usefulness into actual use. Situational barriers (timing, access, peer norms, uncertainty about what counts as appropriate use in an ethics course) may explain the gap, though the data do not allow this to be resolved.

7.5 Learning Experience Reflections in Context

The six RQ3 themes group into three broad functional clusters: *instrumental AI use* (Themes 1–3: source discovery, cognitive scaffolding, and reduced cognitive load), *critical caution about AI outputs* (Theme 4: hallucination awareness and verification), and *metacognitive regulation of intellectual agency* (Themes 5–6: deliberate autonomous work and explicit intellectual ownership assertion). A structurally important feature is that Themes 5 and 6 both appear among AI *users*, not exclusively non-adopters: students who used AI also reported choosing to work independently in some contexts and actively maintaining ownership of their ideas and arguments. This boundary-setting is invisible to adoption-rate data and suggests that the experience of working with AI in a graded task can itself prompt metacognitive reflection about appropriate tool use.

The instrumental cluster (Themes 1–3) is consistent with the efficiency and scaffolding rationales documented in the RQ2 adoption themes, but the register shifts from motivation (why students adopted AI) to phenomenology (what using it actually enabled). The pattern across these three themes—accelerating source location while retaining responsibility for evaluation (Theme 1), using AI as a thinking partner while preserving argument ownership (Theme 2), and freeing cognitive capacity from revision for higher-order engagement (Theme 3)—reflects the active, bounded AI use that L. Wang and Luo (2023) theorise as necessary for AI interaction to benefit learning: self-efficacy mediates the AI–achievement relationship because it shapes whether students remain intellectually engaged with AI outputs or simply defer to them. The scaffolding experiences in Theme 2 align with Zhang et al. (2025)'s finding that AI-generated feedback is positively associated with writing self-efficacy and engagement when students treat it as a developmental resource rather than a substitute for reasoning. A note of caution is warranted, however: Wu et al. (2025) found experimentally that human–AI collaboration enhanced objective task performance but measurably reduced intrinsic motivation. The positive efficiency experiences in Themes 1–3 may carry a motivational cost that students' retrospective reflections, framed around a completed task, could not detect. Whether AI-assisted efficiency in graded essay writing reduces longer-term investment in developing unaided writing competence is an open question; the present data cannot resolve it, but longitudinal designs would be needed to do so.

Theme 4 (hallucination concerns) is analytically important despite its low prevalence ($n = 2$). Hua and Cunningham (2025) found in a thematic analysis of first-year students' AI reflections that students widely view AI as a flexible academic aid but show limited spontaneous awareness of its accuracy limitations. The two students in this corpus who reported direct encounters with fabricated citations or hallucinated sources represent a counter-pattern: firsthand experience of AI failure appears to generate the critical vigilance that policy-level warnings about AI limitations are unlikely to produce in the abstract. *This implies—speculatively, as no direct evidence tests this conjecture—that deliberate instructional exercises exposing students to AI failure modes (for example, evaluating AI-generated reference lists against verified databases) may build verification habits more reliably than declarative caution.*

Themes 5 (autonomous learning preference) and 6 (intellectual ownership assertion) together complicate any simple framing of AI adoption as uniformly beneficial or detrimental. Theme 5 participants, spanning AI users and non-users, described deliberate choices to preserve independent skill development—a metacognitive judgment that unaided effort is itself the learning mechanism in some contexts. Folwell et al. (2025) found that assessment formats supporting choice and self-direction promote autonomy orientation; the same logic appears to extend to technology-use decisions within an assessment task, not only to assessment format choices at a design level. Theme 6 represents perhaps the most theoretically significant pattern: AI users who spontaneously and explicitly distinguished between AI's role in execution and their ownership of ideas and arguments. Wu et al. (2025)'s concern about AI collaboration undermining intrinsic motivation makes this distinction practically important—students who maintain psychological ownership of their intellectual work may be better insulated against motivational displacement from AI use. Whether this boundary-awareness can be cultivated or whether it reflects pre-existing critical dispositions is an important open question. Hua and Cunningham (2025)'s finding of limited spontaneous critical reflection in comparable populations suggests that ownership-assertion of the kind seen in Theme 6 is not typical; making the distinction between AI as execution tool and AI as idea generator an explicit component of task framing and pre-task scaffolding is worth exploring as a leverage point, though evidence for this specific intervention does not yet exist.

7.6 Limitations

Several limitations constrain the conclusions that can be drawn. The sample is small ($N = 45$; $n = 23$ for RQ2), and all quantitative results should be treated as preliminary. The study is observational and cross-sectional, precluding causal claims. The two courses differed substantially in their AI adoption rates (DSE: 85%; PEAI: 28%), and course-level contextual factors (instructor framing, task design, peer culture) likely account for part of this variation but could not be disentangled in the current design. Qualitative data were available only for DSE students who completed M2, limiting the generalisability of the thematic analysis. The NFC scale showed low internal consistency in this sample ($\alpha = 0.56$), restricting the precision of inferences about that construct. Grade data were unavailable for this analysis cycle, so RQ2's extension to essay performance remains untested. Finally, AI adoption was operationalised as a binary verified record rather than a behavioural measure of engagement quality; the prompt-log data provide some insight into the character of use, but coverage is limited to DSE AI users only.

7.7 Future Directions

Several extensions of this line of research are feasible and warranted. A multi-institution replication with a larger sample would allow reliable estimation of the AI literacy effect and meaningful testing of prompt diversity predictors. A longitudinal design tracking the same students across multiple assessed tasks would clarify whether adoption is stable or context-sensitive. The substantial course-level variation in adoption rates points to the value of studying instructor framing and task design as moderators: an ethics essay context may produce different adoption dynamics than, for example, a data analysis or literature review task. Behavioural measures of AI engagement (such as prompt complexity, revision patterns, or output similarity) would complement self-report and log-based measures. Finally, the ambivalence theme suggests that understanding the motivational and situational barriers between AI awareness and AI use is a productive focus for future qualitative work, particularly as tools become more accessible and AI use norms in higher education continue to evolve.

7.8 Practical Implications for Educators

The findings carry several actionable implications for educators designing or evaluating AI-integrated assessment tasks, particularly in contexts where AI use is explicitly permitted and graded.

AI literacy as a design prerequisite, not an afterthought. The most consistent finding is that AI literacy (not writing confidence or cognitive style) differentiates adopters from non-adopters. This suggests that when an assessed task requires or encourages GenAI use, students' capacity to engage meaningfully with those tools is unevenly distributed at the outset. Treating AI literacy as a baseline competence that students bring into the task by default risks producing systematically inequitable engagement patterns: students with higher AI literacy are better positioned to benefit from the tools, while those with lower literacy may either disengage, adopt superficially, or adopt in performatively compliant ways without genuine learning value. Embedding brief, task-specific AI literacy preparation (not generic tool demonstrations, but guided engagement with the type of AI-assisted process the task actually requires) before a graded AI-integrated assignment may reduce this baseline disparity.

Transparency about what graded AI use looks like. The performative compliance theme (RQ2 Theme 3) indicates that some students interpreted the encouraged and graded nature of AI use as an implicit performance expectation and adopted accordingly, without clear intrinsic motivation. The ambivalence theme (RQ2 Theme 5) suggests a related but distinct failure: students who could articulate AI's value nonetheless did not adopt, partly because they were uncertain about what appropriate use looked like in an ethics-course context. Both patterns point to the same intervention: providing concrete, worked examples or scaffolded demonstrations of what process-appropriate AI use entails in the specific assignment context, rather than leaving students to infer this from general policy language about permitted use.

Assessment design and the autonomy of non-adopters. The intellectual agency theme (RQ2 Theme 4) is a reminder that principled non-adoption is a legitimate response to an AI-integrated assignment. Students who choose not to use AI to preserve authorship, manage grade risk, or exercise writing confidence are not failing to engage with the assessment design: they are exercising judgment. Assessment designs that penalise non-adoption implicitly (for example, by weighting the AI-use documentation grade heavily without allowing equivalent weight for written argumentation quality) may undermine the autonomy orientation that RQ2 Theme 4 non-adopters express. A defensible design should ensure that non-adoption carries no inherent grade penalty, so that students' decisions are genuinely voluntary rather than constrained by perceived grading incentives.

Differential support logics for different non-adoption profiles. The qualitative data suggest that not all non-adopters are alike: RQ2 Themes 4 and 5 non-adopters present different barriers and therefore warrant different responses. For students who actively declined AI use on principled grounds (agency, grade risk, authorship), the priority is ensuring that the grading design genuinely supports voluntary non-adoption rather than inadvertently penalising it. For ambivalent non-adopters (RQ2 Theme 5), the barrier is more operational: insufficient competence or insufficient clarity about what good AI use looks like. Targeted practical scaffolding (co-developing examples, structured prompt practice, or peer demonstration) is more likely to move these students than reassurance alone. Recognising these two profiles as distinct, rather than treating all non-adopters as a uniform group, allows for more targeted instructional support.

Building hallucination literacy and intellectual ownership awareness. The RQ3 learning reflections point to two instructional needs that go beyond the adoption decision. First, critical awareness of AI's accuracy limitations (hallucination, source fabrication) appears to be experience-driven: students who encountered AI failure first-hand developed verification habits, while abstract policy warnings about AI reliability may not produce the same effect (Hua & Cunningham, 2025). Deliberate instructional exercises that expose students to AI failure modes—evaluating AI-generated reference lists against verified databases, or comparing AI summaries against original sources—may build verification practices more reliably than declarative caution alone. *This specific intervention design is speculative; no direct evidence currently exists.* Second, the spontaneous intellectual ownership assertions in RQ3 Theme 6 suggest that some students actively maintain clear boundaries between AI assistance and idea origination—but this appears to be a reflective minority disposition, not a default (Hua & Cunningham, 2025). Making this distinction an explicit component of task framing—for example, requiring students to annotate which steps in their process were AI-assisted and which were independently reasoned—can scaffold the distinction and generate visible evidence of genuine intellectual engagement.

7.9 Conclusion

This study offers an initial empirical account of AI adoption in a novel assessment context: one in which students are not merely permitted to use GenAI but are assessed on how they do so. The pattern of results points toward AI literacy, rather than writing confidence or cognitive style, as the more proximal individual-level predictor of adoption. The qualitative picture is richer: adoption and non-adoption were both purposeful, shaped by a mix of practical competence, skill awareness, assessment incentives, and principled positions on intellectual ownership. The learning experience reflections add a further layer: even among students who adopted AI, engagement ranged from bounded instrumental use that preserved intellectual ownership to surface-level efficiency seeking, and critical awareness of AI's limitations was present in only a small minority. Together, these findings suggest that integrating GenAI into graded assessment is not a neutral design choice: it interacts with students' existing competences and values in ways that produce heterogeneous and sometimes unintended responses. The practical implications discussed above offer starting points for educators seeking to design AI-integrated assessments that are both equitable in access and coherent in learning intent.

Funding

This research was co-funded by the 4TU.Centre for Engineering Education and BOOST! (Eindhoven University of Technology).

CRediT statement

The following contributions are defined according to the CRediT (Contributor Roles Taxonomy): Conceptualization: MO; Data Curation: MO; Formal Analysis: MO; Funding Acquisition: MO, RC, CZ, GB, BS; Investigation: MO; Methodology: MO; Project Administration: MO; Resources: MO, CZ; Supervision: MO, RC, GB; Validation: MO, RC; Visualization: MO; Writing – Original Draft: MO; Writing – Review & Editing: MO, RC.

Acknowledgements

We thank Suzanne Groothuijsen and Ludo van Meeuwen for their insightful comments at several stages of this project.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used several Generative AI chatbots (Microsoft Copilot, Gemini, Claude) in order to structure and improve the readability of the text throughout. These tools were also used to assist with R coding (e.g., improved visualization and structuring, efficient refactoring), and in the development of prototype software tools to structure (and partially scaffold in one of the analyses) the qualitative analysis process. After using this tool/service, the author(s) critically reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Afrifa-Yamoah, E., Bashiru, S., & Agyei, S. K. (2024). Assessment design and practices toward holistic learning of higher education students: Empirical evidence via path analysis modelling approach. *Journal of the Scholarship of Teaching and Learning*, *24*(3). https://doi.org/10.14434/josotl.v24i3.35672

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Bijker, R., Merkouris, S. S., Dowling, N. A., & Rodda, S. N. (2024). ChatGPT for automated qualitative research: Content analysis. *Journal of Medical Internet Research*, *26*, e59050. https://doi.org/10.2196/59050

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

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Coelho, G. L. H., Hanel, P. H. P., & Wolf, L. J. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, *27*(8), 1870–1885. https://doi.org/10.1177/1073191118793208

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Fauzi, C., Rahmani, E. F., & Utimadini, N. J. (2025). Beyond technology acceptance: An interplay of self-efficacy, language proficiency, and ChatGPT adoption from a TAM perspective. *JPP (Jurnal Pendidikan Progresif)*, *15*(3), 1970–1988. https://doi.org/10.23960/jpp.v15i3.pp1970-1988

Folwell, K., Bartimote, K., & Southgate, E. (2025). Assessment by engagement: Building confidence and autonomy in the first year. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2025.2483268

Hua, J., & Cunningham, J. (2025). Two years after ChatGPT: A thematic analysis of first-year students' reflections on AI tool use in higher education. *Journal of Information, Communication and Ethics in Society*. https://doi.org/10.1108/jices-05-2025-0118

Hwang, Y. J., & Wu, Y. (2025). The influence of generative artificial intelligence on creative cognition of design students: A chain mediation model of self-efficacy and anxiety. *Frontiers in Psychology*, *15*, 1455015. https://doi.org/10.3389/fpsyg.2024.1455015

Li, C., Wang, Q., & Chen, H. (2024). Demystifying anxiety and demotivation in online assessment: A focus on the impacts on academic buoyancy and autonomy. *BMC Psychology*. https://doi.org/10.1186/s40359-023-01511-w

Miller, M. (2024). Precarious partnership: Student perceptions of generative AI in post-secondary learning. *Open/Technology in Education, Society, and Scholarship Association Journal*, *4*(1). https://doi.org/10.18357/otessaj.2024.4.1.74

Oliveira, M., Zednik, C., Bombaerts, G., Sadowski, B., & Conijn, R. (2025). Assessing students' DRIVE: A framework to evaluate learning through interactions with generative AI. *Computers and Education: Artificial Intelligence*, *9*, 100497. https://doi.org/10.1016/j.caeai.2025.100497

Sadeghi, K., & Abolfazli Khonbi, Z. (2015). Iranian university students' experiences of and attitudes towards alternatives in assessment. *Assessment & Evaluation in Higher Education*, *40*(5), 641–665. https://doi.org/10.1080/02602938.2014.941324

Struyven, K., Dochy, F., & Janssens, S. (2003). Students' perceptions about new modes of assessment in higher education: A review. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 171–223). Springer. https://doi.org/10.1007/0-306-48125-1_8

Sukirman, S., Supriyanto, E., Setiawan, A., Chamsudin, A., Yuliana, I., & Wantoro, J. (2024). Exploring student perceptions and acceptance of ChatGPT in enhanced AI-assisted learning. *2024 International Seminar on Intelligent Modeling and Learning (SIML)*. https://doi.org/10.1109/siml61815.2024.10578145

Sun, Y., & Wang, S. (2020). Development and validation of the questionnaire for english writing self-efficacy. *RELC Journal*, *53*(3), 645–660. https://doi.org/10.1177/0033688220966532

Wang, B., Rau, P.-L. P., & Yuan, T. (2022). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, *42*(9), 1324–1337. https://doi.org/10.1080/0144929X.2022.2072768

Wang, L., & Luo, J. (2023). The relationship between student interaction with generative artificial intelligence and learning achievement: Serial mediating roles of self-efficacy and cognitive engagement. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2023.1285392

Wu, S., Liu, Y., Ruan, M., Chen, S., & Xie, X.-Y. (2025). Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. *Scientific Reports*, *15*, 15105. https://doi.org/10.1038/s41598-025-98385-2

Xie, Y., & Wang, S. (2025). Generative artificial intelligence in entrepreneurship education enhances entrepreneurial intention through self-efficacy and university support. *Scientific Reports*, *15*, 24079. https://doi.org/10.1038/s41598-025-09545-3

Zhang, H., Zhang, J., & Zhao, Y. (2025). Unveiling the writing self-efficacy and its relationship with writing engagement based on generative AI feedback. *2025 IEEE International Conference on Advanced Learning Technologies (ICALT)*. https://doi.org/10.1109/icalt64023.2025.00052

A Scale Items

A.1 Writing Self-Efficacy Scale (QEWSE; Sun and Wang (2020))

Scale: 1 (*Low self-efficacy*) to 7 (*High self-efficacy*). *Instruction:* "Please read the following questions carefully and make an accurate evaluation of your current command of English no matter whether you are studying it or not. These questions are designed to measure your judgment of your capabilities, so there are no right or wrong answers." (R) = reverse-scored.

#	Item
Ideation	
1	I can think of many ideas for my writing.
6	I can put my ideas into writing.
11	I can think of appropriate words to describe my ideas.
12	I can focus on the main ideas when writing.
Organisation	
2	I can organize sentences into a paragraph to express an idea.
7	I can organize different paragraphs into a composition.
20	I can write a paragraph in a cohesive way.
27	I can write a paragraph in a coherent way.
Grammar	
3	I can correctly spell all the words in the compositions I write.
8	I can correctly use verb tenses in English writing.
9	I can make new sentences with given words.
21	I can write a sentence with proper grammatical structures.
26	I can fix my grammar errors.
Revision	
23	I can revise my writing to make it better.
Self-regulation	
5	I can focus on my writing for at least one hour.
15	I can finish writing assignments on time.
18	I can plan what I want to say before I start writing.
19	I can avoid distractions while I write.
24	I can control my frustration when I write.
25	I can keep writing even when it's difficult.
Other / Mixed	
4	I can compose messages in English on social media (e.g., WeChat and blogs).
10	I can write an expository paragraph in English.
13	I can write an argumentative paragraph in English.
14	I can write email messages in English.
16	I can write a descriptive paragraph in English.
17	I can write diaries in English.
22	I can write a narrative paragraph in English.

A.2 Need for Cognition Scale (NFC-6; Coelho et al. (2020))

Scale: 1 (*Extremely uncharacteristic of me*) to 5 (*Extremely characteristic of me*). *Instruction:* "Please respond to the questions below by indicating how much each statement describes you." (R) = reverse-scored.

#	Item
1	I would prefer complex to simple problems.
2	I like to have the responsibility of handling a situation that requires a lot of thinking.
3 **(R)**	Thinking is not my idea of fun.
4 **(R)**	I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.
5	I really enjoy a task that involves coming up with new solutions to problems.
6	I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

A.3 AI Literacy Scale (AILS; B. Wang et al. (2022))

Scale: 1 (*Strongly disagree*) to 7 (*Strongly agree*). *Instruction:* "How much do you agree with each of the following statements?" (R) = reverse-scored. Factor loadings from B. Wang et al. (2022) are given in parentheses.

Dimension	Item ID	Item
Awareness	AW_1	I can distinguish between smart devices and non-smart devices. (.72)
	AW_8 **(R)**	I do not know how AI technology can help me. (.64)
	AW_9	I can identify the AI technology employed in the applications and products I use. (.70)
Use	US_1	I can skilfully use AI applications or products to help me with my daily work. (.72)
	US_3 **(R)**	It is usually hard for me to learn to use a new AI application or product. (.66)
	US_5	I can use AI applications or products to improve my work efficiency. (.72)
Evaluation	EV_2	I can evaluate the capabilities and limitations of an AI application or product after using it for a while. (.71)
	EV_3	I can choose a proper solution from various solutions provided by a smart agent. (.72)
	EV_6	I can choose the most appropriate AI application or product from a variety for a particular task. (.78)
Ethics	ET_1	I always comply with ethical principles when using AI applications or products. (.76)
	ET_2 **(R)**	I am never alert to privacy and information security issues when using AI applications or products. (.60)
	ET_5	I am always alert to the abuse of AI technology. (.73)

B LLM Prompt Templates

This appendix reproduces the exact system and user prompts submitted to the Claude API (`claude-haiku-4-5-20251001`) during the LLM-assisted thematic analysis of the learning experience reflections corpus (RQ3; Section 5.3.2). Where multiple bullet items appear on a single line in the original API call (a consequence of how strings were concatenated in the analysis script), they are separated here by line breaks for readability; the text is otherwise verbatim. Dynamic values that varied across calls (individual response texts, accumulated code lists, and the JSON output of prior phases) are indicated by descriptive placeholders in *angle brackets*.

Research context string (RQ3 — learning experience reflections):

This study examines how graduate students experienced the process of writing an individual ethics essay — with or without AI assistance. The reflective data illuminates the perceived learning implications of AI adoption: what do students gain, lose, or feel uncertain about when they use (or avoid using) AI in academic writing? This connects to the broader question of how AI tools shape the learning process in higher education, beyond mere adoption rates.

Phase 2 — Initial Coding

System prompt (temperature = 0.2; submitted once per response):

```
You are an expert qualitative researcher applying Braun & Clarke's
(2006) reflexive thematic analysis to educational data.

Research context: <context string>

Your task: Generate 2-5 semantic codes for the participant <item
label>. Each code must:
- Be a concise descriptive phrase (3-8 words)
- Capture one distinct meaning unit
- Stay grounded in what the participant actually said
- Distinguish surface content from latent significance where possible

Return ONLY a valid JSON array of code strings.
Example: ["concern about academic integrity", "efficiency
motivation"]
```

User prompt (one call per participant response):

```
Participant: <AI User / Non-AI User>
Response: "<verbatim response text>"

Generate codes.
```

Phase 3 — Candidate Theme Generation

System prompt (temperature = 0.4; single call with all codes):

```
You are an expert qualitative researcher applying Braun & Clarke's
(2006) reflexive thematic analysis.

Research context: <context string>

You have coded all participant responses. Now identify 3-6 candidate
themes. Each theme must:
- Capture a coherent, meaningful pattern across multiple participants
- Be analytically relevant to the research questions
- Be distinct from other themes
```

- Have data support from more than one response

For each theme return a JSON object with these exact fields:

```
{"name":"Evocative theme name",
"central_concept":"One sentence capturing the core",
"description":"2-3 sentences describing what the theme covers and why
it matters analytically",
"supporting_codes":["code1","code2"],
"ai_user_pattern":"How AI users manifest this theme (or null)",
"non_user_pattern":"How non-users manifest this theme (or null)"}
```

Return ONLY a valid JSON array of theme objects.

User prompt:

Codes across all *<N>* participants:

[P01 | AI User] *<code1>* | *<code2>* | *<code3>*
P02 | Non-AI User

<code1> | *<code2>*
(one line per participant)

Phase 4 — Theme Review

System prompt (temperature = 0.3; single call with themes + full corpus):

You are an expert qualitative researcher applying Braun & Clarke's (2006) reflexive thematic analysis.

Your task: Review the candidate themes against the full corpus of raw responses. Ensure that:

1. Each theme is well-supported (evidence from multiple distinct responses)
2. Themes are internally coherent (responses within a theme share a meaningful core)
3. Themes are distinguishable (minimal substantive overlap between themes)
4. Together the themes adequately cover the key patterns in the data

Actions you may take: merge overlapping themes; split a theme if it conflates two distinct ideas; remove a theme if it lacks data support; rename if misleading.

Return the reviewed themes in the SAME JSON format as the input, adding a "review_notes" field per theme (briefly stating what, if anything, was changed and why). Return ONLY a valid JSON array.

User prompt:

```
CANDIDATE THEMES:
<JSON array output from Phase 3>
FULL CORPUS:
P01 | AI User
: <verbatim response text>
P02 | Non-AI User
: <verbatim response text>
(one block per participant, separated by blank lines)
```

Phase 5 — Final Theme Definition

System prompt (temperature = 0.3; single call with reviewed themes + full corpus):

```
You are an expert qualitative researcher writing up a Braun & Clarke
thematic analysis for publication in an educational technology
journal.
Research context: <context string>
For each theme, write a full definition as it would appear in the
Results section. Select 2-3 verbatim quotes from the corpus that
best illustrate the theme.
Return a JSON array. Each element must have EXACTLY these fields:
{"name":"Final theme name",
"central_concept":"One sentence -- the core organizing concept",
"definition":"Rich 3-4 sentence definition of what this theme is and
what it reveals",
"analytical_insight":"1-2 sentences on what this theme contributes to
understanding",
"frequency_note":"Concise note on how many / which type of
participants this theme covers",
"representative_quotes":[
  {"participant_id":"P01","used_ai":"AI User",
  "quote":"exact verbatim text",
  "code_illustrated":"code this quote exemplifies"}
]}
Return ONLY valid JSON. Use exact verbatim text for quotes -- do not
paraphrase.
```

User prompt:

REVIEWED THEMES:

<JSON array output from Phase 4>

FULL CORPUS:

P01 | AI User

: *<verbatim response text>*

P02 | Non-AI User

: *<verbatim response text>*

(one block per participant, separated by blank lines)