

AI Writing Utility, Ethical Concerns, Academic Self-Reliance, and Responsible AI Writing Practices among University Students: An Exploratory Psychometric Study

Zuraini Mat Issa¹, Saidatul Afzan Abdul Aziz^{1*}, Ammar Zaki Abd Aziz¹, Alexson Tuah Anak Demenny Ngumbang¹, Hafizah Hassan¹, Ahmad Ihsan Mohd Yassin²

¹Faculty of Hotel and Tourism Management, Universiti Teknologi MARA Cawangan Selangor, Kampus Puncak Alam, 42300 Bandar Puncak Alam, Selangor, Malaysia.

²Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

*Corresponding Author

DOI : <https://doi.org/10.47772/IJRISS.2026.100800938>

Received: 3 September 2026; Accepted: 8 September 2026; Published: 21 September 2026

ABSTRACT

As the use of generative artificial intelligence (GenAI) in academic writing has grown, there is a need for reliable indicators that reflect the perceived value of GenAI in academic writing as well as students' ethical issues, independence in writing, and responsible use. This study was aimed at assessing and improving a 32-item questionnaire that assesses university students' attitudes and usage of AI-generated text in academic writing. A cross-sectional survey of 135 students at Universiti Teknologi MARA, Malaysia, was performed using a quantitative method. Exploratory factor analysis (EFA) based on polychoric correlation matrix, parallel analysis and oblique rotation was used to explore the underlying structure of the instrument, which was followed by reliability analysis, correlation analysis and multiple regression analysis. The EFA reduced the original five-domain instrument to a four-factor, 21-item instrument, which accounted for 71.1% of the variance, namely AI Writing Utility and Usability (AWUU), Ethical Concerns (EC), Academic Writing Self-Reliance (AWSR), and AI Writing Integrity Practice (AIWIP). The four dimensions demonstrated satisfactory internal consistency with Cronbach's alpha ranging from .822 to .934 and McDonald's omega total ranging from .894 to .964. The students' AIWIP, AWUU, and EC scores were relatively high, while the AWSR score was comparatively low. The model of regression accounted for 34.3% of the variance of AIWIP. In the primary regression model, AWUU was the only construct independently associated with AIWIP ($B = .495$, $p = .003$), while EC and AWSR were not significant. The association involving EC was sensitive to the exclusion of influential observations. These findings provide preliminary support for a parsimonious four-dimensional instrument and suggest that students' practical appraisal of GenAI may be associated with stronger self-reported AI writing integrity practices.

Keywords: generative artificial intelligence; academic writing; responsible AI use; exploratory factor analysis; instrument development

INTRODUCTION

Generative artificial intelligence (GenAI), especially large language model tools such as ChatGPT, has been adopted rapidly in higher education and is already reshaping how students approach academic writing. Students utilize these systems throughout the writing process, from brainstorming ideas to creating outlines, finding or combining information, summarizing the information, drafting, revising the structure, editing grammar and proofreading language (Barrett & Pack, 2023; Chan & Hu, 2023; Kim et al., 2024; Wang, 2024). Recent reviews and empirical studies suggest that students find these tools useful for enhancing efficiency, formal text quality, and writing support and feedback, especially for language learners and novice writers (Imran & Almusharraf, 2023; Y. Liu et al., 2024; Sanz-Tejeda et al., 2026). Meanwhile, the pedagogic value of AI-generated writing is all in the ambiguity: AI can serve as a resource to support writing, yet it also can be a tool to circumvent intellectual work, undermine authorship, and challenge traditional conceptions of assessment and academic

integrity (Barrett & Pack, 2023; Chatto & Logronio, 2025; Perkins, 2023; Wise et al., 2024). The potential for hallucinated or inaccurate information, opaque training data, bias, over-reliance, and the potential for critical engagement to be undermined further indicate that the use of AI-generated text by students is not a niche technology adoption issue, but a core higher-education research question related to writing, learning, and academic standards (Chan & Hu, 2023; Kim et al., 2024; Sanz-Tejeda et al., 2026).

Therefore, it is not possible to conclude that students' attitudes towards AI-generated texts in academic writing are just positive or negative. It is relevant to the technology-acceptance perspectives because perceived usefulness and perceived ease of use are clearly shown to influence willingness to use GenAI tools (Chan & Hu, 2023; Liu & Chaojun, 2023; Niekerk et al., 2025), and students report that they are accessible, convenient, and helpful in generating ideas, improving language, and receiving feedback. However, perceptions are also moral, thinking and feeling in writing contexts. Students differentiate between appropriate and inappropriate uses in various writing tasks, and they do not view brainstorming, outlining, revising, language support and full-text generation in the same way (Barrett & Pack, 2023; Johnston et al., 2024). While students' perceptions vary, several studies have documented ongoing worries about plagiarism, authorship, originality, transparency, and whether or not the use of AI for writing is considered dishonest or not (Liu et al., 2024; Nelson et al., 2025; Perkins, 2023). Perceptions are also related to confidence and self-regulation. While some students find AI to be a writing empowerer, confidence-booster, and anxiety reliever, others are concerned that relying on AI-generated content could lead to a decline in their own writing abilities, critical thinking skills, or the loss of their writing voice (Chatto & Logronio, 2025; Khojasteh et al., 2025; Wang, 2024). This multidimensionality is significant as students' confidence in their own writing seems to be related to their use of GenAI and their stance on its use in academic settings, indicating that beliefs about learning, integrity, and competence should be taken into account in interpreting students' responses to GenAI usage (Johnston et al., 2024; Nelson et al., 2025).

Despite the rapid growth of research on GenAI in higher education, the measurement base for studying students' perceptions remains conceptually uneven. Existing instruments have examined broad attitudes toward generative AI in education, general acceptance, or AI ethics, and recent psychometric studies show that such tools can yield reliable factor structures for those broader constructs (Arlakov & Miklyaeva, 2026; Jang et al., 2022; Marengo et al., 2025). Other emerging instruments address AI and academic writing more directly and indicate that perceptions in this domain may include perceived effectiveness, ethical and authenticity concerns, writing-process support, feedback and enhancement, and affective or motivational impact (Khojasteh et al., 2025). Yet this does not eliminate the present measurement problem. Scales developed for general educational AI use may not adequately represent perceptions of AI-generated text in academic writing, where the focal object is not AI in the abstract but text that can be co-produced, revised, submitted, concealed, or disclosed within high-stakes literacy practices (Arlakov & Miklyaeva, 2026; Kim et al., 2024; Marengo et al., 2025; Van Niekerk et al., 2025). Nor can the validity of adapted items be assumed when questionnaires are transferred across populations, disciplines, languages, technologies, or educational settings, because dimensions that appear theoretically distinct may empirically overlap, fragment, or fail to cohere in a new context (Arlakov & Miklyaeva, 2026; Jang et al., 2022; Marengo et al., 2025). Substantive interpretation, in other words, should only occur after an empirical psychometric evaluation. In particular, exploratory factor analysis (EFA) can be used at this stage to investigate whether the proposed items form interpretable and distinguishable dimensions, whether some of the items have poor loadings or cross loadings, and whether the dimensions exhibit acceptable internal consistency for further analysis (Jang et al., 2022; Khojasteh et al., 2025; Marengo et al., 2025).

In this context, there is a need for a more conceptually precise, reliable and parsimonious tool to investigate students' perception of AI-generated text in academic writing in more detail with an empirical basis. The present study aims to fill this gap by systematically examining the psychometric properties of a proposed 32-item questionnaire that captures students' perceptions and experiences with AI-generated text, its underlying factor structure, internal consistency reliability of the resulting factors, and modification of the measurement structure as warranted by the empirical evidence. The study addresses the persistent demand for more transparent and student-focused evidence to guide institutional policies, pedagogical advice, and the ethical use of AI in the context of writing-intensive higher education (Chan & Hu, 2023; Nelson et al., 2025; Wang, 2024). Its contribution is two-fold: It contributes to the measurement foundations of research on GenAI and academic writing by providing preliminary empirical evidence concerning the dimensional structure and internal consistency of the proposed questionnaire in this particular context, and it uses the refined instrument to explore

students' perceptions of AI-generated text, connecting psychometric evaluation and instrument refinement with substantive inquiry instead of seeing them as separate tasks. This integration is significant as strong findings regarding perceived usefulness, ease of use, ethical issues, academic self-efficacy, and actual use of AI-generated writing rely on the validity of the instrument used to capture the constructs. Therefore, the aims of this study are:

1. To examine the factor structure and psychometric properties of the proposed 32-item questionnaire measuring students' perceptions and practices concerning AI-generated text in academic writing.
2. To refine the questionnaire and derive a parsimonious measurement structure with acceptable internal consistency.
3. To apply the refined instrument to describe students' perceptions and examine associations between the identified dimensions and responsible AI writing practice.

METHODOLOGY

Research Design

A quantitative cross-sectional survey design was used to assess university students' perceptions and practices of the use of AI-generated text in university writing. The study started with the Technology Acceptance Model (TAM) and other literature in the field of education, and the questionnaire was designed to obtain the perceived usefulness, perceived ease of use, ethical issues, academic confidence, and perceived use or impact of AI-generated text in academic writing.

A cross-sectional approach was deemed suitable as the data were gathered from the respondents at a specific timeframe in the study to understand their current perceptions and experiences with GenAI tools. The design also allowed for the analysis of the basic dimensional structure of the questionnaire and the interrelationships between the dimensionally derived constructs. This study was an observational cross-sectional study, and therefore, the results of the analyses were interpreted as associations rather than causation.

Importantly, the dimensional structure of the instrument was empirically evaluated by exploratory factor analysis (EFA), although the initial questionnaire was organised around five theoretically proposed domains. Therefore, the final measurement structure was not necessarily the same as the five-domain structure originally proposed, but was a result of a theoretical understanding and empirical evidence.

Population and Sampling Technique

The target population was students involved in academic writing in Universiti Teknologi MARA (UiTM) Cawangan Selangor, Kampus Puncak Alam, who were undertaking their coursework in the university. The accessible population comprised students from eight faculties on the campus. Purposive sampling was used, where students who were aware of or had some experience with GenAI tools for academic or writing purposes were invited to participate. The participation was voluntary, and there was no set number of responses that would be collected when the questionnaires were distributed.

Data collection took place from the end of October 2025 to the end of January 2026. After data collection, 135 valid responses were kept for data analysis. An a priori power analysis using G*Power was used to determine the minimum sample required for the originally hypothesised multiple regression model involving four predictors. Assuming a medium effect size ($f^2 = .15$), an alpha level of .05, statistical power of .95, and four predictors, the analysis indicated a minimum required sample of 129 respondents. The final sample of 135 therefore exceeded the minimum requirement for the originally proposed regression analysis. Importantly, the G*Power calculation related to the originally proposed regression analysis should not be interpreted as a sample-size justification for EFA. The adequacy of a sample for factor analysis depends on several features of the data and measurement model rather than on a single participant-to-item ratio. Nevertheless, given the modest sample of 135 respondents relative to the initial 32-item pool, the factor structure identified in this study should be regarded as preliminary. The findings therefore provide exploratory evidence for a candidate measurement structure rather than definitive validation of the refined instrument, and replication using a larger, independent sample is required.

Instrument Development

Data were collected using a structured, self-administered questionnaire developed in accordance with the study objectives, the Technology Acceptance Model (TAM), and relevant empirical literature concerning the use of generative artificial intelligence (GenAI) in education and academic writing. Instrument development began by identifying and operationalising five theoretically proposed domains: perceived usefulness (PU), perceived ease of use (PEU), ethical concerns (EC), academic confidence (AC), and students' use or perceived impact of AI-generated text in academic writing.

Items representing PU and PEU were adapted and contextualised from established TAM measures, particularly the conceptualisation of perceived usefulness and perceived ease of use proposed by Davis (1989), and were further informed by subsequent applications of TAM to the use of ChatGPT in academic writing (Zou & Huang, 2023). The PU items were contextualised to capture students' perceptions of the extent to which GenAI tools could enhance writing quality and organisation, improve efficiency, reduce the time required for writing and editing, support task completion, and provide relevant ideas or information. The PEU items addressed students' familiarity and prior experience with GenAI tools, perceived ease of interacting with and providing prompts to AI tools, and confidence in using such tools with limited external assistance.

The EC items were researcher-developed with reference to contemporary literature examining students' perceptions of generative AI and its ethical implications in higher education. In particular, the items were informed by Chan and Hu (2023), who highlighted concerns associated with originality, plagiarism, authorship, attribution and academic integrity in the use of generative AI. Accordingly, the EC items addressed students' concerns regarding academic honesty, originality of their work, similarity between AI-generated and existing content, potential plagiarism, and the need to verify AI-generated text before incorporating it into academic work.

The academic confidence (AC) items were researcher-developed with reference to established conceptualisations and measures of academic-writing self-efficacy (Mitchell et al., 2021; Teng & Wang, 2023). Academic-writing self-efficacy reflects students' beliefs in their capability to successfully undertake academic-writing tasks. In the present study, this concept was contextualised to the emerging use of generative AI by focusing specifically on students' perceived capability to undertake academic writing independently of AI assistance. Accordingly, the items assessed confidence in one's own writing skills, ability to complete academic assignments and produce quality work without AI-generated assistance, and perceptions of whether AI tools provide additional value beyond students' existing writing capabilities.

The eight items assessing students' use or perceived impact of AI-generated text were similarly researcher-developed with reference to contemporary research on the affordances and implications of generative AI for academic writing (Chan & Hu, 2023; Warschauer et al., 2023). These items captured several aspects of students' experiences with AI-generated text, including perceived improvements in writing quality and clarity, ability to meet assignment requirements, writing efficiency and reduced effort, attention to plagiarism and academic-integrity requirements, and perceived confidence and capability in completing academic-writing tasks.

The initial item pool was subsequently contextualised for university students at UiTM Puncak Alam. Item wording was simplified where necessary to improve clarity and ensure suitability for undergraduate respondents while retaining the intended conceptual meaning of each proposed domain. This process resulted in an initial pool of items that combined adapted and contextualised measures for the TAM-related constructs with researcher-developed items informed by relevant empirical literature for the remaining domains.

The questionnaire consisted of two parts. The demographic data and academic data, such as gender, age, semesters studied and faculty, were gathered in section A. The measurement items were included in Section B. The first measurement instrument had a total of 32 items, 6 of which were PU items, 6 were PEU items, 6 were EC items, 6 were AC items, and 8 were items related to students' use or perceived impact of AI-generated text in academic writing.

Each measurement item was expressed as a statement with a four-point Likert response format, from 1 (strongly

disagree) to 4 (strongly agree). The forced-choice format was chosen to get a direction of response to each statement, and the midpoint option was not included. Higher scores represented stronger endorsement of the attribute described by the respective item.

Importantly, the original five-domain, 32-item structure represented the a priori conceptual organisation of the questionnaire rather than an assumed final measurement model. The dimensionality of the items was subsequently evaluated empirically using the main-study data, allowing the measurement structure to be refined according to the observed factor pattern, item performance and theoretical interpretability.

Following initial item development, the questionnaire was reviewed by a member of the research team to assess the clarity of item wording, consistency with the intended constructs, and relevance to the context of AI-generated text in academic writing. Minor wording and contextual adjustments were made where necessary before the questionnaire was administered. No formal expert-panel content validation was undertaken at this stage. Consequently, the present findings should not be interpreted as establishing a comprehensive content validity for the refined instrument. Although the initial domains and items were informed by relevant theoretical and empirical literature, formal evaluation by an independent multidisciplinary expert panel remains necessary to assess item relevance, representativeness, clarity and coverage of the intended constructs. Such content validation should form part of the next stage of instrument development before stronger conclusions can be drawn regarding the psychometric validity of the proposed structure. Accordingly, the present study focused primarily on the empirical evaluation and refinement of the proposed instrument through exploratory factor analysis and internal consistency assessment.

Pilot Testing and Preliminary Reliability

Prior to the main data collection, the questionnaire was pilot-tested among 30 students who were not included in the main study sample. The pilot test was undertaken to evaluate the preliminary performance of the questionnaire and determine whether the proposed item sets demonstrated adequate internal consistency before administration in the main survey.

Pilot data were analysed using IBM SPSS Statistics version 29. Internal consistency was assessed using Cronbach's alpha (α), with coefficients of approximately .70 or higher considered indicative of acceptable internal consistency for preliminary scale evaluation.

The pilot analysis demonstrated satisfactory to high internal consistency across the five initially proposed domains. Cronbach's alpha coefficients were .926 for PU, .879 for PEU, .784 for EC, .947 for AC, and .919 for the perceived impact/use of AI-generated text domain. On the basis of these preliminary results, all 32 items were retained for administration in the main survey.

The pilot reliability findings were treated as preliminary evidence of internal consistency rather than confirmation of construct validity. The dimensional structure and reliability of the instrument were therefore subjected to more comprehensive psychometric evaluation using the main-study sample.

Data Collection Procedure

Data collection was conducted over approximately three months, from late October 2025 to late January 2026, among students enrolled at Universiti Teknologi MARA (UiTM) Cawangan Selangor, Kampus Puncak Alam. The questionnaire was administered to students who met the study eligibility criteria and were familiar with or had experience using GenAI tools. Participation in the study was voluntary, and respondents were informed of the purpose of the study before completing the questionnaire.

As the study employed purposive recruitment and participation was voluntary, the number of responses that would ultimately be obtained was not predetermined during questionnaire distribution. At the end of the data-collection period, 136 response records were initially available. During data screening, one completely blank record containing no valid responses to the study items was identified and excluded. Consequently, 135 valid responses were retained for subsequent analysis.

Prior to statistical analysis, the dataset was examined for completeness and response validity. The retained dataset contained no missing responses across the measurement items, and all responses were within the specified 1–4 Likert response range. The final analytical sample therefore comprised 135 respondents.

Data Analysis

Data analysis was mainly done in the computational environment of Google Colaboratory (Google Colab), and the analysis was performed using Python and R-based statistical software for data management, psychometric evaluation and inferential analysis. The data was analyzed using a sequential analytical procedure to screen the data, analyze the original questionnaire, improve the factor structure of the questionnaire, evaluate the reliability of the factors, and then examine the relationship between the factors.

The data set was screened for completeness, invalid responses and duplicate data before the psychometric evaluation. The responses for the 32 measurement items were reviewed for the allowable range for the four-point Likert scale (1=strongly disagree to 4=strongly agree). A total of 135 trailing records were valid for analysis, with one record being completely blank. The mean, standard deviation, median, range, skewness, kurtosis, and concentration of responses in each of the response categories were then analyzed to determine if there were any distributional irregularities or ceiling or floor effects. The suitability of the data for exploratory factor analysis (EFA) was further assessed using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity.

Instrument refinement was undertaken sequentially. An initial Pearson-correlation-based EFA of the original 32 items was conducted in Python using minimum residual (MINRES) extraction with oblimin rotation. This initial Pearson-based EFA was used solely as a preliminary diagnostic procedure to examine item loadings, cross-loadings, communalities, and potential item-performance problems; it did not determine factor retention or the final factor solution. D1 emerged as the clearest problematic item because it simultaneously demonstrated a weak primary loading, substantial cross-loading, and low communality, and was therefore provisionally removed. The primary factor-retention and subsequent instrument-refinement analyses were then conducted using polychoric correlations, which were considered more appropriate for the four-category ordinal response format.

The 31-item polychoric correlation matrix was estimated using the `polychoric()` function in the *psych* package in R. Because the resulting matrix was non-positive definite, it was smoothed using `cor.smooth()` before factor extraction. This condition was considered likely to reflect instability in the estimated polychoric correlations associated with the combination of a modest sample, substantial overlap among several items, and uneven response distributions. Matrix smoothing was therefore applied to obtain a mathematically admissible correlation matrix while retaining the ordinal-data approach for the primary analysis. Parallel analysis using polychoric correlations, MINRES extraction, and 100 random iterations was then conducted to determine the number of factors to retain. The parallel analysis supported a four-factor solution. Accordingly, a four-factor EFA was subsequently estimated from the smoothed polychoric correlation matrix using MINRES extraction with oblimin rotation.

Further refinement was undertaken iteratively, with statistical and conceptual criteria used to guide item removal rather than applying a single mechanical deletion rule, but not a single mechanical deletion rule. Factor loadings of .40 or greater were deemed desirable as primary factors, and communalities of .30 or greater were deemed desirable as item performance. Items with a difference of less than approximately .20 between their primary and strongest secondary factor loadings were examined for potential factorial ambiguity. Inter-item correlations and reliability diagnostics were also used to determine possible redundancies. These statistical data were considered along with the conceptual meaning of each item, the consistency of each item with the evolving factor interpretation, and the need for adequate content coverage of each construct. Items were thus deleted if they did not fit the emerging measurement structure in terms of their statistical performance and/or conceptual contribution, as this would reduce the clarity, distinctiveness, and parsimony of the measurement structure.

The iterative statistical and conceptual refinement process resulted in a parsimonious 21-item candidate instrument. Parallel analysis was re-run with the 21 retained items and again supported the retention of four

factors. This final polychoric EFA was performed using MINRES extraction with oblimin rotation, which resulted in four interpretable dimensions: AI Writing Utility and Usability (AWUU), Ethical Concerns (EC), Academic Writing Self-Reliance (AWSR), and AI Writing Integrity Practice (AIWIP).

A four-factor EFA based on Pearson correlations was then performed on the same 21 items, which were retained as a sensitivity analysis using MINRES extraction and oblimin rotation. This similarity in the placement of items and the interpretation of factors in the final polychoric and Pearson solutions was regarded as further evidence that the exploratory structure was not just the result of the type of correlation matrix used. However, as the same sample was used for both analyses, the Pearson-based analysis was not considered as a separate validation of the four-factor structure, but rather as a sensitivity analysis.

Ethical Considerations

This study and its consent procedure were reviewed and approved by the Faculty Ethics Review Committee (FERC), Universiti Teknologi MARA [Reference no. FPHP/FREC/810/2025].

RESULTS

Respondent Characteristics

There were 135 valid responses that were used in the final analysis. The female respondents were in the majority ($n = 82$, 60.7%), and there were 53 male respondents (39.3%) as detailed in Table 1. In terms of age, more than half of the respondents were aged 23–27 years ($n = 72$, 53.3%), followed by those aged 18–22 years ($n = 50$, 37.0%). Only a small proportion were aged 28–32 years ($n = 7$, 5.2%) or 33 years and above ($n = 6$, 4.4%).

Respondents included students in all eight semesters of study. The highest number of students was in the 5th semester ($n = 28$, 20.7%) and 6th semester ($n = 27$, 20.0%), followed by the 3rd semester ($n = 23$, 17.0%). The second and fourth semester students had 11.9% ($n = 16$) each, and 13 (9.6%) were in the first semester. Six respondents (4.4%) were from the 7th and 8th semesters.

The sample also included students from eight faculties, showing the involvement of students from various academic fields. The highest proportion of the respondents was from the Faculty of Hotel and Tourism Management ($n = 36$, 26.7%), followed by the Health Sciences ($n = 34$, 25.2%). The remaining respondents were from Education ($n = 17$, 12.6%), Pharmacy ($n = 15$, 11.1%), Business Management ($n = 11$, 8.1%), Built Environment ($n = 8$, 5.9%), Art and Design ($n = 8$, 5.9%), and Accounting ($n = 6$, 4.4%). In general, the sample included both students with varying levels of education and various fields of study, with the age group 23–27 years and the students in Hotel and Tourism Management and Health Sciences being more prevalent.

Table 1. Demographic characteristics of respondents ($N = 135$)

Characteristic	Category	n	%
Gender	Female	82	60.7
	Male	53	39.3
Age	18–22 years	50	37.0
	23–27 years	72	53.3
	28–32 years	7	5.2
	≥33 years	6	4.4
Semester of study	First	13	9.6
	Second	16	11.9
	Third	23	17.0
	Fourth	16	11.9
	Fifth	28	20.7
	Sixth	27	20.0
	Seventh	6	4.4
	Eighth	6	4.4



Faculty	Hotel & Tourism Management	36	26.7
	Health Sciences	34	25.2
	Education	17	12.6
	Pharmacy	15	11.1
	Business Management	11	8.1
	Built Environment	8	5.9
	Art & Design	8	5.9
	Accounting	6	4.4

Preliminary Data Screening and Factorability

Before exploratory factor analysis (EFA), the data were screened to ensure that they were suitable and of good quality for psychometric analysis. The original dataset included 136 rows, but the last row of data was blank and was not used. This led to an analytical sample of 135 respondents. Of the 32 original measurement items, none of the retained cases had any missing responses, and all responses were within the expected range of 1 to 4 for the four-point Likert scale. No duplicate records were found. Some respondents had the same response patterns but were not considered to be duplicate cases since their entire records were different.

Descriptive analysis of items revealed that the original items tended to have relatively high mean scores with adequate distributions of the scores across the response categories. The range of the item was 2.778 to 3.785, and the standard deviations varied from 0.509 to 1.034. The majority of items were negatively skewed, with a higher proportion of responses towards the agree end of the scale. The highest mean score (3.785) and the highest percentage (81.5%) of the respondents in the highest response category were found for D6. D6 also had a considerable negative skewness (-3.052) and positive kurtosis (11.922), suggesting a strong ceiling effect. However, the item was kept at this initial stage since distributional characteristics were not the only criterion used to delete items.

An inter-item correlation matrix was then reviewed to determine if there was too much redundancy. None of the items was an obvious statistical duplicate as no pair of items had an absolute Pearson correlation of .90 or higher. The highest correlations were observed between E4 and E5 ($r = .824$), E2 and E3 ($r = .817$), and B3 and B4 ($r = .812$). The determinant of the correlation matrix was, however, very small (3.49×10^{-16}), which indicated that there was considerable interrelationship among the original items. This was taken into account when considering the subsequent factor analyses, but was not used alone to remove items.

Bartlett's test of sphericity and Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy were also used to assess the suitability of the data set for factor analysis. Bartlett's test was statistically significant, $\chi^2 = 4360.081$, $p < .001$, which means that the correlation matrix was not an identity matrix, and there were enough correlations among the items for factor analysis. The overall KMO value was .719, which is acceptable for the sampling adequacy for EFA.

Because the questionnaire employed a four-point ordinal response scale, a polychoric correlation matrix was subsequently generated for the primary factor analysis. During its estimation, the initial matrix was identified as non-positive definite and was therefore smoothed computationally. The resulting 31×31 polychoric correlation matrix was used in the subsequent factor-retention procedure after the earlier removal of D1 during preliminary refinement. Parallel analysis conducted on this matrix consistently indicated the retention of four latent factors.

Taken together, the significant Bartlett's test, acceptable KMO coefficient, absence of extreme item-pair correlations, and factor-retention evidence supported proceeding with a four-factor EFA. However, the near-singularity of the original correlation matrix and the need to smooth the polychoric matrix were consistent with potential instability in the correlation structure, including substantial overlap among some of the original items. Consequently, subsequent item refinement considered not only factor loadings but also cross-loadings, communalities, conceptual redundancy, and theoretical interpretability. Table 2 summarises the preliminary assessment of the responses gathered.

Table 2. Preliminary assessment of data suitability for exploratory factor analysis

Assessment	Result	Interpretation
Valid sample	135	Cases retained for analysis
Original measurement items	32	Initial questionnaire
Missing item responses	0	No missing data among valid cases
Response range	1–4	All responses within valid Likert range
Exact duplicate records	0	No duplicate respondent records
Highest item-pair correlation	.824	Below .90 threshold
Determinant of correlation matrix	3.49×10^{-16}	Indicates substantial item intercorrelation
KMO	.719	Acceptable sampling adequacy
Bartlett's test	$\chi^2 = 4360.081, p < .001$	Correlation matrix suitable for factor analysis
Parallel analysis	4 factors	Four-factor solution supported

Exploratory Factor Analysis and Instrument Refinement

The dimensional structure of the questionnaire was examined using exploratory factor analysis (EFA). Although the original instrument was organised into five theoretically derived constructs, the empirical factor structure was not assumed a priori. Parallel analysis based on the polychoric correlation matrix consistently indicated that four factors should be retained. Accordingly, a four-factor EFA with oblique rotation was examined, allowing the factors to correlate.

The initial EFA revealed that the original five-construct specification was not fully reproduced. Specifically, the scales that were designed to measure perceived usefulness (PU) and perceived ease of use (PEU) exhibited significant empirical overlap and did not reveal two distinct factors. There were a number of items around the perceived impact of AI-generated text that also fit into this overall factor. In contrast, the responses to questions about ethical issues, academic writing self-reliance, and responsible use of AI showed more unique responses. Primary loadings, cross-loadings, communalities, conceptual redundancy, and substantive meaning of each item were therefore considered in refining the items in an iterative manner.

The preliminary analysis revealed the problem with D1: “I think using AI tools in writing could affect my honesty in completing assignments”. It had a primary loading of .389 and a secondary loading of .340, which resulted in a loading difference of .049 and a communality of .269. D1 was therefore removed because of its comparatively weak primary loading, substantial cross-loading, minimal separation between the primary and secondary loadings, and low communality.

Additional elaboration was directed towards the large factor, which included items from the original PU, PEU and perceived-impact domains. Several items shared very similar features of AI-assisted writing, and some had cross-loading and redundancy. The final retained items were chosen to reflect the range of the developing construct, but not to duplicate items in the list (Table 3). This process resulted in retention of B1, B3, B5, B6, C3, C4, C5, and C6. These were all considered together as perceived improvement in writing quality and efficiency, relevance of AI-generated content, ease of interacting with AI tools and confidence in using AI tools. Considering the above utility and usability characteristics, the factor was named as AI Writing Utility and Usability (AWUU). The results indicated that there was no empirical separation of perceived utility and perceived usability for the present sample and measurement context.

The moral aspect was also more clearly delineated after refinement. The items that were retained, D2, D3, D4, and D5, all had concerns about originality, possible concerns with using AI tools to create the work, similarity to other sources and accidental plagiarism of ideas and/or words. These items were grouped together as a factor, and the label “Ethical Concerns” (EC) was maintained. D6, on the other hand, did not group with these concern-oriented items. D6 (“I feel the need to double-check AI-generated text to make sure it is not plagiarized”), on the other hand, was aligned with F5 and F6, both of which were related to active checking of AI-generated text against academic-integrity expectations. This separation indicated that the dimensions of concern for ethical implications and behaviour designed to control ethical implications were empirically different.

The six E-items were all kept as a separate factor. These statements evaluated the confidence of students in creating academic content without AI, such as their confidence in their own writing, completing assignments without AI, and whether they felt that AI was necessary to achieve satisfactory academic results. The items were reworded as Academic Writing Self-Reliance (AWSR) as the items did not represent academic confidence in general but academic writing confidence in particular. The conceptual breadth of this dimension was also maintained in that all six items were retained.

The fourth factor consisted of D6, F5 and F6. The EC items were not included in these statements, which focused on what the student would do to check the accuracy of the AI-generated content, to reduce plagiarism or academic dishonesty, and to make sure the use of AI-generated text was aligned with academic-integrity expectations. Together, these behaviours are a significant aspect of responsible AI use in academic writing that is practice-oriented. Thus, the factor was termed AI Writing Integrity Practice (AIWIP). AIWIP is not meant to capture all aspects of responsible use of AI, but instead focuses on students' self-reported verification and academic-integrity practices with AI-generated text. This distinction was also substantively important as it distinguished students' concerns regarding the potential ethical implications from the actions they reported they took to deal with those concerns.

Following iterative refinement, the final EFA was conducted on the 21 retained items. The four-factor solution accounted for 71.1% of the variance based on the pattern-matrix sums of squared loadings. The first factor accounted for 27.6%, the second 22.3%, the third 11.8%, and the fourth 9.4%. Communalities ranged from .545 to .919, indicating that the retained items were adequately represented by the four-factor solution. The oblique solution also indicated moderate relationships between selected factors, with factor correlations ranging approximately from .115 to .465, supporting the decision to permit correlated factors.

Table 3. Final 21-item four-factor solution based on polychoric EFA.

Construct	Code	Item statement	Factor loading
AI Writing Utility and Usability (AWUU)	B1	I think the AI tool helps me write in a clearer and more organised way.	.775
	B3	Using the AI tool helps me complete writing tasks faster than doing them on my own.	.883
	B5	I think the content generated by the AI tool matches the requirements of my assignments or tasks.	.870
	B6	The AI tool provides ideas or information that directly support the topic I am working on.	.817
	C3	I feel comfortable exploring different features of the AI tool while working on assignments.	.813
	C4	I find the AI tool simple and easy to use when working on my writing.	.858
	C5	It is easy for me to give instructions or prompts to the AI tool to get useful results.	.848
	C6	I feel confident using the AI tool without needing much help from others.	.807
Ethical Concerns (EC)	D2	I am concerned that heavy reliance on AI might reduce the originality of my work.	.733
	D3	I believe teachers or lecturers might question my work if they know I used AI tools.	.900
	D4	I am worried that using AI tools could produce content that is similar to other sources.	.788
	D5	I think AI-generated text might accidentally copy existing ideas or wording.	.547
Academic Writing Self-Reliance (AWSR)	E1	I feel confident in my own writing skills without using AI tools.	.828



	E2	I can complete my assignments well even without AI assistance.	.874
	E3	I trust my ability to produce quality work without needing AI-generated ideas.	.859
	E4	I think AI tools are not needed for me to do well in my studies.	.873
	E5	I believe I can achieve the same results without using AI tools.	.851
	E6	I feel AI tools add no significant improvement to the work I already produce.	.836
AI Writing Integrity Practice (AIWIP)	D6	I feel the need to double-check AI-generated text to make sure it is not plagiarised.	.779
	F5	Using AI-generated text has made me more careful about avoiding plagiarism or academic dishonesty.	.633
	F6	I review AI-generated text to ensure it meets academic integrity standards.	.777

Note. AWUU = AI Writing Utility and Usability; EC = Ethical Concerns; AWSR = Academic Writing Self-Reliance; AIWIP = AI Writing Integrity Practice. Factor loadings are from the final 21-item polychoric four-factor EFA solution. The four-factor solution accounted for 71.1% of the total variance, with all retained items demonstrating salient primary loadings

F7 and F8 were subsequently examined because both demonstrated substantial cross-loading in the 31-item four-factor solution. F7 loaded .528 on its primary factor and .404 on a secondary factor, yielding a loading difference of only .124, while F8 loaded .509 and .410 on its two strongest factors, respectively, yielding an even smaller difference of .099. Although both items had acceptable communalities (.443 and .431, respectively), the small differences between their primary and secondary loadings indicated insufficient factorial distinctiveness. Conceptually, both items assessed confidence or capability arising from the use of AI-generated text, thereby combining elements of perceived AI benefit with students' writing capability rather than representing a clearly differentiated dimension. They were therefore removed on the basis of combined statistical and conceptual evidence.

The final solution therefore represented a substantial refinement of the original measurement model. A total of 21 of the original 32 items were retained, and 11 were excluded, resulting in four empirically interpretable dimensions: AWUU (8 items), EC (4 items), AWSR (6 items) and AIWIP (3 items) (Table 3). Importantly, it was not just a scaling down of the original five scales. Instead, the analysis revealed a reorganisation of the instrument, with the original PU and PEU domains not being empirically separated in this sample, and AI writing integrity practice as a distinct dimension from ethical concern. Removal of items was not necessarily due to poor factor loadings. Some items had acceptable primary loadings but were excluded as they did not add significantly new conceptual information compared to other items that were kept in the same emerging factor. To this end, the process of refinement involved not only statistical performance but conceptual contribution as well, and an attempt to maintain sufficient construct coverage and a more parsimonious measurement structure.

Table 4. Items Removed During Exploratory Factor Analysis and Instrument Refinement

Item	Original construct	Brief item content	Statistical issue	Conceptual consideration	Decision
B2	PU	Improved writing quality/style	Primary loading acceptable (.638), but lowest corrected item-total correlation in the 16-item factor (.493); α remained essentially	Broad writing-quality benefit overlapped with benefit-oriented content retained within AWUU	Removed to improve parsimony while preserving construct coverage

			unchanged if deleted (.951)		
B4	PU	Reduced editing/correction time	Strong primary loading (.743), but very high correlation with B3 ($r = .812$)	Substantial overlap with B3 in representing efficiency/time-saving benefits	Removed because of statistical and conceptual redundancy
C1	PEU	Familiarity with how AI tools work	Strong primary loading (.721), but comparatively lower corrected item-total correlation (.548); α remained .950 if deleted	Familiarity reflects prior knowledge/exposure more than perceived usability during academic writing	Removed to improve conceptual focus and parsimony
C2	PEU	Previous AI use for academic/personal tasks	Strong primary loading (.802) and CITC (.705); correlations with C4 ($r = .668$) and B1 ($r = .664$) indicated some overlap	Previous AI use primarily reflects exposure/experience rather than perceived utility or usability itself	Removed primarily to sharpen conceptual interpretation of AWUU
D1	EC	AI use may affect honesty in assignments	Weak primary loading (.389), secondary loading (.340), small loading difference ($\Delta = .049$), and low communality (.269)	Broad wording concerning honesty may have permitted multiple interpretations	Removed because of weak and factorially ambiguous performance
F1	Perceived impact/use	Improved writing quality/clarity	Acceptable primary loading (.595), but highly correlated with F2 ($r = .760$) and C4 ($r = .719$)	Writing-improvement benefit overlapped with content already represented within AWUU	Removed to reduce redundancy
F2	Perceived impact/use	Help in meeting assignment requirements	Strong primary loading (.679), but substantial correlations with C4 ($r = .771$), F3 ($r = .762$), F1 ($r = .760$), and B3 ($r = .745$)	Practical/task-related benefits were already represented within AWUU	Removed because of substantial statistical and conceptual redundancy
F3	Perceived impact/use	Quicker/more efficient writing	Strong primary loading (.693), but highly correlated with F2 ($r = .762$) and B6 ($r = .704$)	Efficiency benefits overlapped with retained AWUU content	Removed to reduce redundancy and improve parsimony
F4	Perceived impact/use	Reduced effort in completing assignments	Acceptable but comparatively weaker primary loading (.520) and CITC (.588); correlated with C3 ($r = .667$)	Reduced effort overlapped with ease/efficiency content and contributed limited distinctive content	Removed because of comparatively weaker contribution and conceptual overlap

F7	Perceived impact/use	Increased confidence in academic writing	Cross-loading between the utility/usability and confidence-related dimensions, with insufficient separation between competing loadings	Combined perceived AI benefit with AI-related writing confidence, producing ambiguous construct membership	Removed because of factorial and conceptual ambiguity
F8	Perceived impact/use	Increased capability to produce high-quality work	Cross-loading between the utility/usability and confidence-related dimensions, with insufficient separation between competing loadings	Combined perceived AI benefit with perceived writing capability, producing ambiguous construct membership	Removed because of factorial and conceptual ambiguity

Note. AWUU = AI Writing Utility and Usability; EC = Ethical Concerns; AWSR = Academic Writing Self-Reliance; AIWIP = AI Writing Integrity Practice. Statistical information reported for excluded items was derived from the relevant preliminary and intermediate analyses conducted during the iterative refinement process.

Final Four-Factor Structure and Reliability

Following the iterative EFA procedure, the refined instrument comprised 21 items distributed across four empirically derived constructs: AI Writing Utility and Usability (AWUU), Ethical Concerns (EC), Academic Writing Self-Reliance (AWSR), and AI Writing Integrity Practice (AIWIP). The retained items demonstrated satisfactory primary factor loadings and communalities in the final polychoric EFA, supporting the interpretability of the four-factor structure.

To examine whether the factor structure was dependent on the use of polychoric correlations, an additional EFA based on Pearson correlations was conducted as a sensitivity analysis. The Pearson-based analysis reproduced the same four-factor configuration. AWUU items loaded between .690 and .833, AWSR items between .791 and .878, EC items between .543 and .816, and AIWIP items between .593 and .824. None of the 21 retained items exhibited a primary loading below .40, a loading difference of less than .20 between the primary and secondary factors, or a communality below .30 (Table 5). The correspondence between the polychoric and Pearson solutions provided supplementary evidence of the consistency of the exploratory four-factor configuration across alternative correlation specifications.

Table 5. Sensitivity assessment of the final factor structure using Pearson EFA

Construct	No. of items	Loading range	Communality range
AWUU	8	.690–.833	.572–.711
EC	4	.543–.816	.334–.691
AWSR	6	.791–.878	.643–.788
AIWIP	3	.593–.824	.406–.690

As summarised in Table 5, the retained items demonstrated satisfactory primary factor loadings across the four constructs. Factor loadings ranged from .543 to .878, while communalities ranged from .334 to .788. AWSR showed the strongest overall loading range (.791–.878), whereas the lowest retained loading was observed within EC (.543). Collectively, these results supported retention of the 21 items within the final four-factor exploratory solution

Internal consistency was subsequently evaluated using Cronbach's alpha and McDonald's omega total. As shown in Table 6, all four constructs demonstrated satisfactory internal consistency. Cronbach's alpha coefficients

ranged from .822 to .934, with the highest coefficient observed for AWSR ($\alpha = .934$), followed by AWUU ($\alpha = .929$), EC ($\alpha = .825$), and AIWIP ($\alpha = .822$).

McDonald's omega total provided further support for score reliability, with coefficients ranging from .894 to .964. AWSR again demonstrated the highest reliability ($\omega_t = .964$), followed by AWUU ($\omega_t = .960$), AIWIP ($\omega_t = .917$), and EC ($\omega_t = .894$). The consistency of the alpha and omega results indicated satisfactory internal coherence across all four refined dimensions.

Mean inter-item correlations were .628 for AWUU, .538 for EC, .722 for AWSR, and .621 for AIWIP. The relatively strong inter-item associations were consistent with the high internal-consistency coefficients. Nevertheless, the particularly high mean inter-item correlation for AWSR (.722) suggests substantial similarity among its six items and should be considered when further validating or shortening the instrument in an independent sample.

Table 6. Internal consistency reliability of the refined 21-item instrument

Construct	Items	k	Cronbach's α	McDonald's ω_t	Mean inter-item r
AI Writing Utility and Usability (AWUU)	B1, B3, B5, B6, C3, C4, C5, C6	8	.929	.960	.628
Ethical Concerns (EC)	D2, D3, D4, D5	4	.825	.894	.538
Academic Writing Self-Reliance (AWSR)	E1, E2, E3, E4, E5, E6	6	.934	.964	.722
AI Writing Integrity Practice (AIWIP)	D6, F5, F6	3	.822	.917	.621

Note. k = number of items; α = Cronbach's alpha; ω_t = McDonald's omega total.

Construct-Level Descriptive Statistics and Correlations

Following establishment of the refined measurement structure, composite scores were calculated by averaging the retained items within each construct. Because all items were measured on the same four-point response scale, the resulting scores remained interpretable within the original range of 1 to 4.

As shown in Table 7, respondents' scores were concentrated towards the agreement end of the scales across the four dimensions. AIWIP recorded the highest mean ($M = 3.637$, $SD = 0.533$), indicating generally strong self-reported engagement in practices intended to ensure responsible use of AI-generated text. AWUU also recorded a relatively high mean ($M = 3.449$, $SD = 0.578$), suggesting generally favourable perceptions of the utility and usability of AI tools for academic writing. Ethical concerns were similarly high ($M = 3.465$, $SD = 0.601$). AWSR showed the lowest mean of the four constructs ($M = 3.068$, $SD = 0.761$), although the average remained above the midpoint of the response scale.

The distribution of AIWIP was notably negatively skewed (skewness = -2.373) and leptokurtic (kurtosis = 8.090), reflecting the concentration of respondents toward the upper end of the response scale. AWUU and EC also exhibited negative skewness, whereas AWSR showed a comparatively more symmetrical distribution (skewness = -0.161). These distributional characteristics were considered when conducting the subsequent regression analyses.

Table 7. Descriptive statistics for the four refined constructs

Construct	M	SD	Median	Min	Max	Skewness	Kurtosis
AWUU	3.449	0.578	3.50	1.00	4.00	-1.337	3.089
EC	3.465	0.601	3.50	1.25	4.00	-1.206	1.619
AWSR	3.068	0.761	3.00	1.50	4.00	-0.161	-1.224
AIWIP	3.637	0.533	4.00	1.00	4.00	-2.373	8.090

Bivariate correlation analysis showed that there were a number of different relationships between constructs (Table 8). The highest correlation was observed between AWUU and AIWIP ($r = .563, p < .001$), indicating that students who perceived greater utility and usability of AI for academic writing also reported stronger AI writing integrity practices.

EC was moderately and positively related with AWSR ($r = .489, p < .001$) and was slightly but positively related with AIWIP ($r = .230, p = .007$). AWSR was also found to be weakly but significantly related to AIWIP ($r = .203, p = .018$). By contrast, AWUU was not significantly associated with either EC ($r = .162, p = .061$) or AWSR ($r = .120, p = .167$).

Table 8. Correlations among the refined constructs

Construct	1	2	3	4
1. AWUU	1			
2. EC	.162	1		
3. AWSR	.120	.489***	1	
4. AIWIP	.563***	.230**	.203*	1

Note. $N = 135$. $p < .05^*$, $p < .01^{**}$, $p < .001^{***}$.

Factors Associated with AI Writing Integrity Practice

Due to the bounded four-point composite outcome and the strong ceiling effects found for AIWIP, the regression analysis was used in an exploratory fashion, and the focus was on the direction, magnitude, and strength of the association between constructs, not causal or predictive inferences. Multiple regression analysis was conducted to examine the independent associations of AWUU, EC, and AWSR with AIWIP. All three predictors were entered simultaneously into the model.

The overall regression model was statistically significant, $F(3, 131) = 21.59, p < .001$, explaining 34.3% of the variance in AIWIP ($R^2 = .343$; adjusted $R^2 = .328$). No problematic multicollinearity was detected, with VIF values ranging from 1.029 to 1.333.

As shown in Table 9, AWUU was the only construct significantly and independently associated with AIWIP in the primary model ($B = .495, \beta = .536$, robust $SE = .163, t = 3.025, p = .003, 95\% CI [.171, .818]$). Thus, higher perceived utility and usability of AI for academic writing was associated with stronger self-reported AI writing integrity practice after controlling for EC and AWSR. EC was positively but not significantly associated with AIWIP ($B = .087, \beta = .098, p = .236$), while AWSR was also non-significant ($B = .064, \beta = .091, p = .436$). Therefore, although EC and AWSR were associated with AIWIP at the bivariate level, neither explained significant unique variance after simultaneous adjustment for the other predictors.

Table 9. Multiple regression predicting responsible AI writing practice

Predictor	B	Robust SE	β	t	p	95% CI	VIF
Constant	1.434	.387	-	3.708	<.001	[.669, 2.199]	-
AWUU	.495	.163	.536	3.025	.003	[.171, .818]	1.029
EC	.087	.073	.098	1.191	.236	[-.058, .232]	1.333
AWSR	.064	.082	.091	0.781	.436	[-.098, .225]	1.316

Note. B = unstandardised regression coefficient; β = standardised regression coefficient; robust SE = HC3 heteroscedasticity-consistent standard error; CI = confidence interval; VIF = variance inflation factor

Robustness and Sensitivity Analyses

The stability of the regression estimates was further examined using 5,000 non-parametric bootstrap resamples. The results of the bootstrap analysis corroborated the main results. AWUU was also positively associated with AIWIP with a mean bootstrap coefficient of .470 ($SE = .145$) and a 95% bootstrap confidence interval that did

not contain 0 [.150, .708]. In contrast, the bootstrap confidence intervals included zero for both EC ($B = .095$, 95% CI $[-.044, .248]$) and AWSR ($B = .070$, 95% CI $[-.071, .228]$). The bootstrap analysis therefore provided further support for AWUU as the only construct showing a statistically reliable independent association with AIWIP in the full sample.

Seven observations had Cook's distance values above the screening value of $4/n$, and four observations had absolute studentised residuals > 3 . These cases were included in the main analysis as no evidence of invalid, duplicated or incorrectly entered responses was found in the data screening. However, a sensitivity analysis was performed with the four observations that have the most extreme studentised residuals omitted.

The reduced-sample model ($N = 131$) remained statistically significant and explained 32.8% of the variance in AIWIP ($R^2 = .328$; adjusted $R^2 = .312$). AWUU was still significantly and positively associated with AIWIP, as can be seen in Table 10 ($B = .302$, $p < .001$), while AWSR was still not significant ($B = .038$, $p = .376$). However, EC was significant ($B = .164$, $p = .020$) in the secondary model, whereas it was not in the primary model ($B = .087$, $p = .236$). Therefore, the independent association between EC and AIWIP was sensitive to extreme observations and should be interpreted with caution. In the primary, bootstrap and sensitivity analyses, AWUU was the most consistently independently related to AIWIP.

Table 10. Comparison of primary and sensitivity regression analyses

Predictor	Primary B	Primary p	Sensitivity B	Sensitivity p	Interpretation
AWUU	.495	.003	.302	<.001	Stable positive association
EC	.087	.236	.164	.020	Sensitive to extreme observations
AWSR	.064	.436	.038	.376	Consistently non-significant

Note. Primary analysis: $N = 135$ with HC3 robust standard errors. Sensitivity analysis: $N = 131$ following exclusion of four observations with the most extreme studentised residuals.

Overall interpretation of the regression findings

Taken together, the primary regression, bootstrap analysis, and sensitivity analysis provide consistent evidence for AWUU as a positive independent correlate of AIWIP. In contrast, the evidence for EC was less robust. EC was not significantly associated with AIWIP in the primary full-sample model, and its bootstrap confidence interval included zero, but it became statistically significant after four observations with extreme studentised residuals were excluded. The independent association between EC and AIWIP should therefore be regarded as sensitive to influential observations rather than as a robust finding. AWSR showed no significant independent association with AIWIP in either the primary or sensitivity model.

DISCUSSION

The results show that the students' perception of the AI-generated texts was not organized as expected by the original five-domain questionnaire. The analysis, however, provided evidence for a four-factor solution with AI Writing Utility and Usability (AWUU), Ethical Concerns (EC), Academic Writing Self-Reliance (AWSR) and AI Writing Integrity Practice (AIWIP). It is interesting to note that although the five theoretically proposed domains showed acceptable to high internal consistency in the pilot testing, the EFA did not replicate the original five domain structure. This finding illustrates that satisfactory internal consistency alone does not establish the dimensional validity of a measurement instrument. Rather, the EFA provided evidence that some items originally conceptualised as representing distinct domains were empirically more closely related, resulting in the emergence of a four-factor structure.

Most notably, items originally developed to represent perceived usefulness and perceived ease of use converged within AWUU, suggesting that students in the present sample may experience the practical value and usability of generative AI as an integrated judgement when engaging with academic-writing tasks. This empirical convergence should not be interpreted as indicating that perceived usefulness and perceived ease of use are theoretically equivalent or indistinguishable within TAM. Rather, the adapted PU and PEU items were not sufficiently differentiated in the present sample and academic-writing context to form separate factors. The

finding may reflect the conversational and task-oriented nature of generative AI, whereby students' perceptions of how easily they can interact with the technology are closely intertwined with their judgements of whether it helps them accomplish academic-writing tasks efficiently and effectively. Previous TAM-informed research similarly identifies usefulness and ease of use as central to GenAI adoption, while suggesting that ease of use may operate partly through perceived usefulness in task-oriented contexts (Duong et al., 2023; Liu & Chaojun, 2023; Shahzad et al., 2024; Susanto et al., 2026).

The refinement of the original perceived impact/use domain also suggests that this initially proposed domain may have encompassed conceptually distinct aspects of students' experiences with AI-generated text. The original items addressed perceived improvements in writing quality and efficiency, reduced effort, confidence and capability, as well as attention to plagiarism and academic integrity. These were not identified as a single dimension in the EFA. Some items, however, regarding perceived writing benefits were part of the general AWUU dimension, and some items regarding plagiarism and academic integrity were part of AIWIP. Other items were deleted due to cross-loading and/or conceptual duplication, such as items related to writing confidence and capability with AI. The results indicate that when evaluating students' interactions with AI-generated writing, perceived benefits, AI-related writing confidence, and integrity-oriented practices are not necessarily considered as indicators of a single underlying construct.

The rise of the new dimensions EC, AWSR, and AIWIP also suggests that students' interactions with AI-generated text are more than just about technology acceptance. EC was not part of general assessments of the usefulness of AI, as students can appreciate the benefits of using GenAI for efficiency, drafting, and language support, while at the same time expressing concerns about plagiarism, misinformation, authorship, dependence, and reduced critical engagement (Acosta-Enriquez et al., 2024; Cotton et al., 2023; Nelson et al., 2025; Vetter et al., 2024). Thus, the concept of ethical concern is not necessarily the inverse of acceptance of AI; students can see the potential advantages of AI, but remain aware of the potential risks.

AIWIP was also empirically distinct from EC, highlighting an important distinction between recognising ethical issues and reporting responsible practices. In the present instrument, AIWIP primarily reflected verification of AI-generated material and adherence to academic-integrity expectations. Its emergence as a separate factor is consistent with the view that responsible AI use involves behavioural and procedural practices rather than ethical attitudes alone. Students may recognise concerns surrounding plagiarism, originality, or authorship without necessarily translating those concerns into verification, revision, disclosure, or other responsible writing behaviours (Alshehri et al., 2026; Oravec, 2023; Vetter et al., 2024).

AWSR, meanwhile, captured students' confidence in their capacity to undertake academic writing without relying on AI. Although the construct demonstrated strong internal consistency, its high mean inter-item correlation suggests that some retained items may capture closely overlapping aspects of writing independence. The construct may therefore benefit from further refinement to encompass more differentiated aspects of self-reliance, such as independent idea generation, control over revision, critical decision-making, and maintenance of authorial voice. This distinction is relevant because AI-assisted writing may involve negotiation between technological support and authorship rather than a simple contrast between dependence and independence (Wang, 2024).

The consistency of the four-factor configuration in the Pearson-based sensitivity EFA provides supplementary evidence of the consistency of the exploratory four-factor structure across alternative correlation specifications. However, this finding should not be interpreted as independent validation because both analyses were conducted using the same sample. Further validation of the structure through CFA with an independent and more diverse sample is needed, especially given the dynamic and ever-changing nature of GenAI tools, institutional policies and student practices (Acosta-Enriquez et al., 2024; Parameswara et al., 2026).

Students' AWUU, EC, AIWIP and AWSR were all relatively high at the construct level. This indicates that positive attitudes towards AI, knowledge about AI's ethical aspects and self-reported responsible AI practices can coexist. Students might view GenAI as a helpful writing tool, and at the same time, be aware of the potential dangers of using GenAI and trying to adhere to academic integrity standards. This mixed attitude has also been found in the latest studies, which show that students appreciate the use of AI for brainstorming, language support,

drafting, and revision, but are worried about over-reliance, academic dishonesty, and loss of their own voice (AbdAlgane et al., 2025; Nelson et al., 2025; Wang, 2024).

However, AIWIP scores are high and must be interpreted with caution. The construct was skewed towards the ceiling ($M = 3.637$, skewness = -2.373 , kurtosis = 8.090), and this may be partly due to the academically desirable nature of the three items that were retained. Respondents are easily able to agree with behaviors like checking the AI-generated content and preventing plagiarism, potentially leading to a social-desirability bias and limited discrimination among students who moderately or highly practice responsible behaviors. Therefore, the findings should not be viewed as conclusive evidence that students have consistently responsible behavior in the use of GenAI. Future development of AIWIP should include more complex and challenging behaviours and maintain the construct's coherence. The strong ceiling effect should also be taken into account when interpreting the results of the regression analyses, since this limited range of AIWIP may have affected the construct-level associations.

The results of the regression are an added substantive insight from the use of the instrument that has been refined. In the primary model, AWUU was the only independent correlate of AIWIP that was significant, and this did not change in the bootstrap or sensitivity analyses. Similarly, students who felt that AI was helpful and controllable for academic writing also tended to have higher scores on the verification- and academic-integrity-oriented practices after controlling for EC and AWSR. This finding challenges a simplistic assumption that favourable perceptions of GenAI necessarily correspond with less responsible academic-writing practices. It could be argued that knowing about AI could help students to be more critical of AI-generated content and establish routines for checking, revision and incorporation of AI-generated content. Some evidence to support this interpretation comes from studies that correlate AI literacy, SRL, and higher levels of AI engagement (Alshehri et al., 2026; Jin et al., 2024; Kim et al., 2025; Shi et al., 2025). But, AWUU does not equal AI literacy, and the cross-sectional results do not mean that awareness of the usefulness of AI leads to more responsible behaviour. The results of EC were more complex. EC was also bivariate positively associated with AIWIP, but in the primary regression, when AWUU and AWSR were both included, the association of EC with AIWIP was not significant. However, EC was only important in the sensitivity analysis, with the four observations having extreme studentised residuals omitted. Furthermore, there was no evidence that EC was different from zero in the bootstrap confidence interval. Therefore, it is considered to be sensitive to influential observations and not robust in association with AIWIP. This pattern nonetheless helps to perpetuate the conceptual divide between ethical awareness and responsible practice. While awareness of the importance of recognising concerns around originality, plagiarism, authorship and inappropriate use of AI may be important, it does not guarantee that students will consistently translate this awareness into responsible writing behaviours. Ethical awareness can also co-exist with other factors, such as AI literacy, institutional guidance, transparency, and opportunities for the practice of responsible use (Acosta-Enriquez et al., 2024; Shishakly & Nachouki, 2026).

Although AWSR was weakly correlated with AIWIP at the bivariate level, it did not demonstrate a significant independent association with AIWIP after adjustment for AWUU and EC. This suggests that greater confidence in writing without AI does not necessarily correspond with stronger integrity-oriented AI-writing practices once the other dimensions are considered. Students can have good independent writing skills and opt to use GenAI strategically for planning, feedback, language refinement or revision. However, students who are not as self-reliant in their writing can still use AI responsibly, such as by checking the results, understanding academic-integrity guidelines, and setting clear lines of authorship. The use of AI tools in writing effectively, therefore, may not be so much about avoiding the use of AI but more about maintaining control and agency, using critical thinking, and deciding when and how to use AI in writing (Hadi et al., 2026; Hongxia & Razali, 2026; Wang, 2024).

The implications of these findings for teaching and learning are discussed. Responses to GenAI by institutions must go beyond a dichotomy of ban and free use and develop critical and responsible AI literacy. In addition to addressing the use of AI, academic-writing instruction needs to highlight how students use AI. This can involve intentional prompting, critical reading of produced work, checking and verifying facts and sources, substantial editing, academic-integrity standards, disclosure of AI support and retention of students' intellectual work. The difference between EC and AIWIP also implies that ethical instruction must go hand-in-hand with opportunities to apply the ethical principles to real-world writing tasks, such as comparing AI-generated text with

independently sourced evidence, explaining how AI was used to create the final product, and evaluating AI-generated text. These activities can help shift the focus of responsible AI education from an understanding of the ethical principles to their implementation in academic writing practice.

Meanwhile, the absence of a significant independent association between AWSR and AIWIP suggests that responsible AI education should not be framed as a binary choice between using AI and maintaining self-reliance. The educational goal should be to help students learn to identify when they need AI support, when they need to check and/or make significant edits to AI-generated content, and when they should write on their own. In this way, the integration of responsible GenAI is more aptly thought of as a conscious, self-regulated writing process where technological support is still considered to be under the control of human judgment, authorship and academic responsibility.

In terms of measurement, the results suggest that the four-factor structure should be further investigated and not taken as conclusive. Future research is needed to evaluate the proposed measurement structure using CFA in larger, independent and more diverse samples, to expand AIWIP to include more than just the three items focused on verification and academic integrity, and to modify AWSR to minimize possible duplication while maintaining conceptual breadth. Additionally, longitudinal and behavioural studies could be conducted to determine if there is a correlation between the self-reported responsible behaviors and the actual observed use of GenAI in students' academic writing activities. Overall, the results indicate that students' interaction with AI-generated text is multifaceted, ranging from practical evaluation and ethical considerations to autonomous writing and self-reported integrity-oriented practices, and that responsible use can be as much a product of informed and critical interaction with AI as of being aware of the potential risks.

CONCLUSION

This study offers preliminary empirical evidence for a candidate instrument to assess university students' perceptions and practices with regard to AI text in academic writing. The original 5-domain questionnaire was empirically evaluated, resulting in a 21-item questionnaire with 4 factors: AI Writing Utility and Usability (AWUU), Ethical Concerns (EC), Academic Writing Self-Reliance (AWSR), and AI Writing Integrity Practice (AIWIP), with satisfactory internal consistency across the refined factors. Results of the exploratory analyses using both polychoric and Pearson correlation matrices showed the same four-factor configuration, which provided additional evidence that the four-factor configuration was not a result of the type of correlation matrix used, but further independent confirmation is required. In addition to instrument refinement, AWUU showed the strongest positive association with AIWIP and remained independently associated with AIWIP across the primary, bootstrap and sensitivity analyses. In contrast, AWSR showed no significant independent association in the adjusted models, while the independent association between EC and AIWIP was sensitive to influential observations. These results suggest that the perception of usefulness and usability of AI can co-occur with responsible academic-writing practices and does not necessarily imply inappropriate use.

The results also indicate that GenAI should be addressed pedagogically rather than just restricted or generally made aware of its ethical implications. Academic-writing instruction should foster practical and critical AI literacy to enable students to assess and validate AI-generated content, engage in substantive revisions of AI-generated material, uphold academic integrity, acknowledge AI support if necessary, and ensure their own intellectual contribution. This separation between ethical concerns and reported responsible practice further points to the need for not only awareness of ethical concerns, but also opportunities to practice responsible use of AI in real academic-writing tasks.

Several limitations should be considered when interpreting the findings. First, the study involved a relatively modest purposive sample of 135 students from a single UiTM campus. Although the data demonstrated acceptable factorability and yielded an interpretable four-factor solution, the sample limits the stability and generalisability of the exploratory structure. The need to smooth the initially non-positive-definite polychoric correlation matrix provides an additional reason for caution, notwithstanding the reproduction of the same broad factor configuration in the Pearson-correlation sensitivity analysis. As both EFA approaches were conducted using the same dataset, the latter should be regarded as a sensitivity analysis rather than independent validation. Second, formal content validation by an independent expert panel was not undertaken; therefore, the present

findings provide evidence concerning empirical structure and internal consistency rather than comprehensive evidence of content validity. Third, AIWIP comprised only three self-reported items focused primarily on verification and academic-integrity practices and exhibited a pronounced ceiling tendency, potentially limiting its conceptual breadth and ability to discriminate among respondents. The high inter-item association within AWSR also suggests potential item redundancy. Finally, the forced-choice four-point response format may have contributed to the concentration of responses towards particular response categories.

Accordingly, the refined 21-item instrument should be regarded as a candidate measure requiring further psychometric validation rather than as a fully validated instrument. Future research should evaluate the proposed four-factor structure using confirmatory factor analysis in a larger, independent and more diverse sample. Subsequent validation should assess convergent and discriminant validity and examine whether instrument scores are associated with relevant external measures or behaviours (criterion-related validity). With sufficiently large and diverse samples, future studies should also evaluate measurement invariance to determine whether the instrument functions consistently across different student groups, such as academic disciplines and levels of study. Formal content validation involving experts in academic writing, educational measurement, academic integrity and generative AI is also recommended to assess item relevance, clarity and construct coverage. Particular attention should be given to broadening AIWIP to encompass aspects such as transparency, authorship, critical verification, appropriate attribution and policy-compliant AI use, while examining potential redundancy within AWSR. Future studies could also investigate alternative response formats and use longitudinal or behavioural measures to determine whether self-reported responsible practices correspond with students' actual use of GenAI during academic writing.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. Data sharing is subject to ethical and confidentiality constraints.

ACKNOWLEDGEMENT

The authors would also like to acknowledge all UiTM Kampus Puncak Alam students who participated in this study

Conflict Of Interests

The authors declared that they have no conflict of interest.

REFERENCES

1. AbdAlgane, M., Ali, R., Othman, K., Ibrahim, I. Z. A., Alhaj, M. K. M., Ali, E., & Allehyani, F. (2025). Exploring AI-Generated Texts vs. Human-Written Texts in EFL Academic Writing: A Case Study of Qassim University in Saudi Arabia. *World Journal of English Language*, 16(2), 114-131. <https://doi.org/10.5430/wjel.v16n2p114>
2. Acosta-Enriquez, B. G., Arbulú Ballesteros, M. A., Perez Vargas, C. A., Ulloa, M. N. O., Gutiérrez Ulloa, C. R., Pizarro Romero, J., Gutiérrez Jaramillo, N. D., Cuenca Orellana, H. U., Ayala Anzoátegui, D. X., & López Roca, C. (2024). Knowledge, attitudes, and perceived Ethics regarding the use of ChatGPT among generation Z university students. *International Journal for Educational Integrity*, 20, 1–23. <https://doi.org/10.1007/s40979-024-00157-4>
3. Alshehri, O., Lamouchi, A., Abdellatif, M., Al-Dosari, M., Mekheimer, M., & Nemet-Allah, M. A. (2026). From tool to scaffold: structured human–AI collaboration and its effects on academic writing and digital critical thinking among Saudi EFL learners. *Frontiers in Psychology*, 17, 830103. <https://doi.org/10.3389/fpsyg.2026.1830103>
4. Arlakov, E. A., & Miklyaeva, A. (2026). Questionnaire “Students’ attitudes towards the use of artificial intelligence technologies in educational activities”: development and psychometric characteristics.” *Psychological Science and Education*, 31(3), 5–20. <https://doi.org/10.17759/pse.2026310301>
5. Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: student and teacher perspectives on the use of

- generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20, 1–24. <https://doi.org/10.1186/s41239-023-00427-0>
6. Chan, C., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(43), . <https://doi.org/10.1186/s41239-023-00411-8>
 7. Chatto, M., & Logronio, F. (2025). Navigating AI in Academia: Undergraduate Experiences with ChatGPT and the Redefinition of Academic Writing. *Journal of Education and Learning Reviews*, 24(2), 27–42. <https://doi.org/10.60027/jelr.2025.2061>
 8. Cotton, D., Cotton, P., & Shipway, J. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61, 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
 9. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340. <https://doi.org/10.2307/249008>
 10. Duong, C. D., Vu, T., & Ngo, T. (2023). Applying a modified technology acceptance model to explain higher education students' usage of ChatGPT: A serial multiple mediation model with knowledge sharing as a moderator. *The International Journal of Management Education*, 21(3), 100883. <https://doi.org/10.1016/j.ijme.2023.100883>
 11. Hadi, A., Aribowo, K., & Prasetyo, D. E. (2026). Students' Self-Regulation in Learning Academic Writing with AI Mediation Tools. *Journal of English Language and Education*, 11(3), 355–365. <https://doi.org/10.31004/jele.v11i3.2437>
 12. Hongxia, H., & Razali, A. B. (2026). Impact of AI-Assisted Peer Feedback on EFL Students' Self-Regulated Learning, Self-Efficacy, and Motivation in English Academic Writing. *International Journal of Instruction*, 19(2), 163-182. <https://doi.org/10.29333/iji.2026.1929a>
 13. Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15(4), ep464. <https://doi.org/10.30935/cedtech/13605>
 14. Jang, Y., Choi, S., & Kim, H. (2022). Development and validation of an instrument to measure undergraduate students' attitudes toward the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of AI education. *Education and Information Technologies*, 27, 11635–11667. <https://doi.org/10.1007/s10639-022-11086-5>
 15. Jin, F., Lin, C.-H., & Lai, C. (2024). Modeling AI-assisted writing: How self-regulated learning influences writing outcomes. *Computers in Human Behavior*, 165, 108538. <https://doi.org/10.1016/j.chb.2024.108538>
 16. Johnston, H., Wells, R., Shanks, E., Boey, T., & Parsons, B. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity*, 20, 1–21. <https://doi.org/10.1007/s40979-024-00149-4>
 17. Khojasteh, L., Karimian, Z., Nasiri, E., Kafipour, R., & Farahmandi, A. Y. (2025). Artificial intelligence and academic writing questionnaire (AI-AWQ): development and validation among medical students' experiences using exploratory factor analysis. *BMC Medical Education*, 25(1), 1697. <https://doi.org/10.1186/s12909-025-08288-z>
 18. Kim, J., Lee, S.-S., Detrick, R., Wang, J., & Li, N. (2025). Students-Generative AI interaction patterns and its impact on academic writing. *Journal of Computing in Higher Education*, 38, 504–525. <https://doi.org/10.1007/s12528-025-09444-6>
 19. Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*, 30, 1265–1300. <https://doi.org/10.1007/s10639-024-12878-7>
 20. Liu, G., & Chaojun. (2023). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 18, 125–138. <https://doi.org/10.1080/17501229.2023.2240316>
 21. Liu, Y., Park, J., & McMinn, S. (2024). Using generative artificial intelligence/ChatGPT for academic communication: Students' perspectives. *International Journal of Applied Linguistics*, 34(4), 1437–1461. <https://doi.org/10.1111/ijal.12574>
 22. Marengo, A., Yilmaz, F., Yilmaz, R., & Ceylan, M. (2025). Development and validation of generative artificial intelligence attitude scale for students. *Frontiers in Computer Science*, 7, 1528455.

- <https://doi.org/10.3389/fcomp.2025.1528455>
23. Mitchell, K. M., McMillan, D. E., Lobchuk, M. M., Nickel, N. C., Rabbani, R., & Li, J. (2021). Development and validation of the situated academic writing self-efficacy scale (SAWSES). *Assessing Writing*, 48, 100524. <https://doi.org/10.1016/j.asw.2021.100524>
 24. Nelson, A., Santamaría, P., Javens, J., & Ricaurte, M. (2025). Students' Perceptions of Generative Artificial Intelligence (GenAI) Use in Academic Writing in English as a Foreign Language. *Education Sciences*, 15(5), 611. <https://doi.org/10.3390/educsci15050611>
 25. Oravec, J. A. (2023). Artificial Intelligence Implications for Academic Cheating: Expanding the Dimensions of Responsible Human-AI Collaboration with ChatGPT. *Journal of Interactive Learning Research*, 34(2), 213–237. <https://doi.org/10.70725/304731gmmvhw>
 26. Parameswara, D. A. D., Berlilana, B., & Saputro, R. E. (2026). Utilitarian vs Human-Centered AI Acceptance: Explaining Students' Adoption of ChatGPT in Higher Education. *Edumatic: Jurnal Pendidikan Informatika*, 10(1), 190–199. <https://doi.org/10.29408/edumatic.v10i1.34218>
 27. Perkins, M. (2023). Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2), 1–24. <https://doi.org/10.53761/1.20.02.07>
 28. Sanz-Tejeda, A., Domínguez-Oller, J. C., Baldaquí-Escandell, J.-M., Gómez-Díaz, R., & García-Rodríguez, A. (2026). The impact of generative AI on academic reading and writing: a synthesis of recent evidence (2023–2025). *Frontiers in Education*, 10, 1711718. <https://doi.org/10.3389/feduc.2025.1711718>
 29. Shahzad, M. F., Xu, S., & Javed, I. (2024). ChatGPT awareness, acceptance, and adoption in higher education: the role of trust as a cornerstone. *International Journal of Educational Technology in Higher Education*, 21, 1–23. <https://doi.org/10.1186/s41239-024-00478-x>
 30. Shi, J., Liu, W., & Hu, K. (2025). Exploring How AI Literacy and Self-Regulated Learning Relate to Student Writing Performance and Well-Being in Generative AI-Supported Higher Education. *Behavioral Sciences*, 15(5), 705. <https://doi.org/10.3390/bs15050705>
 31. Shishakly, R., & Nachouki, M. (2026). Academic Integrity and Students' Ethical Use of ChatGPT in Higher Education. *Journal of Information Technology Education: Research*, 25, 8. <https://doi.org/10.28945/5730>
 32. Susanto, H., Mutalif, A. Q., Setiana, D., & Besar, N. (2026). From usefulness to motivation: a TAM–SDT perspective on AI adoption and problematic internet use in higher education. *Frontiers in Education*, 10, 1699827. <https://doi.org/10.3389/feduc.2025.1699827>
 33. Teng, M. F., & Wang, C. (2023). Assessing academic writing self-efficacy belief and writing performance in a foreign language context. *Foreign Language Annals*, 56(1), 144–169. <https://doi.org/10.1111/flan.12638>
 34. Van Niekerk, J., Delpont, P., & Sutherland, I. (2025). Addressing the use of generative AI in academic writing. *Computers and Education: Artificial Intelligence*, 8, 100342. <https://doi.org/10.1016/j.caeai.2024.100342>
 35. Vetter, M., Lucia, B., Jiang, J., & Othman, M. (2024). Towards a framework for local interrogation of AI ethics: A case study on text generators, academic integrity, and composing with ChatGPT. *Computers and Composition*, 71, 102831. <https://doi.org/10.1016/j.compcom.2024.102831>
 36. Wang, C. (2024). Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. *Technology, Knowledge and Learning*, 30, 1825–1846. <https://doi.org/10.1007/s10758-024-09744-3>
 37. Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing*, 62, 101071. <https://doi.org/10.1016/j.jslw.2023.101071>
 38. Wise, B., Emerson, L., Van Luyn, A., Dyson, B., Bjork, C., & Thomas, S. (2024). A scholarly dialogue: writing scholarship, authorship, academic integrity and the challenges of AI. *Higher Education Research & Development*, 43, 578–590. <https://doi.org/10.1080/07294360.2023.2280195>
 39. Zou, M., & Huang, L. (2023). To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model. *Frontiers in Psychology*, 14, 1259531. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1259531>