Exploring Multimodal LLM Engagement on Prompt Engineering Competency Among Non-Tech Students in India

Meltem Alkoyak-Yildiz, Sandrapuri Aswin Tejasvi, Venkatesh Viswanathan, and Luis Felipe Hernandez Mora
Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India

meltemalkoyak@am.amrita.edu
tejashyd23@gmail.com
venkatesh14518@gmail.com
hm_luisfelipe@am.amrita.ed

Abstracts: Large Language Models (LLMs) have become increasingly prevalent in the daily use of university students' educational contexts, yet the specific perceptual LLM engagement patterns that contribute to effective LLM usage remain weakly understood. This exploratory cross-sectional study (N = 75) examines which factors of LLM engagement most strongly predict self-perceived prompt engineering competency among non-tech university students in India. Academic participants completed a survey measuring: Self-Perceived Prompt Engineering Competency (dependent variable), Basic Modality Use of LLM Input (text, images, audio, video, code), Strategic Integration & Application (purposeful combination of modalities), Depth of Use (usage of basic queries vs complex multi-modal tasks), and Self-Readiness & Access to AI tools, Primary Adoption Barrier to use AI. Non-parametric Wilcoxon signed-rank tests, multiple regression, and barrier analyses through Chi-square tests were conducted. Wilcoxon signed-rank tests determined whether scores significantly exceeded the theoretical scale midpoint of 3.0. Three of five constructs were significantly above the midpoint: Self-Readiness & Access ($W = 1323$, $p < .001$, $r = .782$, large), Prompt Engineering Competency ($W = 2197$, $p < .001$, $r = .768$, large), and Integration & Application ($W = 1008$, $p = .002$, $r = .463$, medium). Basic Modality Use ($W = 516$, $p = .997$) and Depth of Use ($W = 496$, $p = .963$) did not exceed the midpoint. These results indicate that despite high perceived competency, multimodal and complex LLM engagement remains limited. Results of the multiple regression analysis show that the four-predictor regression model explains 41.6% of the variance in competency ($F(4, 70) = 12.5$, $p < .001$). Integration & Application was the strongest unique predictor ($B = 0.51$, $p < .001$), followed by Self-Readiness & Access ($B = 0.34$, $p = .006$). Neither Basic Modality Use nor Depth of Use contributed unique variance. A Chi-square goodness-of-fit test revealed a significantly unequal barrier distribution ($\chi^2(5, N = 75) = 37.6$, $p < .001$, $w = 0.708$). Lack of knowledge was the most prevalent Adoption Barrier to use AI (37.3%) and scored significantly lower on competency than students reporting privacy concern ($U = 158$, $p = .034$, $r = .373$). This study demonstrates that Prompt Engineering Competency is driven specifically by the purposeful integration of LLMs into academic workflows, rather than frequent exposure or access alone. Implications for AI literacy curriculum design are further discussed.

Keywords*:* Prompt Engineering, Multimodal Interaction, LLM Competency, AI Literacy, Higher Education

1. Introduction

The rapid integration of Large Language Models (LLMs) into academic workflows has fundamentally altered the landscape of scholarly communication and productivity. Tools such as ChatGPT, Claude, Gemini, and Copilot have become embedded features of the academic toolkit, with a growing number of faculty, researchers, and students relying on them for tasks ranging from literature synthesis to data analysis and creative ideation. Yet, despite this widespread adoption, a critical gap persists between the availability of these tools and the ability of users to leverage them effectively (Idris et al., 2024; Zhou et al., 2026).

Prompt engineering is recognized as a learnable skill that requires explicit instruction and deliberate practice. Interestingly, even technically literate users frequently struggle to construct effective prompts, underscoring the need for empirical research into the factors shaping prompt competency in academic settings. Non-expert users often struggle to engage in systematic prompt engineering, instead exploring designs opportunistically and relying heavily on the model's ability to interpret ambiguous instructions. Their difficulties stem from both misconceptions about model capabilities and the inherent challenge of translating cognitive intent into effective execution (Zamfirescu-Pereira et al., 2023). This suggests that prompt engineering is not an intuitive ability but a structured competency, motivating a closer investigation into its underlying indicators.

The present study addresses this gap through an exploratory cross-sectional investigation of perceived multimodal LLM engagement among 75 academic participants. Rather than treating multimodal interaction as a unitary construct, we approach five dimensions: Self-Perceived Prompt Engineering Competency (dependent variable), Basic Modality Use of LLM input (text, images, audio, video, code), Strategic Integration & Application (purposeful combination of modalities), Depth of Use (usage of basic queries vs complex multi-modal tasks), and

Perceived Self-Readiness & Access to AI tools. Further we investigated the Primary reported Barrier in using AI and LLM.

The following research questions are investigated: To what extent do the levels of Perceived Self-Readiness & Access to AI Tools, Strategic Integration & Application (purposeful combination of modalities), and Self-Perceived Prompt Engineering Competency among non-technical university students significantly deviate from a neutral baseline? It is hypothesized that students' levels of Self-Readiness & Access, purposeful Integration & Application, and Self-Perceived Prompt Engineering Competency will significantly exceed the neutral baseline midpoint.

Further, we investigate whether, to what extent do LLM engagement factors—specifically Basic Modality Use, Depth of Use, Integration & Application, and Self-Readiness & Access—collectively predict self-perceived Prompt Engineering Competency among non-technical university students? It is hypothesized that the combination of LLM engagement factors (Basic Modality Use, Depth of Use, Integration & Application, and Self-Readiness & Access) significantly predicts self-perceived prompt engineering competency.

And finally, what are the primary barriers to LLM adoption reported by non-tech university students, and how do these specific barriers differ in their impact on students' perceived prompt engineering competency?

2. Methodology

2.1 Research Design

This study employed an exploratory cross-sectional survey design to examine the relationships between multimodal LLM engagement patterns and self-reported Prompt Engineering Competency among academic users. Data was collected at a single time point from 75 participants. A cross-sectional design was selected to establish baseline associations between constructs prior to longitudinal investigation.

2.2 Participants

Seventy-five academic participants were recruited for this study. All participants reported prior experience with AI tools (e.g., ChatGPT, Gemini, Copilot). Nearly half (46.7%) reported daily AI use, with an additional 14.7% using AI tools weekly. One-third (33.3%) reported occasional use, while 5.3% reported rare use. The majority of participants (58.7%) primarily used mobile devices for academic work, followed by laptop users (37.3%). Most participants (94.7%) had stable internet access, and 69.4% reported being comfortable or very comfortable with digital tools and platforms.

While the sample size is modest relative to large-scale survey research, it exceeds the minimum threshold recommended for multiple regression with four predictors. Stevens (1996) recommends a minimum of 15 participants per predictor, yielding a threshold of 60. The present sample of 75 satisfies this requirement.

An a priori power analysis using G*Power 3.1 indicated that a sample of 67 participants would be required to detect a medium-to-large effect size ($f^2 = 0.30$; Cohen, 1988). The obtained sample size of 75 exceeds this threshold and achieved an actual power of .95, confirming adequate statistical power for the primary regression analysis.

2.3 Instruments and Constructs

All items employed 5-point scales measuring either agreement or frequency, and the dependent variable is Perceived Prompt Engineering (see Table I).

Table 1: Construct Definitions and Measurement Properties

Construct	Items	Scale	α	Role
Perceived Prompt Engineering Competency	5	Agreement	.877	Dependent Variable
Basic Modality Use of LLM input (AI Input Diversity text, images, audio, video etc.)	5	Frequency	.733	Predictor
Integration & Application (purposeful combination of modalities)	3	Agreement	.815	Predictor
Depth of Use (usage of basic queries vs complex multi-modal tasks)	2	Frequency	.712	Predictor
Self-Readiness & Access to AI tools	2	Agreement	.652	Predictor
Primary Adoption Barrier	1	Nominal	—	Categorical

Perceived Prompt Engineering Competency served as the dependent variable and was operationalized as prompt engineering capability and output evaluation skill, grounded in approach of Zamfirescu-Pereira et al. (2023). Five items assessed the participant's ability to craft effective prompts, decompose complex tasks, request diverse output types, iteratively refine prompts, and critically evaluate AI-generated outputs. Internal consistency was good to very good ($\alpha = .877$).

Basic Modality Use of LLM Input measured the range of input modalities participants reported using when interacting with LLM tools. Five items captured the frequency of text, image, audio, video, and code/structured data input, drawing on Rogers' (2003) Diffusion of Innovation theory which claims that users gradually adopt new technology features. Reliability was acceptable ($\alpha = .733$).

Strategic Integration & Application captured higher-order capabilities involving the purposeful combination and application of multiple modalities. Three items assessed the participant's ability to combine input types effectively, interpret multimodal outputs, and apply multimodal interactions for specific academic tasks. Reliability shows a good value ($\alpha = .815$).

Depth of Use measured the progression from basic to complex task engagement using two items. The items distinguished between using multimodal features for simple queries versus combining multiple modalities for complex, integrated tasks. Reliability was acceptable ($\alpha = .712$).

Perceived Self-Readiness & Access to AI tools were operationalized as a two-item composite capturing both material access to multimodal-capable devices and perceived technical capability. The two items were significantly correlated ($\rho = .485$, $p < .05$), supporting their aggregation. The lower reliability coefficient ($\alpha = .652$) is attributable to the two-item scale length; the Spearman-Brown prophecy formula predicts attenuated reliability estimates for brief scales even when inter-item correlation is adequate (Field, 2013).

Primary Adoption Barrier was measured as a single-choice item with six unordered categories: lack of tools, lack of knowledge, lack of time, connectivity issues, privacy concerns, and none. This item was included as an exploratory measure to contextualize the engagement findings with practical adoption information and was analysed using categorical methods.

2.4 Data Screening and Assumptions

All continuous variables were screened for distributional properties. Shapiro-Wilk tests indicated statistically significant departures from normality for four of five constructs ($p < .05$), while Basic Modality use approached significance ($W = .968$, $p = .051$). Non-parametric methods were adopted across all constructs for analytical consistency. For regression analyses, which are relatively robust to moderate non-normality with sufficient sample sizes (Norman, 2010)

2.5 Factor Structure

An exploratory factor analysis using principal axis factoring with oblimin rotation and parallel analysis was conducted on all 17 items. Two factors were extracted. Factor 1 comprised the five Perceived Prompt Competency items (loadings: .654–.853), confirming the unidimensionality of the dependent variable. Factor 2 comprised the remaining 12 items spanning all four predictor constructs, with Integration & Application items

cross-loading on both factors. The two-factor solution suggests that while the dependent variable is factorially distinct, the four predictor constructs share substantial variance and form a single behavioural engagement dimension at the observed sample size. However, as the regression results demonstrate below, these constructs exhibit meaningfully different predictive relationships with the dependent variable, providing evidence of functional distinctiveness that the exploratory factor analysis, constrained by n = 75 and as few as two items per construct, lacked statistical power to detect. The four predictor constructs are therefore retained as theoretically derived dimensions, consistent with their grounding in distinct frameworks (Bloom, 1956; Bandura, 1977; Dreyfus and Dreyfus, 1980; Rogers, 2003; van Dijk, 2005).

2.6 Analytical Strategy

The analysis proceeded in two phases. Phase 1 addressed the primary research questions through one-sample Wilcoxon Signed-Rank tests (testing engagement levels against the scale midpoint of 3.0), Spearman rank-order correlations (establishing bivariate associations between predictors and the dependent variable), and simultaneous multiple linear regression. Phase 2 comprised exploratory analyses of adoption barrier distribution (chi-square goodness-of-fit) and barrier-competency relationships (Kruskal-Wallis and Mann-Whitney U tests). All analyses were conducted using Jamovi, with significance set at $\alpha = .05$. Effect sizes were interpreted using Cohen's (1988) conventions.

3. Results

3.1 Descriptive Statistics

Table 2: Descriptive Statistics for All Continuous Constructs

Construct	Mdn	M	SD	IQR	Skew	S-W p	n
Perceived Prompt Engineering Competency	3.80	3.64	0.758	0.60	−1.01	< .001	75
Basic Modality Use of LLM input (AI Input Diversity text, images, audio, video etc.)	2.80	2.75	0.770	0.80	0.49	.051	75
Strategic Integration & Application (purposeful combination of modalities)	3.00	3.26	0.740	0.67	0.08	.008	75
Depth of Use (usage of basic queries vs complex multi-modal tasks	3.00	2.79	0.923	1.25	0.02	.011	75
Perceived Self-Readiness & Access to AI tools	3.50	3.48	0.665	1.00	0.08	.003	75

Note. Mdn = Median; M = Mean; IQR = Interquartile Range; S-W = Shapiro-Wilk. All constructs measured on 1–5 scales.

Examination of the descriptive statistics revealed a notable pattern. The perceptual constructs — Perceived Prompt Competency (Mdn = 3.80) and Perceived Self-Readiness & Access to AI tools (Mdn = 3.50) scored well above the scale midpoint, indicating that academic participants generally perceived themselves as competent and equipped. In contrast, the behavioural frequency measures — Basic Modality Use of LLM Input (Mdn = 2.80) and Depth of Use Level (Mdn = 3.00) — was placed at or below the midpoint, suggesting that actual multimodal behaviour remained limited despite positive self-perceptions. This perception-behaviour gap is explored further in the analyses below.

Examination of the descriptive statistics revealed a notable pattern. The perceptual constructs — Prompt Engineering Competency (Mdn = 3.80) and Self-Assessment & Access (Mdn = 3.50) — scored well above the scale midpoint, indicating that academic participants generally perceived themselves as competent and equipped. Integration & Application had a median at the midpoint (Mdn = 3.00), but the Wilcoxon test indicated a significant upward tendency in the distribution (M = 3.26, W = 1008, p = .002, r = .333), suggesting that participants tended to score above neutral despite the median falling at 3.00. In contrast, the two behavioural frequency measures — Basic Modality Use (Mdn = 2.80) and Depth of Use (Mdn = 3.00, M = 2.79) — did not exceed the midpoint, suggesting that actual multimodal behaviour remained limited despite positive self-perceptions. This perception-behaviour gap is explored further in the analyses below

3.2 Above-Midpoint Engagement Testing

Wilcoxon Signed-Rank tests were conducted to assess whether participants' scores on each construct significantly exceeded the theoretical scale midpoint (3.0). Three of five constructs scored significantly above the midpoint. Perceived Self-Readiness & Access to AI tools exhibited the largest effect ($W = 1323$, $p < .001$, $r = .782$), followed by Perceived Prompt Competency ($W = 2197$, $p < .001$, $r = .768$), both representing large effects. Strategic Integration & Application was significantly above the midpoint with a medium effect ($W = 1008$, $p = .002$, $r = .463$).

In contrast, Basic Modality Use of LLM input ($W = 516$, $p = .997$, $r = -.417$) and Depth of Use Level ($W = 496$, $p = .963$, $r = -.281$) did not exceed the midpoint, indicating that participants' behavioural engagement with multiple modalities remained limited. This pattern is consistent with early-majority adoption as described by Rogers (2003). Academic users have developed competency through text-based interaction but have not yet explored the full multimodal capability space (Zamfirescu-Pereira et al., 2023). The finding that 58.7% of participants primarily used mobile devices for academic work may partially explain this pattern, as multimodal features such as code input, structured data upload, and complex document combination are substantially more difficult on mobile platforms.

3.3 Bivariate Associations

Spearman rank-order correlations confirmed significant positive associations between all four predictors and Perceived Prompt Competency. Strategic Integration & Application exhibited the strongest bivariate association ($\rho = .667$, $p < .001$), representing a large effect, followed by Perceived Self-Readiness & Access to AI tools ($\rho = .514$, $p < .001$), also a large effect. Depth of Use Level demonstrated a medium association ($\rho = .371$, $p = .001$), while Basic Modality Use of LLM input exhibited the weakest relationship ($\rho = .265$, $p = .022$), indicating a small association.

The magnitude ranking of these bivariate associations foreshadowed the regression results: the two strongest correlates (Strategic Integration & Application and Perceived Self-Readiness & Access to AI tools) subsequently emerged as the only significant unique predictors, while the two weaker correlates (Depth of Use and Basic Modality Use of LLM input) lost significance when competing with the others. Inter-predictor correlations ranged from $\rho = .261$ to $\rho = .473$, indicating related but non-redundant constructs.

3.4 Multiple Regression

A simultaneous multiple linear regression was conducted with Perceived Prompt Competency as the dependent variable and all four continuous predictors entered in a single block. The overall model was statistically significant and explained 41.6% of the variance in competency scores ($F(4, 70) = 12.5$, $p < .001$, $R^2 = .416$, adjusted $R^2 = .382$).

Table 3: Multiple Regression Coefficients

Predictor	B	SE	t	p
(Intercept)	0.941	0.416	2.264	.027
Predictor	**B**	**SE**	**t**	**p**
Basic Modality Use of LLM input (AI Input Diversity text, images, audio, video etc.)	−0.018	0.111	−0.163	.871
Strategic Integration & Application (purposeful combination of modalities)	0.506	0.112	4.522	< .001
Depth of Use (usage of basic queries vs complex multi-modal tasks)	−0.031	0.096	−0.325	.746
Perceived Self-Readiness & Access to AI tools	0.343	0.122	2.817	.006

Note. DV = Perceived Prompt Engineering Competency. *$N = 75$. $R^2 = .416$, adjusted $R^2 = .382$.*

Strategic Integration & Application emerged as the strongest unique predictor ($B = 0.51$, $SE = 0.11$, $t = 4.52$, $p < .001$). For each unit increase in Strategic Integration & Application, Perceived Prompt Competency increased by approximately half a scale unit, holding all other predictors constant. Perceived Self-Readiness & Access to AI

tools was the second significant predictor (B = 0.34, SE = 0.12, t = 2.82, p = .006), indicating that foundational access and technical confidence contributed to unique variance beyond behavioural engagement patterns.

Neither Basic Modality Use of LLM input (B = −0.02, SE = 0.11, t = −0.16, p = .871) nor Depth of Use Level (B = −0.03, SE = 0.10, t = −0.33, p = .746) contributed significant unique variance. Both coefficients were near zero and slightly negative, indicating that their bivariate associations with competency are entirely subsumed by Strategic Integration & Application and Perceived Self-Readiness & Access to AI tools. This pattern of results suggests that the breadth of modality use, and the basic-to-depth task progression do not independently contribute to Perceived Prompt Competency once the capacity for purposeful integration and technical readiness is accounted for.

3.5 Exploratory Barrier Analyses

A chi-square goodness-of-fit test revealed that primary adoption barriers were significantly and unevenly distributed among the six categories ($\chi^2(5, N = 75) = 37.6$, $p < .001$, Cohen's $w = 0.708$). The most frequently reported barrier was lack of knowledge (n = 28, 37.3%), followed by privacy concerns (n = 18, 24.0%) and connectivity issues (n = 15, 20.0%). Lack of tools was the least reported barrier (n = 3, 4.0%), and only 9.3% of participants reported no barriers. This distribution is consistent with Van Dijk's (2005) digital divide model, which predicts that knowledge-based access deficits override material barriers in populations with adequate technological infrastructure.

A Kruskal-Wallis test comparing Perceived Prompt Competency across all six barrier groups did not reach statistical significance ($\chi^2(5) = 5.73$, $p = .333$), though a medium effect was observed ($\varepsilon^2 = .078$). The non-significant result is attributable to the highly unequal and in some cases very small cell sizes (ranging from n = 3 to n = 28), which substantially reduced statistical power.

To provide a more balanced test of whether barrier type relates to perceived Prompt Engineering Competency, a targeted Mann-Whitney-U comparison was conducted between participants reporting lack of knowledge (n = 28) and those reporting privacy concerns (n = 18) — the two largest barrier groups. Knowledge-barrier participants scored significantly lower on Perceived Prompt Engineering Competency (Mdn = 3.60) than privacy-barrier participants (Mdn = 3.90), U = 158, p = .034, r = .373, representing a medium effect. This finding indicates that knowledge-based barriers are associated with measurable competency deficits, while attitudinal barriers (privacy concerns) are not. Privacy-concerned users appear to possess adequate prompting skills but limit their engagement due to data safety considerations.

An additional Mann-Whitney-U comparison between participants reporting any barrier (n = 68) and those reporting none (n = 7) revealed a consistent pattern across all five constructs. Barrier-free participants scored higher on every measure, with the largest effect observed for Basic Modality Use (r = .330, medium effect). However, no comparison reached statistical significance due to the small barrier-free group size (n = 7). The consistent directionality suggests a systematic barrier effect that calls for replication with larger barrier-free samples.

4. Discussion

The central finding of this study is that Perceived Prompt Competency among academic users is predicted by how purposefully they combine multiple input modalities, not by how many modalities they have tried. Strategic Integration & Application was the dominant predictor (B = 0.51, p < .001), while Basic Modality Use of LLM Input contributed no unique variance whatsoever (B = −0.02, p = .871). This distinction carries significant implications for both theory and practice.

Theoretically, the finding provides empirical support for taxonomic hierarchy to the context of human-AI interaction. Just as Bloom (1956) distinguished between lower-order knowledge acquisition (exposure, recognition) and higher-order application and synthesis (purposeful deployment), our results distinguish between lower-order modality exploration and higher-order modality integration, implying a pedagogical integration of Prompt Engineering skills.

The results indicate a notable gap between perceptual and behavioural constructs. Participants reported above-midpoint competency (Mdn = 3.80) and readiness (Mdn = 3.50) but below-midpoint diversity in modality use (Mdn = 2.80). Non-tech university student users feel competent and equipped, yet their actual reported multimodal behaviour remained limited. This perception-behaviour gap is consistent with Bandura's (1977)

observation that self-efficacy beliefs often outpace behavioural engagement, particularly during early adoption phases when users have developed confidence in familiar modalities (text) without extending that confidence to unfamiliar ones (images, audio, code).

The predominance of mobile device usage (58.7%) may partially explain this gap. Multimodal features such as code input, structured data upload, and complex document combination are substantially more difficult on mobile platforms than on laptops or desktops. Users may perceive themselves as capable of multimodal engagement while being structurally constrained by their primary device. This has direct implications for institutional AI strategy: providing students and faculty with laptop access for AI-intensive tasks may be a simple but effective intervention to bridge the perception-behaviour gap.

Perceived Self-Readiness & Access to AI tools remained a significant predictor ($B = 0.34$, $p = .006$). This finding supports both Bandura's (1977) self-efficacy theory and Van Dijk's (2005) digital divide framework. Material access and self-perceived readiness are not merely pre-requisites, rather, they continue to independently shape Perceived Prompt Engineering Competency among active users. The implication is that institutional investment in tool provision and technical support may yield ongoing returns in competency development, not merely initial adoption.

The exploratory barrier analyses revealed that lack of knowledge was the most prevalent barrier (37.3%), and that knowledge-barrier users scored significantly lower on competency than privacy-barrier users ($U = 158$, $p = .034$, $r = .373$). This convergence of barrier prevalence with competency deficits identifies a clear intervention target. The largest barrier group is also the group most in need of skill development.

5. Conclusion

This study sets out to identify which dimensions of multimodal LLM engagement predict Perceived Prompt Engineering Competency among non-tech academic users. Results indicate that it is not about how many tools are used, but about how purposefully they are synchronized together. Strategic Integration & Application -the purposeful combination of modalities— was the strongest predictor of competency, while the mere breadth of modalities explored contributed nothing unique.

For academic institutions designing AI literacy programmes, the practical implication is direct: curricula should move beyond feature-by-feature exposure toward integrative, project-based activities that require students to combine text, images, data, code and other modalities in purposeful workflows. Simultaneously, foundational readiness — both material access and technical confidence — must be addressed as an ongoing support structure, not merely as an entry requirement. And for the 37.3% of academic users who identify knowledge as their primary barrier, targeted training in multimodal workflow integration may address both the perceived obstacle and the underlying competency gap in a single intervention. The present findings suggest that the path to competency is not determined through breadth of exploration but through depth of integration indicating the need of relevant pedagogical frameworks in curriculum integration of prompt engineering (Mishra and Koehler, 2006).

6. Limitations and Future Directions

Several limitations should be noted when interpreting these findings. First, the cross-sectional design captures a single point in time, which prevents making definitive claims about cause and effect. Future longitudinal research is recommended to track how these user habits develop over time. Second, while the regression model explains a meaningful portion of the variance ($R^2 = .416$), a significant amount remains unexplained; future studies should explore additional variables like intrinsic motivation and trust in automation. Third, the reliance on self-reported data introduces potential method bias, which future research could address by incorporating objective performance metrics. Fourth, while the sample size was sufficient for the primary analysis, it limited the capacity for detailed subgroup comparisons and broader model validation, highlighting the need for larger, more diverse samples in future studies.

Ethics Declaration

This research received approval from the Ethics Committee of Amrita Vishwa Vidyapeetham under Ref No.: IHEC/2025/302, ensuring that all ethical protocols were followed during the data collection process involving participants.

AI Declaration

The authors acknowledge the use of generative artificial intelligence tools and large language models from Anthropic (2026), specifically Claude Sonnet 4.5, Claude Opus 4.6, and Claude Opus 4.7, alongside tools from Google (2026), specifically Gemini Flash 3.1 and Gemini Flash 3.5. These models were utilized solely for grammatical corrections, stylistic editing, inconsistency identification, and formatting alignment with Harvard referencing requirements. All original conceptual synthesis, empirical interpretations of the statistical data, and final intellectual contributions remain entirely the responsibility of the human authors.

References

Bandura, A. (1977) 'Self-efficacy: Toward a unifying theory of behavioural change', Psychological Review, 84(2), pp. 191–215.

Bloom, B. S. (1956) Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. New York: Longmans, Green.

Cohen, J. (1988) Statistical power analysis for the behavioural sciences. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum.

Dreyfus, H. L. and Dreyfus, S. E. (1980) A five-stage model of the mental activities involved in directed skill acquisition. Technical Report. Operations Research Centre, University of California, Berkeley.

Field, A. (2013) Discovering statistics using IBM SPSS Statistics. 4th edn. London: Sage.

Hayes, A. F. (2018) Introduction to mediation, moderation, and conditional process analysis. 2nd edn. New York: Guilford Press.

Idris, M. D., Feng, X. and Dyo, V. (2024) 'Revolutionizing higher education: Unleashing the potential of large language models for strategic transformation', IEEE Access, 12, pp. 67738–67757.

Mishra, P. and Koehler, M. J. (2006) 'Technological pedagogical content knowledge: A framework for teacher knowledge', Teachers College Record, 108(6), pp. 1017–1054.

Norman, G. (2010) 'Likert scales, levels of measurement and the ‘laws’ of statistics', Advances in Health Sciences Education, 15(5), pp. 625–632.

Rogers, E. M. (2003) Diffusion of innovations. 5th edn. New York: Free Press.

Stevens, J. P. (1996) Applied multivariate statistics for the social sciences. 3rd edn. Mahwah, NJ: Lawrence Erlbaum.

van Dijk, J. (2005) The deepening divide: Inequality in the information society. Thousand Oaks, CA: Sage.

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B. and Yang, Q. (2023) 'Why Johnny can’t prompt: How non-AI experts try (and fail) to design LLM prompts', in Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Hamburg, Germany, pp. 1–21.

Zhou, L., Zhang, R., Dai, X., Hershcovich, D. and Li, H. (2026) 'ScholarLens: Tracking the growing penetration of large language models in scholarly writing and peer review', Engineering Applications of Artificial Intelligence.