



Validation of the Perceived Usefulness of Artificial Intelligence Scale among Preservice Mathematics Teachers

Yusuf Feyisara Zakariya^{1,2*}, Jamilu Abdullahi¹, Kazaik Benjamin Danlami³

¹Department of Science Education, Ahmadu Bello University, Zaria, Kaduna State Nigeria

²TETFund Centre of Excellence in Pedagogy, Ahmadu Bello University, Zaria, Kaduna State Nigeria

³Department of Science Education, University of Maiduguri, Borno State, Nigeria

*Corresponding Author: yfzakariya@abu.edu.ng

Abstract

The rapid integration of artificial intelligence (AI) into education has heightened the need to understand preservice teachers' perceptions of AI usefulness in mathematics learning, particularly in Nigerian teacher-education contexts. This study validated the perceived usefulness of AI scale among preservice mathematics teachers in selected higher institutions in the Federal Capital Territory (FCT), Abuja, focusing on the psychometric properties of an adapted six-item Perceived Usefulness of AI Scale and differences by institution and gender. A cross-sectional survey design was employed with 221 preservice mathematics teachers. Data were analysed using confirmatory factor analysis, coefficient omega, Kruskal–Wallis H test, and Mann–Whitney U test. The initial one-factor model demonstrated inadequate fit, but improved after specifying a theoretically plausible correlated residual between Items 5 and 6 (comparative fit index – CFI = .939, standardised root mean square residual – SRMR = .052, root mean square error of approximation – RMSEA = .093, Tucker–Lewis index – TLI = .885), with satisfactory reliability ($\omega = .804$). No significant differences were found by institution or gender. The findings provide preliminary evidence supporting the scale's validity and reliability.

Keywords: Artificial intelligence; perceived usefulness; preservice mathematics teachers; technology acceptance; mathematics teacher education; higher education.

Introduction

Artificial intelligence (AI) is reshaping education by expanding access to information, learning resources, problem-solving support, personalised learning, and instructional assistance. The emergence of generative AI has accelerated this transformation through systems that can produce explanations, examples, instructional materials, and solutions to complex tasks. This development is particularly significant for teacher education because preservice teachers are preparing to teach in increasingly AI-mediated learning environments and will be expected to use, evaluate, and regulate such technologies in professional practice (Ng et al., 2025). Mathematics education is especially relevant because AI can support activities involving problem solving, reasoning, representation, mathematical communication, lesson preparation, and feedback. Yet the growing availability of AI does not necessarily translate into educational usefulness. Concerns about inaccurate or fabricated outputs,

overdependence, academic integrity, learner autonomy, and privacy have raised questions about when and how AI should be used in education (Cotton et al., 2024; Niu et al., 2024). Consequently, the educational value of AI depends not only on its capabilities but also on whether prospective teachers perceive it as useful and can critically evaluate its outputs.

Perceived usefulness offers a focused way of examining this issue. Davis (1989) defined perceived usefulness as the extent to which an individual believes that using a particular system will enhance task performance. Within the Technology Acceptance Model (TAM), perceived usefulness is a central determinant of technology acceptance and behavioural intention (Venkatesh & Davis, 2000). In mathematics teacher education, perceived usefulness can therefore be understood as preservice teachers' belief that AI can improve learning, facilitate mathematical problem solving,

support lesson preparation, or enhance instructional effectiveness. Recent studies have examined perceived usefulness alongside attitudes, acceptance, and intentions to use generative AI and ChatGPT, demonstrating its continuing relevance to understanding technology adoption in educational settings (e.g., Sallam et al., 2024). What remains less clear, however, is whether perceived usefulness of AI has been measured with sufficient psychometric rigour among preservice mathematics teachers.

This measurement issue represents an important gap. Much of the emerging literature has examined AI attitudes, acceptance, behavioural intention, AI literacy, readiness, and general perceptions rather than establishing whether perceived usefulness itself is adequately operationalised (Bergdahl & Sjöberg, 2025; Guan et al., 2025; Taşkın, 2026). These constructs are related but not interchangeable. Moreover, adapting an established technology-acceptance measure to a new technology, population, or educational context does not automatically preserve its original measurement properties. Evidence of dimensionality, reliability, and construct validity is therefore necessary before scores from an adapted instrument can be interpreted as meaningful indicators of perceived usefulness (Zakariya, 2022). Without such evidence, apparent differences between groups may reflect weaknesses in measurement rather than genuine differences in perceptions.

The measurement problem becomes particularly consequential when perceived usefulness is compared across institutional settings and gender. Differences in technological infrastructure, institutional support, lecturer practices, access to digital resources, and exposure to AI may influence how preservice teachers evaluate its usefulness. At the same time, similarities in scores across groups cannot be confidently interpreted unless the instrument represents the underlying construct adequately across those groups (Vandenberg & Lance, 2000). Evidence concerning gender differences in technology acceptance has also been inconsistent across contexts (Xue et al., 2026), suggesting that assumptions about who is more likely to perceive AI as useful should not be generalised without empirical evidence. Thus, two questions remain intertwined: does an adapted measure adequately capture perceived usefulness of AI, and, once this is established, do perceptions differ across institutional and gender groups?

The present study addresses these questions among preservice mathematics teachers in selected higher institutions in the Federal Capital Territory (FCT), Abuja, Nigeria. It evaluated the factorial

structure and internal consistency of an adapted six-item Perceived Usefulness of AI (PU-AI) Scale and examined whether perceived usefulness differs across institutional types and gender. The study therefore moved beyond simply asking whether preservice teachers view AI positively by first establishing whether perceived usefulness is measured adequately and then determining whether meaningful group differences exist. Specifically, the study explored the one-factor structure and reliability of the PU-AI Scale and tests differences in perceived usefulness across institutions and gender. This measurement-first approach provides a more defensible basis for interpreting preservice mathematics teachers' perceptions of AI in teacher education.

Research Questions

The following research questions guided the study:

1. Do the items of the adapted instrument measure the dimension of perceived usefulness of AI they are purported to measure?
2. What is the strength of the relationship between the items and their purported construct of perceived usefulness of AI?
3. How accurately do the items measure the construct of perceived usefulness of AI they are purported to measure?
4. What is the difference in perceived usefulness of AI among preservice mathematics teachers across higher institutions in FCT, Abuja?
5. What is the difference in perceived usefulness of AI between male and female preservice mathematics teachers in FCT, Abuja?

Methodology

Research Design

The study adopted a quantitative cross-sectional survey design. Data were collected from preservice mathematics teachers to examine the construct validity and reliability of the Perceived Usefulness of AI (PU-AI) Scale and to investigate differences in perceived usefulness across institutional type and gender.

Study Context and Participants

The study was conducted among preservice mathematics teachers enrolled in three higher institutions in the Federal Capital Territory (FCT), Abuja: a university, a college of education, and an open higher institution. The sample comprised 221 preservice mathematics teachers, including 71

(32.1%) from the university, 73 (33.0%) from the college of education, and 77 (34.8%) from the open higher institution. By gender, 109 (49.3%) participants were male and 112 (50.7%) were female.

Description of the Instrument

Data were collected using the six-item Perceived Usefulness of AI (PU-AI) Scale, adapted from the perceived usefulness subscale of Davis

(1989). An adaption was done as developing a new instrument may lack a strong theoretical footing as in TAM that has stood the test of the time. The original 12-item instrument comprised six perceived usefulness and six perceived ease-of-use items. For this study, the six perceived usefulness items were adapted by replacing the original technology and work-related context with an AI and mathematics-learning context. Table 1 presents the original and adapted items.

Table 1

Comparison of the Original Perceived Usefulness Subscale and the Adapted PU-AI Scale

Original item statements	Adapted item statements
Using CHART-MASTER in my job would enable me to accomplish more quickly	Using AI tools in learning mathematics would enable me to accomplish more quickly.
Using CHART-MASTER would improve my job performance.	Using AI tools would improve my performance in mathematics.
Using CHART-MASTER in my job would increase my productivity	Using AI tools in learning mathematics would increase my productivity.
Using CHART-MASTER would enhance my effectiveness on the job	Using AI tools would enhance my effectiveness in solving mathematical problems.
Using CHART-MASTER would make it easier to do my job	Using AI tools would make it easier to learn mathematics
I would find CHART-MASTER useful in my job	I would find AI tools useful for learning mathematics

The adapted items were rated on a five-point Likert scale ranging from 1 = Strongly Disagree to 5 = Strongly Agree, with higher scores indicating greater perceived usefulness of AI for learning mathematics. The six items yielded a total score ranging from 6 to 30.

Construct Validity and Reliability Procedures

Construct validity was examined using a one-factor confirmatory factor analysis (CFA) conducted in Mplus. Model fit was evaluated using the chi-square statistic (χ^2), chi-square-to-degrees-of-freedom ratio (χ^2/df), root mean square error of approximation (RMSEA) with its 90% confidence interval, comparative fit index (CFI), Tucker–Lewis index (TLI), and standardised root mean square residual (SRMR). Standardised factor loadings were examined to assess the contribution of individual items to the latent construct. Reliability coefficient was evaluated using coefficient omega (ω).

Data Collection and Analysis

Data were collected from pre-service mathematics teachers in three participating higher institutions in FCT, Abuja. Following institutional

approval, researchers administered paper-based questionnaires directly to eligible participants. Participants were informed about the study purpose, voluntary participation, confidentiality, and the absence of academic consequences. Clear completion instructions were provided. Completed questionnaires were collected immediately and screened for completeness. Of the returned questionnaires, 221 valid responses were retained for analysis, coded, and entered into statistical software for subsequent statistical analysis. Meanwhile, some datasets were still missing which explains the discrepancies reported in the result section. Descriptive statistics, including means and standard deviations, were used to summarise PU-AI scores across institutional type and gender, with bar charts used to present group comparisons. Differences across the three institutions were examined using the Kruskal–Wallis H test, while gender differences were examined using the Mann–Whitney U test. Analyses were conducted using IBM SPSS Statistics, with statistical significance set at $\alpha = .05$. Missing PU-AI cases were excluded from the relevant analyses using available cases.

Results

The results are presented according to the five research questions. Research Questions One to Three address the construct validity and measurement properties of the PU-AI Scale, while Research Questions Four and Five examine differences in perceived usefulness across institutional type and gender, respectively.

To address Research Question One, a six-item, one-factor confirmatory factor analysis (CFA) was conducted to determine whether the six observed items adequately represented the latent construct of perceived usefulness of AI.

Table 2

The goodness of fit statistics for a 1-factor PU scale

Global fit statistics	Model 1	Model 2
Chi-square		
Estimate (χ^2)	39.739	23.312
Degrees of freedom (df)	9	8
p-value	<.001	.003
χ^2 / df	4.415	2.914
RMSEA (Root Mean Square Error of Approximation)		
Estimate	.124	.093
90 percent confidence interval	[.086, .165]	[.050, .138]
Probability RMSEA $\leq$.05	.001	.049
CFI/TLI		
CFI (Comparative Fit Index)	.877	.939
TLI (Tucker-Lewis index)	.795	.885
SRMR (Standardized Root Mean Square Residual)	.063	.052

The results in Table 2 show that the initial one-factor model (Model 1) did not fit the data adequately, $\chi^2(9) = 39.739$, $p < .001$, $\chi^2/df = 4.415$, RMSEA = .124, CFI = .877, TLI = .795, and SRMR = .063. After specifying a correlated residual between Items 5 and 6, Model 2 showed improved fit, $\chi^2(8) = 23.312$, $p = .003$, $\chi^2/df = 2.914$, RMSEA = .093, CFI = .939, TLI = .885, and SRMR = .052. Although CFI and SRMR indicated acceptable fit, RMSEA and TLI suggested residual misfit. Thus, Model 2 provided a substantially improved and reasonably adequate representation of the data, but

not unequivocally optimal fit (Hu & Bentler, 1999). The findings provide preliminary support for the one-factor structure and warrant further validation in independent samples.

Research Question Two was addressed by examining the standardized factor loadings of the six PU-AI items on the latent perceived usefulness factor. The CFA results showed that the items loaded appropriately on the perceived usefulness construct. The structural model of the instrument and selected local fit statistics are presented in Figure 1.

Figure 1

Standardised estimates of model parameters for the 1-factor PU scale

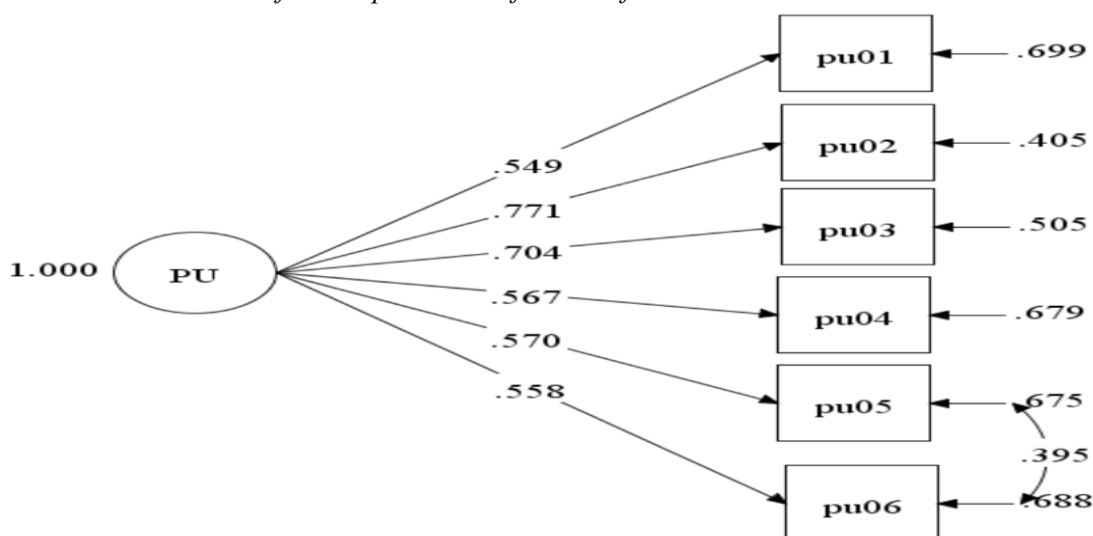
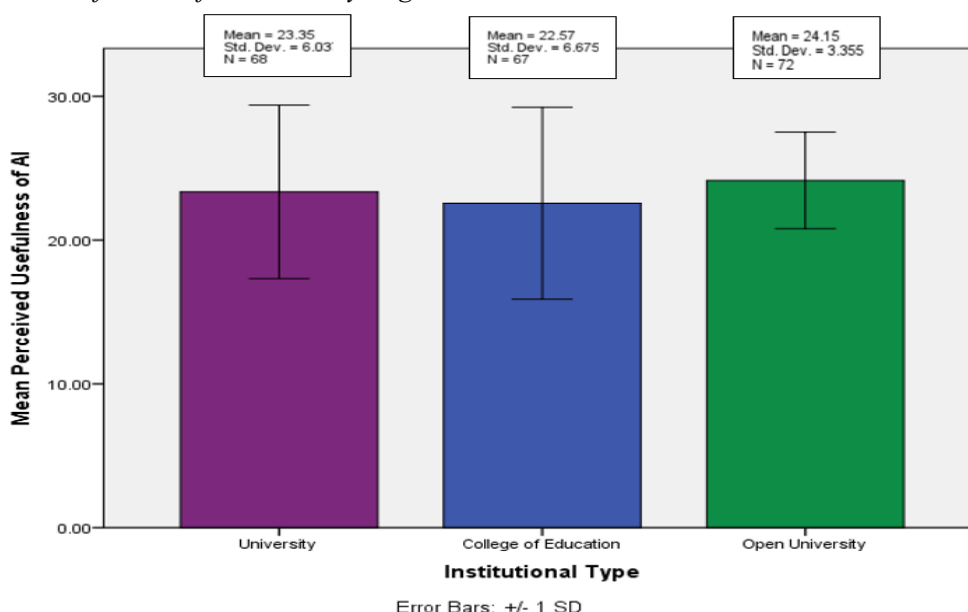


Figure 1 shows the standardised estimates of the local fit statistics (e.g., factor loadings), suggesting that each item measures the corresponding dimension of perceived usefulness of AI scale appropriately. This follows from the assertion that a factor loading of at least .40 is acceptable in educational and social science research (Brown, 2015). Thus, the global (Table 1) and the local (Figure 2) fit statistics suggest that PU-AI Subscale demonstrated an acceptable construct validity (Brown, 2015).

To address Research Question Three, the reliability coefficient of the six-item PU-AI Scale was estimated using coefficient omega. The scale

produced a coefficient omega of .804, indicating satisfactory reliability among the items. The coefficient omega suggests that the six items demonstrate sufficient shared variance to function as indicators of the same underlying perceived usefulness of AI construct. The findings therefore provide preliminary evidence of satisfactory reliability for the PU-AI Scale in the present sample. We explored the generated dataset for its descriptive characteristics across the three different higher institutions and visualised the dataset using a simple bar chart. The results of this analysis are presented in Figure 2.

Figure 2
Mean Perceived Usefulness of AI Scores by Higher Institution



The descriptive results presented in Figure 2 indicate that preservice mathematics teachers from the Open University recorded the highest perceived usefulness of AI score ($M = 24.15$, $SD = 3.355$), followed by those from the University ($M = 23.35$, $SD = 6.03$), while participants from the College of Education recorded the lowest mean score ($M = 22.57$, $SD = 6.675$). Furthermore, we examined

whether the observed differences in mean scores across the higher institutions were statistically significant using an inferential statistical test. Prior to this analysis, the normality assumption was assessed to determine whether a parametric or non-parametric test was appropriate. Table 3 presents the results of the normality tests.

Table 3
Tests of Normality of PU-AI Scale across institutional type

	Institutional Type	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	p-value	Statistic	Df	p-value
Perceived Usefulness of AI	University	.154	68	<.001	.900	68	<.001
	College of Education	.183	67	<.001	.890	67	<.001
	Open University	.100	72	.073	.970	72	.079

a. Lilliefors Significance Correction

The results presented in Table 3 show that the normality assumption was violated for the University and College of Education groups, as both the Kolmogorov-Smirnov and Shapiro-Wilk tests were statistically significant ($p < .05$). Although the Open University scores did not significantly depart from normality, the assumption of normality was not

satisfied across all three groups. Consequently, the Kruskal-Wallis H test was conducted to determine whether the distribution of PU-AI differed significantly among preservice mathematics teachers across the three higher institutions. The results of this analysis are presented in Table 4.

Table 4
Kruskal-Wallis H Test of Perceived Usefulness of AI by Institution

Comparison	n	Kruskal-Wallis H	df	p-value	ϵ^2
University of Abuja, COE Zuba, and NOUN	207	0.408	2	.816	<.001

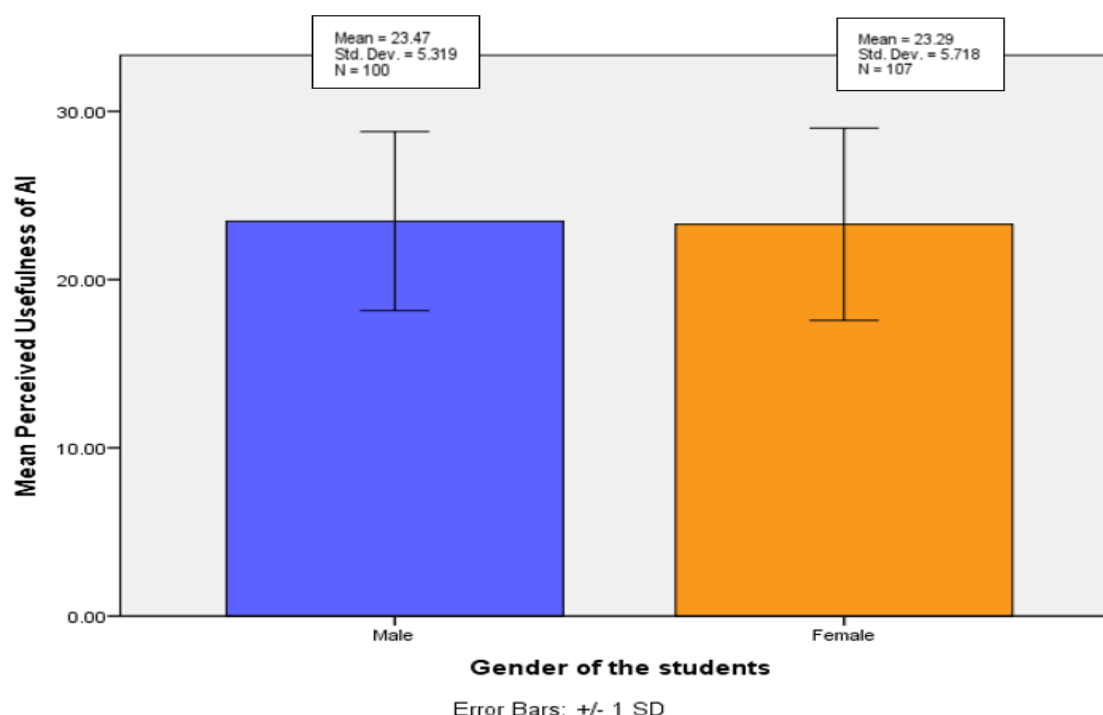
As shown in Table 4, the Kruskal-Wallis H test revealed no statistically significant difference in the distribution of perceived usefulness of AI among preservice mathematics teachers across the three higher institutions, $H(2) = 0.408$, $p = .816$. Since the p-value was greater than the .05 significance level, the null hypothesis was retained. The effect size was negligible ($\epsilon^2=0.00$), indicating that institutional affiliation accounted for virtually none of the variation in perceived usefulness of AI. Therefore, although the descriptive statistics showed some

variation in mean scores across the institutions, these differences were neither statistically significant nor practically meaningful. Consequently, no post-hoc pairwise comparisons were warranted.

Addressing Research Question Five

We explored the generated dataset for its descriptive characteristics across male and female preservice mathematics teachers and visualised the dataset using a simple bar chart. The results of this analysis are presented in Figure 3.

Figure 3
Mean Perceived Usefulness of AI Scores by Gender



The descriptive results in Figure 3 show that male preservice mathematics teachers had a mean perceived usefulness of AI score of $M = 23.47$ ($SD = 5.32$), compared with $M = 23.29$ ($SD = 5.72$) for females. Although males recorded a slightly higher

mean, the difference was small. To determine whether this difference was statistically significant, an inferential test was conducted. Normality was first assessed to determine the appropriate statistical procedure, with the results presented in Table 5.

Table 5
Tests of Normality of PU-AI Scale across institutional type

Gender of students	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Perceived Usefulness of AI	.121	100	.001	.916	100	<.001
Male						
Female	.129	107	<.001	.906	107	<.001

a. Lilliefors Significance Correction

Table 5 shows that PU-AI scores violated the normality assumption for both male and female participants, as indicated by significant Kolmogorov–Smirnov and Shapiro–Wilk tests ($p \leq .001$). Therefore, a non-parametric test was

appropriate, and the Mann–Whitney U test was used to examine gender differences in perceived usefulness of AI. The results are presented in Table 6.

H_{02} : There is no significant difference in perceived usefulness of AI between male and female preservice mathematics teachers in FCT, Abuja.

Table 6
Mann–Whitney U Test of PU-AI by Gender

Comparison	N	Mann–Whitney U	Z	p-value	Effect size (r)
Male vs. Female	207	5352.500	0.006	.995	.001

As shown in Table 6, the Mann–Whitney U test indicated no statistically significant difference in perceived usefulness of AI between male and female

preservice mathematics teachers, $U = 5352.50$, $z = 0.006$, $p = .995$. The null hypothesis was therefore retained. The effect size was negligible ($r = .00$),

indicating virtually no practical effect of gender on perceived usefulness of AI. Thus, male and female preservice mathematics teachers reported statistically and practically similar levels of perceived usefulness.

Discussion

The study yielded two main findings. First, the adapted six-item PU-AI Scale demonstrated preliminary factorial validity and satisfactory internal consistency ($\omega = .804$), although some residual misfit remained. This supports perceived usefulness as a relevant construct for examining AI acceptance among preservice mathematics teachers while underscoring the need for contextual validation of adapted technology-acceptance measures (Ma & Lei, 2024; Yao & Wang, 2024). Second, perceived usefulness did not differ significantly across institutions or gender, with negligible effect sizes. This suggests that institutional type and gender contributed little to variation in perceived usefulness within the study context. However, these findings do not preclude differences in other AI-adoption factors, such as self-efficacy, trust, perceived risk, behavioural intention, or actual use.

The findings extend the application of the Technology Acceptance Model (TAM) to AI in mathematics teacher education. The initial model misfit highlights the importance of contextual validation when TAM-based measures are adapted to AI. The study therefore supports a measurement-first approach in which instrument adequacy is established before interpreting substantive group differences. The negligible institutional and gender differences further indicate relatively stable perceptions of AI usefulness across the groups studied. The findings suggest that AI preparation in mathematics teacher education should emphasise common competencies rather than primarily differentiating provision by institution or gender. Preservice teachers should develop skills in evaluating AI-generated mathematical outputs, identifying errors, using effective prompts, protecting learner privacy, recognising AI limitations, and integrating AI with professional mathematical and pedagogical judgement. Researchers should likewise establish adequate validity and reliability before using adapted AI-acceptance measures for group comparisons or for programme evaluation.

Conclusion

The study provides preliminary evidence that the adapted six-item PU-AI Scale is a potentially

reliable measure of perceived usefulness of AI among preservice mathematics teachers, although further validation is warranted. Perceived usefulness did not differ significantly across institutions or gender, and the practical effects were negligible. The findings therefore suggest broadly similar perceptions of AI usefulness among the participating preservice teachers. More importantly, they demonstrate the need for rigorous measurement validation before adapted technology-acceptance instruments are used to draw substantive or group-comparison conclusions. Future studies should validate the scale across diverse samples, establish measurement invariance, and incorporate additional AI-acceptance constructs.

References

- Bergdahl, N., & Sjöberg, J. (2025). Attitudes, perceptions and AI self-efficacy in K-12 education. *Computers and Education: Artificial Intelligence*, 8, 100358.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.
- Guan, L., Zhang, Y., & Gu, M. M. (2025). Pre-service teachers preparedness for AI-integrated education: An investigation from perceptions, capabilities, and teachers' identity changes. *Computers and Education: Artificial Intelligence*, 8, 100341. <https://doi.org/10.1016/j.caeai.2024.100341>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Ma, S., & Lei, L. (2024). The factors influencing teacher education students' willingness to adopt artificial intelligence technology for information-based teaching. *Asia Pacific Journal of Education*, 44(1), 94–111. <https://doi.org/10.1080/02188791.2024.2305155>

- Ng, D. T. K., Chan, E. K. C., & Lo, C. K. (2025). Opportunities, challenges and school strategies for integrating generative AI in education. *Computers and Education: Artificial Intelligence*, 8, 100373.
- Niu, W., Zhang, W., Zhang, C., & Chen, X. (2024). The role of artificial intelligence autonomy in higher education: A uses and gratification perspective. *Sustainability*, 16(3), 1276.
- Sallam, M., Elsayed, W., Al-Shorbagy, M., Barakat, M., El Khatib, S., Ghach, W., Alwan, N., Hallit, S., & Malaeb, D. (2024). ChatGPT usage and attitudes are driven by perceptions of usefulness, ease of use, risks, and psycho-social impact: A study among university students in the UAE. *Frontiers in Education*, 9, 1414758. <https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2024.1414758/full>
- Taşkın, T. (2026). Artificial intelligence in teacher education: Exploring the role of AI literacy on attitudes. *Journal of Research in Innovative Teaching & Learning*, 1–14.
- Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, 3(1), 4–70. <https://doi.org/10.1177/109442810031002>
- Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.
- Xue, L., Mahat, J., & Ghazali, N. (2026). Technology Acceptance Model in Artificial Intelligence in Education: A Meta-Analysis. *Sage Open*, 16(1), 21582440251409441. <https://doi.org/10.1177/21582440251409441>
- Yao, N., & Wang, Q. (2024). Factors influencing pre-service special education teachers' intention toward AI in education: Digital literacy, teacher self-efficacy, perceived ease of use, and perceived usefulness. *Heliyon*, 10(14). [https://www.cell.com/heliyon/fulltext/S2405-8440\(24\)10925-5](https://www.cell.com/heliyon/fulltext/S2405-8440(24)10925-5)
- Zakariya, Y. F. (2022). Cronbach's alpha in mathematics education research: Its appropriateness, overuse, and alternatives in estimating scale reliability. *Frontiers in Psychology*, 13, 1074430.