

Generative AI Adoption in Statistics and Data Science Higher Education

Noha Youssef

nayoussef@aucegypt.edu

American University in Cairo

Research Article

Keywords: Generative AI, Quantitative STEM education, Statistics education, Data science education, UTAUT2, PLS-SEM

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11132628/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Generative AI Adoption in Statistics and Data Science Higher Education

Noha Youssef

Department of Mathematics and Actuarial Science, School of Sciences and Engineering

The American University in Cairo, New Cairo, Egypt

Corresponding author: nayoussef@aucegypt.edu

Abstract

Generative AI (GenAI) is increasingly embedded in advanced quantitative STEM learning, where students use it for conceptual explanation, coding, debugging, model interpretation, and analytical writing. Yet the conditions that shape adoption in courses requiring statistical judgment and evidence evaluation remain underexamined. We surveyed 72 senior undergraduates enrolled in advanced statistics and data science courses at a private university in Egypt. An adapted UTAUT2 model was estimated using PLS-SEM and supplemented with descriptive analysis of open-ended responses. Effort expectancy was the strongest predictor of behavioral intention ($\beta = 0.562, p < .001$), followed by facilitating conditions ($\beta = 0.329, p = .003$) and self-efficacy ($\beta = 0.195, p = .006$); behavioral intention strongly predicted self-reported use ($\beta = 0.709, p < .001$). The model explained 73.2% of behavioral intention and 50.3% of use behavior. Effect sizes and sensitivity analyses indicated that the structural pattern was stable, although discriminant-validity concerns remained for adjacent constructs. Out-of-sample predictive evidence was weak, so the model is interpreted primarily as explanatory. Open-ended responses showed that students valued immediacy and coding support while remaining concerned about inaccurate outputs, overreliance, verification, and policy clarity. The findings distinguish adoption from responsible disciplinary use: ease, support, and confidence encourage uptake, but responsible use additionally requires verification, statistical judgment, evidence evaluation, and transparent disclosure.

Keywords: Generative AI; Quantitative STEM education; Statistics education; Data science education; UTAUT2; PLS-SEM

1. Introduction

Generative AI has moved rapidly into academic work, including explanation, coding, troubleshooting, and writing. In science and STEM education, its significance extends beyond information retrieval because it enters reasoning, modeling, evidence evaluation, and explanation. Recent scholarship therefore asks not only whether AI supports learning, but how it reshapes disciplinary practice and what guidance is needed for responsible use (Lee et al., 2026; Tang & Putra, 2026).

This issue is especially important in statistics and data science, where conceptual reasoning, programming, model diagnostics, uncertainty, and communication are tightly coupled. GenAI can reduce procedural friction by explaining concepts, suggesting code, or interpreting output (Ellis & Slade, 2023; Bien & Mukherjee, 2025), but the same fluency can conceal incorrect code, unjustified model choices, or misleading interpretations (DeLuca et al., 2025; Rao et al., 2026). In these courses, a tool that makes work easier can also change how students approach evidence and judgment.

Much higher-education research examines general perceptions, writing support, or integrity concerns. Science-education research has begun to address GenAI-influenced scientific practice and argumentation (Cheung & Zhang, 2025; Tang & Putra, 2026), but adoption in advanced quantitative courses, where AI directly intersects with coding, model selection, diagnostics, and interpretation, remains less understood. This matters because responsible-use instruction operates only after AI has entered students' disciplinary workflow.

Three gaps motivate the present study. First, we know relatively little about the acceptance factors associated with GenAI use in advanced quantitative STEM courses rather than in general academic settings. Second, the meaning of familiar technology-acceptance constructs may shift when the technology generates

code, analysis, and explanations that appear authoritative. Ease of use and institutional support may facilitate adoption, but neither guarantees that students verify outputs or preserve disciplinary reasoning. Third, adoption studies often stop at intention or use, whereas science education increasingly emphasizes how students evaluate evidence and challenge AI outputs after adoption.

Accordingly, the study addresses two overarching questions: (1) Which acceptance-related factors are associated with students' behavioral intention and self-reported GenAI use in advanced statistics and data science coursework? (2) How do students' open-ended accounts of benefits, risks, and policy expectations help interpret those quantitative relationships? We use an adapted UTAUT2 framework to answer the first question and descriptive qualitative evidence to contextualize the second.

The study makes three contributions. First, it provides discipline-specific evidence from advanced quantitative STEM courses where GenAI is used in modeling, coding, diagnostics, and interpretation. Second, it evaluates the retained relationships with effect sizes, predictive assessment, bootstrapped discriminant-validity diagnostics, and sensitivity analyses rather than path significance alone. Third, it distinguishes adoption from responsible disciplinary use: whether GenAI enters students' workflow is different from whether students verify outputs, exercise model judgment, evaluate evidence, and disclose AI assistance appropriately.

The remainder of the paper is organized as follows. Section 2 reviews GenAI in quantitative and science education and positions the study within UTAUT2-based adoption research. Section 3 describes the conceptual framework, learning context, sample, instrument, and PLS-SEM procedures. Section 4 presents the measurement and structural results, predictive assessment, robustness checks, and descriptive analysis of the open-ended responses. Section 5 discusses the theoretical and pedagogical implications, limitations, and directions for future research. Section 6 concludes the paper.

2. Literature Review

Generative AI now supports explanation, programming, summarization, and drafting in higher education, while instructional policies and assessment practices continue to evolve (Chan & Hu, 2023; Ellis & Slade, 2023; Kasneci et al., 2023). Its immediacy can support self-directed learning, but plausible errors, overreliance, and academic-integrity concerns make students' ability to evaluate outputs central (Celik et al., 2022).

These tensions are especially visible in statistics and data science, where students coordinate conceptual understanding, computation, statistical judgment, and communication (Nolan & Temple Lang, 2010; Wild & Pfannkuch, 1999). AI can explain models, generate or debug code, and interpret output (Bien & Mukherjee, 2025; Ellis & Slade, 2023), yet procedural help without statistical judgment can produce runnable code or fluent explanations while masking unsuitable assumptions or incorrect interpretations (DeLuca et al., 2025; Rao et al., 2026). The educational issue is therefore not simply whether AI can complete a task, but whether students retain responsibility for the reasoning that makes the task valid.

Discipline-specific studies indicate that GenAI can support project-based statistics work and engagement (Al Labadi & Ly, 2025; Ahmed et al., 2025). Science-education research similarly emphasizes reasoning-centered and responsible use rather than treating AI as an unquestioned authority (Lee et al., 2026; Tang & Putra, 2026). Cheung and Zhang (2025) further show that undergraduates' views of GenAI are connected to how they understand scientific practice and knowledge production. Together, this literature suggests that adoption should be studied in relation to the disciplinary practices into which AI is being introduced.

UTAUT and UTAUT2 provide a useful starting point for examining adoption because they distinguish several mechanisms through which users form intentions and translate them into behavior (Venkatesh et al., 2003, 2012, 2016). Performance expectancy captures perceived usefulness, effort expectancy captures perceived ease, facilitating conditions capture the support and resources available for use, and hedonic motivation reflects enjoyment. In the present study, self-efficacy is included as an extension because advanced quantitative use also requires confidence in interacting with AI for coding, analysis, and interpretation.

These constructs may take somewhat different meanings in a generative-AI setting. Effort expectancy can be understood as the friction involved in bringing AI into an existing analytical workflow. Facilitating conditions may extend beyond technical access to include instructor guidance, permitted-use boundaries, and expectations for disclosure. Self-efficacy may matter because students need enough confidence to use the

tool interactively, but confidence alone does not establish that they can evaluate its output. Performance expectancy is also theoretically relevant, although its discriminating role may weaken if students already take AI’s usefulness for granted. These possibilities justify evaluating UTAUT2 empirically in advanced quantitative coursework rather than assuming that its conventional ordering transfers unchanged.

Prior adoption studies have identified performance expectancy, facilitating conditions, trust, and institutional legitimacy as relevant to GenAI use (Sergeeva et al., 2025; Namatovu & Kyambade, 2025; Rahi et al., 2025). In science education, Ramnarain et al. (2025) combined PLS-SEM with qualitative evidence to examine pre-service teachers’ intentions to use GenAI in inquiry-based teaching, illustrating the value of connecting adoption constructs with disciplinary educational practice. The present study extends this line of work to advanced statistics and data science, where coding and statistical reasoning are closely intertwined.

Accordingly, the study examines which acceptance-related factors are associated with students’ intention to use GenAI and whether intention is associated with self-reported use. Open-ended responses are then used to contextualize the quantitative results by showing how students describe the benefits, risks, and instructional boundaries of AI use in their own coursework.

3. Methodology

3.1 Research Framework

UTAUT2 served as the theoretical starting point because the study focuses on students’ intention to adopt and use a digital technology within an educational setting. The model is particularly suitable for examining technology adoption because it distinguishes between beliefs about usefulness, ease of use, enabling conditions, motivation, and behavioral intention (Venkatesh et al., 2012). However, rather than applying the full UTAUT2 model mechanically, the framework was adapted to the specific characteristics of generative AI use in advanced quantitative coursework.

The candidate structural model therefore included performance expectancy (PE), effort expectancy (EE), facilitating conditions (FC), hedonic motivation (HM), and self-efficacy (SE) as antecedents of behavioral intention (BI), with BI predicting use behavior (UB). Each construct was included for a distinct theoretical reason. Performance expectancy was included because students may be more willing to adopt GenAI when they believe it improves their learning, coding, or analytical performance. Effort expectancy was included because ease of interaction is especially relevant for conversational AI tools, which students often use to reduce friction in coding, debugging, and interpretation. Facilitating conditions were included because access, instructor guidance, course policy, and institutional support can shape whether students feel able to incorporate AI into their coursework. Hedonic motivation was included because enjoyment and interest are part of the original UTAUT2 framework and may influence continued engagement with interactive AI tools. Self-efficacy was added as a context-specific extension because advanced statistics and data science tasks require students to feel confident not only in using AI generally, but also in applying it to coding, analysis, and interpretation.

Performance expectancy was represented initially by two dimensions to better reflect the dual role of GenAI in quantitative education. PE_Learn captured perceived support for understanding statistical ideas, concepts, and explanations, whereas PE_Tech captured perceived support for computational tasks such as coding, implementation, and debugging. This distinction was theoretically important because students may perceive AI as useful for conceptual learning without necessarily perceiving it as equally useful for technical problem solving, or vice versa. The performance-expectancy hypothesis was therefore evaluated through these two operational dimensions rather than through a single undifferentiated PE score.

Some original UTAUT2 constructs were not retained because they were less appropriate for the study context. Price value was omitted because most students had access to free or institutionally available AI tools and therefore showed little meaningful variation in perceived cost. Habit was excluded because GenAI use was still relatively recent and had not yet become sufficiently routinized to justify treating habit as a stable behavioral determinant. Social influence was considered initially, but preliminary measurement diagnostics indicated weak measurement quality and limited relevance to the largely self-directed use examined in these

courses. These exclusions were therefore based on contextual relevance and measurement quality rather than convenience.

The expanded model was estimated before structural refinement so that the theoretically plausible relationships could first be evaluated empirically. Hedonic motivation was unsupported. The performance-expectancy measures showed substantial discriminant overlap with effort expectancy, behavioral intention, and use behavior, while one learning-oriented performance-expectancy path showed nominal significance. Because these effects could not be interpreted as sufficiently distinct construct contributions, performance expectancy was not retained in the final specification. The refinement decision was therefore based on a combination of theoretical relevance, measurement quality, discriminant validity, and parsimony rather than on statistical significance alone. Figure 1 presents the framework examined before refinement. The retained model included EE, FC, and SE as predictors of BI, with BI predicting UB.

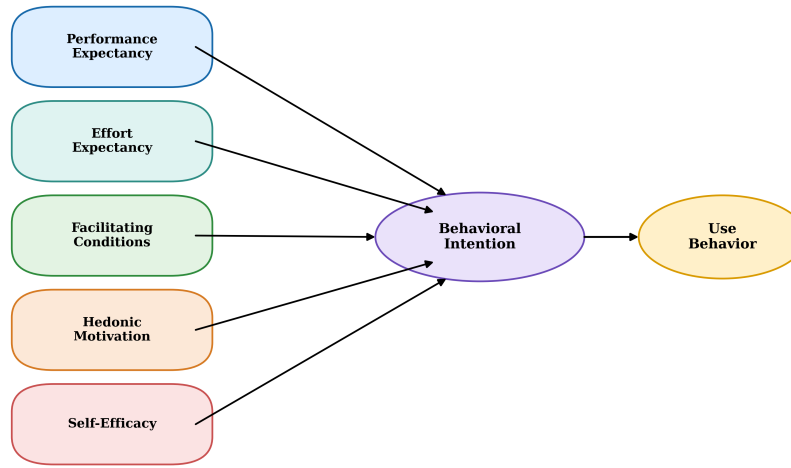


Fig. 1. Conceptual adapted UTAUT2 framework examined before empirical model refinement

3.2 Learning Context

Participants were drawn from two senior-level quantitative STEM courses at a private university in Egypt: Time Series Analysis and Categorical Data Analysis. Most were Data Science or Actuarial Science students with prior coursework in statistics, mathematics, and programming. Both courses combine statistical reasoning, software implementation, diagnostics, evidence evaluation, and written interpretation, making them suitable for examining AI adoption where computational support and disciplinary judgment are intertwined.

Time Series Analysis uses R for autoregressive and moving-average models, ARIMA-type models, diagnostics, forecasting, and volatility modeling. Categorical Data Analysis covers models for discrete outcomes, including logistic regression, Poisson regression, and loglinear models, with applied work completed in R or Python. In both courses, students must choose appropriate models, justify assumptions, interpret coefficients and diagnostics, and communicate results.

AI use was therefore consequential rather than peripheral: students could seek help with concepts, code, diagnostic plots, model choice, or interpretation. Assistance may remove procedural barriers, whereas faulty code or polished but incorrect explanations can disrupt reasoning. Table 1 summarizes the course activities most closely connected to AI use.

Table 1. Course context and AI-relevant learning activities.

Course	Program context	Core statistical work	AI-relevant learning activities
Time Series Analysis	Senior-level statistics / actuarial science course	R-based modeling and forecasting; AR/MA/ARIMA-type models; diagnostics; forecasting; volatility models	Coding assistance; interpreting model output; checking diagnostics; explaining forecasts
Categorical Data Analysis	Senior-level data science / statistics course	Models for discrete outcomes; logistic regression; Poisson regression; loglinear models; applied projects in R or Python	Debugging R/Python code; interpreting regression output; explaining assumptions; drafting analysis narratives

3.3 Hypotheses

The six hypotheses below were derived from the adapted UTAUT2 framework and prior higher-education adoption research (Venkatesh et al., 2012; Xue et al., 2024). They represent the relationships tested in the expanded model before refinement. Performance expectancy was operationalized through the PE_Learn and PE_Tech dimensions described above.

H1. Performance expectancy, represented by learning-oriented and technical/task-oriented dimensions, positively influences behavioral intention to use generative AI.

H2. Effort expectancy positively influences behavioral intention to use generative AI.

H3. Facilitating conditions positively influence behavioral intention to use generative AI.

H4. Hedonic motivation positively influences behavioral intention to use generative AI.

H5. Self-efficacy positively influences behavioral intention to use generative AI.

H6. Behavioral intention positively influences use behavior.

3.4 Sample and Participants

Data were pooled across two comparable senior-level course offerings. Approximately 100 students were eligible overall, and 72 valid responses remained after screening, corresponding to about 70% coverage of the attainable population. Pooling was judged reasonable because the offerings occupied similar program positions and shared comparable learning outcomes, computational demands, and AI-use contexts.

The absolute sample remains modest for structural modeling, so adequacy was not justified by a generic sample-size rule. Instead, the analysis emphasized the parsimony of the retained model, the high coverage of the finite eligible population, 3,000 bootstrap resamples, effect-size estimation, and targeted sensitivity analyses. These procedures assess the stability of the observed relationships but do not remove the limitations of $n = 72$, particularly for detecting small effects or examining subgroup heterogeneity.

Respondents were mainly senior undergraduates with substantial preparation in statistics, mathematics, and programming. No controlled AI intervention was introduced; the survey captured perceptions and self-reported use within normal coursework. The findings therefore describe associations in an authentic educational setting rather than causal effects of a prescribed treatment. Because the subgroup sizes were small, the study did not attempt formal multi-group or measurement-invariance testing across offerings; this remains an important direction for larger replications.

3.5 Questionnaire Development

Survey items were adapted from validated UTAUT2 measures (Venkatesh et al., 2012) and rewritten for advanced statistics coursework. The wording referred to activities familiar to participants, including interpreting output, debugging R or Python code, evaluating analytical results, and clarifying statistical concepts. Closed-ended items used a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

Four open-ended prompts asked how students used AI, what problems they encountered, and what guidance

or restrictions they considered appropriate. These responses were used to illuminate the quantitative results rather than to support a standalone qualitative claim. The full instrument appears in Appendix A.

3.6 Data Analysis

PLS-SEM was estimated in R with the `sempr` package (Ray et al., 2022). Outer loadings were used to assess indicator reliability, that is, how strongly each observed item represented its intended construct. Cronbach’s α , Dijkstra-Henseler’s ρ_A , and composite reliability were used to assess internal consistency, with composite reliability allowing indicator loadings to differ rather than assuming equal contributions. Average variance extracted (AVE) assessed convergent validity by estimating the proportion of indicator variance captured by a construct. The Fornell-Larcker criterion assessed discriminant validity by comparing the square root of each construct’s AVE with its correlations with other constructs, while the heterotrait-monotrait ratio (HTMT) provided a more direct assessment of similarity between constructs (Fornell & Larcker, 1981; Henseler et al., 2015). Variance inflation factors (VIFs) were used to detect problematic collinearity among structural predictors (Hair & Alamer, 2022). Internal-consistency coefficients were not interpreted for the single-item self-efficacy construct.

Structural paths were evaluated with 3,000 bootstrap resamples and 95% percentile confidence intervals, which quantify sampling uncertainty without relying on normal-theory standard errors. The coefficient of determination, R^2 , summarized the proportion of variance explained in each endogenous construct, and f^2 quantified the incremental explanatory contribution of each structural predictor. PLSpredict compared out-of-sample prediction errors with a linear-model benchmark, while Q^2_{predict} summarized predictive relevance, with positive values indicating improvement over the benchmark (Shmueli et al., 2019). Bootstrapped HTMT confidence intervals were also inspected to test inferential discriminant validity; an interval that includes 1 indicates that clear empirical separation between the two constructs cannot be established in the sample.

Refinement was theory-guided. The final model retained effort expectancy, facilitating conditions, and self-efficacy as predictors of behavioral intention, with intention predicting use behavior. Sensitivity analyses removed FC2, UB3, and each BI indicator in turn to evaluate AVE-HTMT trade-offs and item-driven overlap. Alternative specifications were diagnostic checks, not automatic deletion rules; the main model was retained when improvement in one diagnostic came at the cost of construct coverage or explanatory power.

4. Results

4.1 Descriptive Statistics

The 72 respondents reported generally favorable perceptions (Table 2). Facilitating conditions and behavioral intention had the highest means, whereas hedonic motivation had the lowest, indicating that students approached AI primarily as an academic resource rather than an enjoyable activity.

Table 2. Descriptive statistics for raw construct scores.

Construct	n	Mean	SD	Min	Max
PE_Learn	72.00	3.70	0.82	1.00	5.00
PE_Tech	72.00	3.63	0.93	1.00	5.00
EE	72.00	3.84	1.03	1.00	5.00
HM	72.00	3.28	1.10	1.00	5.00
SE	71.00	3.66	1.09	1.00	5.00
FC	72.00	4.27	0.81	2.00	5.00
BI	72.00	3.99	0.88	1.00	5.00
UB	72.00	3.60	0.95	1.00	5.00

Note. Means and standard deviations are computed from raw item averages using the original 1-5 Likert response scale.

4.2 Measurement Model Assessment

The measurement model was assessed for indicator reliability, internal consistency, convergent validity, discriminant validity, and collinearity. Outer loadings indicate how strongly each observed item represents the construct it is intended to measure; higher loadings therefore provide stronger evidence of indicator reliability. Internal consistency indicates whether the items assigned to a construct behave coherently as a scale and was examined using Cronbach's α , ρ_A , and composite reliability. Convergent validity asks whether a construct captures sufficient variance from its own indicators and was assessed using AVE. Discriminant validity asks whether conceptually different constructs are also empirically distinguishable and was examined with both the Fornell-Larcker criterion and HTMT. Finally, VIF assesses whether structural predictors are so strongly correlated that their path estimates may become unstable.

Composite reliability exceeded 0.70 for all multi-item retained constructs (Table 3). Cronbach's α was lower for FC (0.516) and UB (0.599), indicating that those short scales should not be described as uniformly strong on every reliability criterion. Their composite reliabilities were nevertheless acceptable (0.731 and 0.786, respectively), so the reliability evidence was interpreted jointly rather than from a single coefficient. FC was the only construct below the conventional AVE = 0.50 guideline, with AVE = 0.479 and $\rho_C = 0.731$. Removing FC2 raised AVE to 0.583 but worsened FC-BI HTMT from 0.831 to 1.178, while structural estimates changed little. FC2 was therefore retained and the AVE shortfall treated as a limitation. Self-efficacy was represented by a single indicator. Its loading of 1.000 is a consequence of single-item specification and should not be interpreted as evidence of perfect measurement; internal-consistency statistics are not applicable to that construct.

Table 3. Reliability and convergent validity for the final trimmed measurement model.

Construct	Cronbach's α	ρ_A	Composite reliability	AVE
EE	0.823	0.823	0.919	0.850
FC	0.516	0.523	0.731	0.479
SE	Not applicable	Not applicable	Not applicable	Not applicable
BI	0.823	0.826	0.895	0.740
UB	0.599	0.635	0.786	0.554

Most retained loadings were at or near conventional thresholds (Table 4). FC2 and UB3 were weaker but retained for content coverage and interpretability.

Table 4. Outer loadings for the final trimmed measurement model.

Indicator	Construct	Loading
EE1	EE	0.920
EE2	EE	0.923
FC1	FC	0.793
FC2	FC	0.583
FC3	FC	0.684
SE1	SE	1.000
BI1	BI	0.903
BI2	BI	0.861
BI3	BI	0.814
UB1	UB	0.848
UB2	UB	0.714
UB3	UB	0.656

The Fornell-Larcker criterion supports discriminant validity when the square root of a construct's AVE exceeds its correlations with other constructs. By this criterion, the retained constructs showed acceptable separation (Table 5). HTMT provides a stricter check by comparing average cross-construct indicator correlations with within-construct correlations; values approaching or exceeding conventional thresholds

signal possible construct overlap. HTMT identified overlap for EE-BI (0.926) and BI-UB (0.981; Table 6). Bootstrapped HTMT intervals add an inferential check, because an interval that includes 1 means that empirical distinctness cannot be established with confidence. The 95% intervals for EE-BI, [0.793, 1.053], and BI-UB, [0.812, 1.170], both included 1. The constructs therefore cannot be treated as unequivocally distinct in this sample. Because they are theoretically adjacent, the paths were retained, examined through item-level sensitivity analyses, and interpreted cautiously.

Table 5. Fornell-Larcker criterion for the final trimmed model.

Construct	EE	FC	SE	BI	UB
EE	0.922				
FC	0.364	0.692			
SE	0.415	0.340	1.000		
BI	0.763	0.600	0.540	0.860	
UB	0.656	0.362	0.478	0.709	0.744

Table 6. HTMT ratios for the final trimmed model.

Construct	EE	FC	SE	BI	UB
EE					
FC	0.426				
SE	0.458	0.434			
BI	0.926	0.831	0.596		
UB	0.900	0.541	0.597	0.981	

Note. HTMT values above conventional thresholds indicate limited discriminant separation. Bootstrapped intervals for EE-BI and BI-UB included 1; these relationships are therefore interpreted with caution.

4.3 Structural Model Refinement

The expanded candidate model was estimated before the retained model was defined. Hedonic motivation did not show a meaningful direct contribution to BI. Performance expectancy required a more nuanced decision. The learning-oriented PE path showed nominal significance in the expanded model ($p = .035$), but the PE measures were strongly entangled with EE, BI, and UB on discriminant-validity diagnostics. A statistically significant path under those conditions cannot be interpreted confidently as a distinct usefulness effect.

PE was therefore not retained simply because one path crossed a significance threshold. The refinement decision prioritized whether the constructs were empirically distinguishable and theoretically interpretable. This does not imply that perceived usefulness is unimportant to GenAI adoption; rather, its incremental contribution could not be separated cleanly from closely related perceptions of ease and intended use in this sample. The retained model linked EE, FC, and SE to BI, with BI predicting UB.

4.4 Final Structural Model

The coefficient of determination, R^2 , measures the proportion of variance in an endogenous construct explained by its predictors. The final model explained 73.2% of behavioral intention and 50.3% of use behavior (Table 7), indicating strong in-sample explanatory power that must be interpreted alongside the weaker predictive results below.

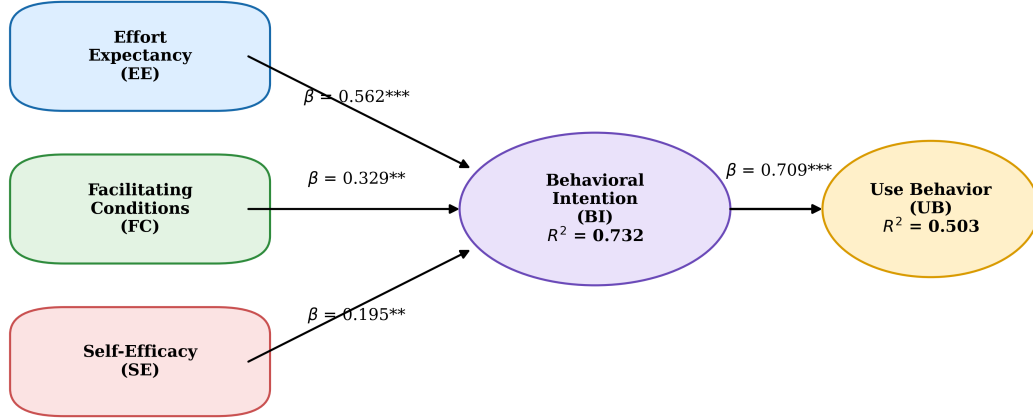


Fig. 2. Final trimmed PLS-SEM model estimated on the expanded sample ($n = 72$)

Table 7. Explanatory power of the final trimmed model.

Endogenous construct	R^2	Adjusted R^2
Behavioral Intention (BI)	0.732	0.720
Use Behavior (UB)	0.503	0.496

The standardized path coefficient, β , indicates the direction and relative strength of each structural relationship, while the bootstrap confidence interval indicates the uncertainty around that estimate. Effort expectancy was the strongest predictor of behavioral intention ($\beta = 0.562$, 95% CI [0.411, 0.682], $p < .001$), followed by facilitating conditions ($\beta = 0.329$, 95% CI [0.127, 0.480], $p = .003$) and self-efficacy ($\beta = 0.195$, 95% CI [0.060, 0.363], $p = .006$). Intention strongly predicted use ($\beta = 0.709$, 95% CI [0.549, 0.833], $p < .001$). The f^2 effect size shows how much each predictor contributes to explained variance beyond the other predictors in the equation. Effect sizes confirmed the same ordering: $EE \rightarrow BI$ was large ($f^2 = 0.905$), $FC \rightarrow BI$ was medium ($f^2 = 0.330$), $SE \rightarrow BI$ was small ($f^2 = 0.112$), and $BI \rightarrow UB$ was large ($f^2 = 1.012$).

Table 8. Bootstrapped structural path estimates for the final trimmed model.

Path	β	Bootstrap SD	t	2.5% CI	97.5% CI	p	f^2
$EE \rightarrow BI$	0.562	0.071	7.953	0.411	0.682	$< .001$	0.905
$FC \rightarrow BI$	0.329	0.088	3.759	0.127	0.480	.003	0.330
$SE \rightarrow BI$	0.195	0.077	2.546	0.060	0.363	.006	0.112
$BI \rightarrow UB$	0.709	0.073	9.735	0.549	0.833	$< .001$	1.012

Note. Two-tailed bootstrap tests based on 3,000 resamples. f^2 values are interpreted descriptively as approximately 0.02 = small, 0.15 = medium, and 0.35 = large.

Figures 3 and 4 summarize the model's explained variance and the relative size of the structural effects.

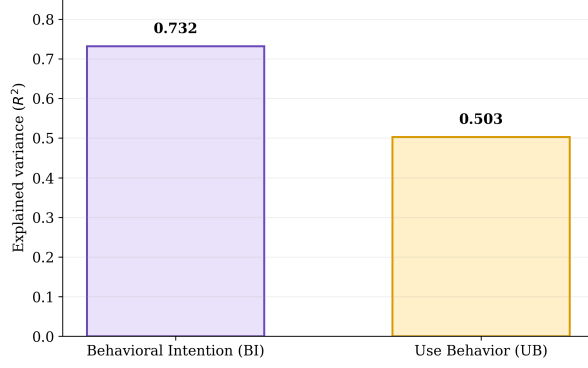


Fig. 3. Explained variance for the endogenous constructs.

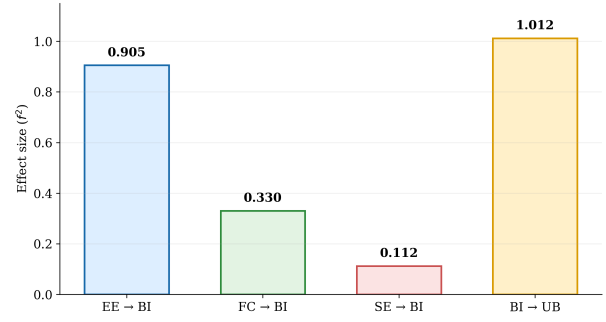


Fig. 4. Effect sizes for the structural paths.

VIF quantifies how much the variance of a structural coefficient is inflated by correlation among predictors; low values indicate that the path estimates are not being destabilized by multicollinearity. VIF values for the predictors of behavioral intention ranged from 1.213 to 1.296 (Table 9), well below the conservative threshold of 3.3.

Table 9. Structural collinearity diagnostics.

Predictor	VIF for BI
EE	1.296
FC	1.213
SE	1.271

4.5 Predictive Assessment

PLSpredict evaluates out-of-sample prediction by comparing the PLS-SEM model's prediction errors with those from a simpler benchmark model; lower RMSE or MAE values indicate better predictive performance. The PLS-SEM model showed slightly lower errors than the linear benchmark for most BI and UB indicators, except UB3 (Table 10). The gains were modest, so predictive evidence was limited rather than strong.

Table 10. PLSpredict prediction errors for behavioral intention and use behavior indicators.

Indicator	Metric	PLS	LM benchmark
BI1	RMSE	0.670	0.693
BI2	RMSE	0.771	0.836
BI3	RMSE	0.724	0.739
UB1	RMSE	0.943	0.970
UB2	RMSE	1.066	1.101
UB3	RMSE	1.325	1.314
BI1	MAE	0.521	0.538
BI2	MAE	0.587	0.620
BI3	MAE	0.557	0.579
UB1	MAE	0.790	0.814
UB2	MAE	0.799	0.815
UB3	MAE	1.104	1.105

Q^2_{predict} summarizes predictive relevance relative to the benchmark, with values above zero indicating that the PLS-SEM predictions improve on the benchmark for the corresponding indicator or construct. Construct-level Q^2_{predict} was 0.086 for BI and 0.033 for UB; all item-level values were positive except UB3 (-0.018; Table 11). Thus, strong in-sample R^2 and f^2 did not translate into strong out-of-sample prediction. The model is therefore interpreted as explanatory for this sample, not as a forecasting model for other student populations.

Table 11. Q^2_{predict} values for item-level and construct-level predictive relevance.

Indicator	Construct	Q^2_{predict}	Level
BI1	BI	0.065	Item
BI2	BI	0.151	Item
BI3	BI	0.041	Item
UB1	UB	0.054	Item
UB2	UB	0.062	Item
UB3	UB	-0.018	Item
BI	BI	0.086	Construct
UB	UB	0.033	Construct

Table 12. Summary of model refinement and supported relationships.

Hypothesis	Path	Decision	Interpretation
H1	Performance expectancy → Behavioral intention	Not retained in final model	Performance expectancy was not retained because the expanded specification showed substantial discriminant overlap with EE/BI/UB; a learning-oriented PE path showed nominal significance, but its distinct structural contribution was not cleanly interpretable.
H2	Effort expectancy → Behavioral intention	Supported	Largest positive effect on BI.
H3	Facilitating conditions → Behavioral intention	Supported	Positive effect on BI; guidance and support matter.
H4	Hedonic motivation → Behavioral intention	Not supported / not retained	Hedonic motivation did not show a meaningful direct contribution to behavioral intention in the expanded model.
H5	Self-efficacy → Behavioral intention	Supported	Smaller but significant positive effect.
H6	Behavioral intention → Use behavior	Supported	Strong positive effect on self-reported use.

4.6 Robustness and Sensitivity Analyses

Sensitivity analyses tested whether conclusions depended on specific measurement choices. Removing FC2 increased FC AVE from 0.479 to 0.583 but worsened FC-BI HTMT from 0.831 to 1.178; structural coefficients changed little and BI R^2 remained high (0.719 versus 0.732). This trade-off supported retaining FC2.

Item-level checks removed UB3 or one BI indicator at a time. Removing UB3, BI1, or BI2 did not resolve the relevant HTMT overlap. Removing BI3 reduced EE-BI HTMT from 0.926 to 0.889 and BI-UB from 0.981 to 0.905, while the structural pattern remained similar ($EE \rightarrow BI = 0.555$; $FC \rightarrow BI = 0.288$; $SE \rightarrow BI = 0.184$; $BI \rightarrow UB = 0.669$). However, R^2 declined to 0.664 for BI and 0.447 for UB, and BI-UB overlap remained slightly above 0.90. The prespecified three-item BI measure was therefore retained, with the discriminant-validity limitation reported rather than removed through post hoc trimming.

Table 13. Targeted measurement sensitivity analyses.

Sensitivity check	Main specification	Alternative specification	Structural stability	Decision
Remove FC2	FC AVE = 0.479; FC-BI HTMT = 0.831	FC AVE = 0.583; FC-BI HTMT = 1.178	$EE \rightarrow BI = 0.547$; $FC \rightarrow BI = 0.310$; $SE \rightarrow BI = 0.206$; $BI \rightarrow UB = 0.709$; $R_{BI}^2 = 0.719$	Retain FC2; report AVE limitation
Remove BI3	EE-BI HTMT = 0.926; BI-UB HTMT = 0.981; $R_{BI}^2/R_{UB}^2 = 0.732/0.503$	EE-BI HTMT = 0.889; BI-UB HTMT = 0.905; $R_{BI}^2/R_{UB}^2 = 0.664/0.447$	$EE \rightarrow BI = 0.555$; $FC \rightarrow BI = 0.288$; $SE \rightarrow BI = 0.184$; $BI \rightarrow UB = 0.669$	Retain BI3; report discriminant-validity limitation

4.7 Descriptive Thematic Analysis of Open-Ended Responses

Open-ended responses were reviewed inductively and grouped by recurring content. Because they were intended to contextualize the survey rather than support a standalone qualitative claim, the analysis is descriptive.

Students valued AI for patient explanation, examples, coding, and organizing written work. One respondent described using it for ‘understanding and simplifying concepts’; another said it helped ‘immensely in coding and summarizing the points’ for project reports.

Students also reported inaccurate answers, wrong calculations, and explanations that conflicted with course material, especially at advanced levels. Dependence was a deeper concern: one respondent called AI ‘addictively convenient’ and warned that it could reduce critical thinking.

Several students supported AI for assignments or checking work but not closed-book examinations, and emphasized disclosure. Overall, they valued immediacy while wanting clearer expectations, verification routines, and safeguards against overreliance. This pattern complements the SEM results: factors that encourage adoption do not by themselves ensure statistical judgment, verification, or self-regulation.

5. Discussion

5.1 Interpretation of the Structural Findings

The retained model provides a deliberately bounded account of GenAI adoption in advanced quantitative STEM coursework. Students were more likely to report intention to use AI when the tool fitted easily into their workflow, when the course environment supported its use, and when they felt capable of using it. Intention was then strongly associated with reported use. These results concern the conditions under which AI enters disciplinary practice; they do not show that AI improves learning or that the observed paths are causal.

Effort expectancy was the strongest incremental contributor to behavioral intention in this sample. Its large effect size, $f^2 = 0.905$, indicates that effort expectancy contributed substantially to the model’s ability to explain variation in behavioral intention. In other words, when effort expectancy was removed from the model, the proportion of variance explained in behavioral intention decreased markedly. This suggests that students’ intention to use GenAI was shaped not only by whether they viewed the technology as useful, but especially by how easy it was to incorporate into their academic workflow.

This finding is particularly meaningful in advanced statistics and data science courses, where students frequently move between conceptual understanding, coding, model fitting, diagnostic checking, and interpretation. In such settings, a tool that can quickly explain an idea, suggest code, identify an error, or help interpret output reduces the effort required to move from one stage of the analytical process to the next. That reduction in effort may make students more willing to adopt GenAI as part of their regular coursework.

At the same time, ease of use has an important educational downside. When AI produces an answer quickly and fluently, students may be tempted to accept the output before checking whether the code is correct, whether model assumptions are satisfied, or whether the interpretation is supported by the evidence.

Thus, low friction may promote adoption while also increasing the risk of uncritical reliance. This tension is consistent with science-education research emphasizing that productive GenAI use requires active evaluation and disciplinary judgment rather than passive acceptance of generated content (Tang & Putra, 2026). Facilitating conditions also contributed meaningfully to BI. In this context, support is not only technical access. It includes knowing whether AI is permitted for a task, how its use should be disclosed, and how generated code, calculations, and interpretations should be checked. This broadens the educational meaning of facilitating conditions and is consistent with work that treats responsible AI integration as a pedagogical and disciplinary issue (Lee et al., 2026). Self-efficacy had a smaller positive association. Confidence may make students more willing to engage with AI, but the open-ended responses show why confidence cannot be equated with sound judgment.

The performance-expectancy result should not be summarized as “usefulness does not matter.” A learning-oriented PE path showed nominal significance in the expanded model, but PE was not sufficiently distinct from EE, BI, and UB to support a clean interpretation of an independent usefulness effect. One plausible explanation is that usefulness was already widely assumed among these experienced students, whereas ease of integration differentiated intention more clearly. That interpretation remains tentative because the measurement overlap itself limits what can be concluded. Hedonic motivation, by contrast, was unsupported and students’ comments portrayed AI primarily as a functional academic resource rather than a source of enjoyment.

5.2 Adoption Versus Responsible Disciplinary Use

The central theoretical implication is that adoption and responsible disciplinary use are related but different outcomes. UTAUT2 helps explain whether GenAI enters students’ workflow. It does not establish whether students subsequently verify code, challenge an answer, test assumptions, interpret uncertainty appropriately, or disclose AI assistance. The open-ended responses make this distinction visible: students valued immediacy and coding support while also reporting wrong calculations, inaccurate explanations, dependence, and uncertainty about acceptable use.

This distinction matters for science and quantitative education because technology uptake is not itself evidence of epistemically productive use. A student can be highly willing to use AI and still fail to evaluate the output. Future acceptance research could therefore place adoption models alongside separate measures of AI literacy, verification behavior, self-regulation, and disciplinary judgment. The study therefore does not propose that UTAUT2 explains responsible AI learning; it identifies conditions associated with AI entering the learning process.

The measurement results reinforce the need for that distinction. EE, BI, and UB were theoretically adjacent and also showed empirical overlap, particularly for EE-BI and BI-UB. The BI3 sensitivity analysis improved HTMT while leaving the directional structural story intact, but it reduced explained variance and did not fully resolve BI-UB overlap. This pattern can reflect both genuine closeness between intention and contemporaneous self-reported use and limitations of the present measures. The study therefore treats the overlap as a substantive caution, not as something to be hidden through post hoc deletion.

5.3 Pedagogical Implications

The strongest pedagogical implication is to make the reasoning around AI use visible, not simply to make AI easier to use. Assignments can require students to compare an AI-generated solution with their own, locate an error, justify a model choice, diagnose an assumption violation, or document how a generated suggestion was verified. Such tasks turn AI output into material for critique rather than an invisible shortcut.

Guidance should also be tied to specific disciplinary tasks. A general statement that AI is “allowed” or “not allowed” offers little help with questions such as whether generated code may be used, whether model interpretations must be independently checked, or how AI assistance should be disclosed. Explicit verification and disclosure routines can therefore be treated as part of the learning environment itself. This is particularly important in quantitative work because incorrect output can be syntactically polished and numerically plausible.

5.4 Limitations and Future Research

The study was conducted at one institution in two advanced quantitative course offerings, so it should be treated as a focused case in statistics, data science, and actuarial science rather than generalized to all STEM students. Although 72 responses represented about 70% of the eligible population, the absolute sample remains modest for structural modeling. Bootstrapping, effect sizes, and sensitivity analyses support stability within this sample but not broad population generalizability. The sample was also too small for meaningful multi-group or measurement-invariance analysis across offerings, which should be addressed in larger replications.

All key quantitative variables were self-reported and measured at one time. The BI-to-UB path therefore represents association, not evidence that intention preceded or caused later behavior. Future studies should combine surveys with AI-use logs, assignment artifacts, longitudinal measurements, or direct performance tasks. The open-ended component was descriptive rather than a standalone qualitative study, so future work could use interviews or focus groups to examine how students actually verify AI-generated analyses and how institutional rules shape those practices.

Measurement limitations also remain important. SE was represented by a single item, FC had $AVE = 0.479$, Cronbach's α was below 0.70 for FC and UB, and bootstrapped HTMT intervals for EE-BI [0.793, 1.053] and BI-UB [0.812, 1.170] included 1. Item-deletion checks improved some point estimates but did not justify post hoc restructuring without losses in content coverage or explanatory power. Finally, predictive relevance was weak ($Q^2_{\text{predict}} = 0.086$ for BI and 0.033 for UB), reinforcing that the present model should be read primarily as explanatory for this context rather than as a forecasting model for other student populations.

6. Conclusion

In this sample, GenAI adoption was associated most strongly with low-friction integration, followed by supportive conditions and confidence in using the tool. EE, FC, and SE were positively associated with behavioral intention, and intention was strongly associated with self-reported use. Effect-size and sensitivity analyses showed that this structural pattern was reasonably stable across plausible measurement alternatives, although the measurement and predictive limitations remain important.

The broader contribution is the distinction between uptake and responsible disciplinary use. Making AI easy to adopt may help it enter students' quantitative workflow, but responsible use requires a different set of practices: checking code and calculations, evaluating assumptions and evidence, interpreting uncertainty, and disclosing assistance transparently. For advanced statistics and data science education, the instructional challenge is therefore not simply whether students use GenAI, but how they reason with it once they do.

Future research should test this distinction directly with larger multi-institution samples, stronger multi-item measures, longitudinal designs, and observable indicators of verification and disciplinary reasoning. Such work can determine whether the factors that encourage adoption also support, or sometimes undermine, high-quality quantitative learning.

Declarations

Funding

No external funding was received for this research.

Ethical approval

The study was reviewed and approved under the expedited category by the Institutional Review Board (IRB) of The American University in Cairo, CASE #2026-54, on March 31, 2026. All procedures involving human participants were performed in accordance with relevant institutional guidelines and regulations and with the ethical principles of the Declaration of Helsinki.

Consent to participate

Participation was voluntary. Informed consent was obtained from all participants by their proceeding with and completing the online survey after being provided with the study information; no separate signed consent form was used.

Consent to publish

Not applicable.

Availability of data and materials

The de-identified data supporting the findings of this study are not included in the manuscript or supplementary files and are not publicly available because they were collected from human participants under institutional ethical oversight. Requests for access may be directed to the corresponding author and will be considered subject to applicable ethical and institutional requirements.

Use of Generative AI in Manuscript Preparation

The authors used generative AI tools, including ChatGPT, for language refinement, readability, and formatting assistance during manuscript preparation. The tools were not used to generate or alter the study data or statistical results. All AI-assisted text was critically reviewed, checked against the original analyses and source literature, and substantively edited by the authors, who take full responsibility for the final content, interpretations, and conclusions.

References

- Ahmed, M., Mohd Zaid, N., & Abdullah, A. H. (2025). The impact of integration of AI in statistics learning on student motivation and learning engagement in higher education sector. *International Journal of Academic Research in Progressive Education and Development*, 14(1), 426–438. <https://doi.org/10.6007/IJARPED/v14-i1/24353>
- Al Labadi, L., & Ly, A. (2025). Enhancing statistics education through project-based learning (PBL) and the emergence of ChatGPT. *Teaching Statistics*, 47, 200–218. <https://doi.org/10.1111/test.12405>
- Bien, J., & Mukherjee, G. (2025). Generative AI for data science 101: Coding without learning to code. *Journal of Statistics and Data Science Education*, 33(2), 129–142. <https://doi.org/10.1080/26939169.2024.2432397>
- Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, 66(4), 616–630. <https://doi.org/10.1007/s11528-022-00715-y>
- Cheung, K. K. C., & Zhang, W. (2025). Undergraduate students' perceptions of the nature of generative artificial intelligence (GenAI)-influenced science: A structural equation modelling approach. *Journal of Science Education and Technology*, 34, 1495–1507. <https://doi.org/10.1007/s10956-025-10230-3>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- DeLuca, L. S., Reinhart, A., Weinberg, G., Laudénbach, M., Miller, S., & Brown, D. W. (2025). Developing students' statistical expertise through writing in the age of AI. *Journal of Statistics and Data Science Education*, 33(3), 266–278. <https://doi.org/10.1080/26939169.2025.2497547>
- Ellis, A. R., & Slade, E. (2023). A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education*, 31(2), 128–133. <https://doi.org/10.1080/26939169.2023.2223609>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.2307/3151312>
- Hair, J., & Alamer, A. (2022). Partial least squares structural equation modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research Methods in Applied Linguistics*, 1(3), Article 100027. <https://doi.org/10.1016/j.rmal.2022.100027>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-3403-8>
- Lee, G., Yun, M., Zhai, X., & Crippen, K. (2026). Artificial intelligence in science education research: Current states and challenges. *Journal of Science Education and Technology*, 35, 110–127. <https://doi.org/10.1007/s10956-025-10239-8>

- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Namatovu, A., & Kyambade, M. (2025). Leveraging AI in academia: University students' adoption of ChatGPT for writing coursework (take home) assignments through the lens of UTAUT2. *Cogent Education*, 12(1), Article 2485522. <https://doi.org/10.1080/2331186X.2025.2485522>
- Nolan, D., & Temple Lang, D. (2010). Computing in the statistics curricula. *The American Statistician*, 64(2), 97–107. <https://doi.org/10.1198/tast.2010.09132>
- Nouredine, R., Boote, D., & Campbell, L. O. (2025). Assessing the validity of UTAUT among higher education instructors: A meta-analysis. *Education and Information Technologies*, 30, 16687–16719. <https://doi.org/10.1007/s10639-025-13449-0>
- Rahi, B., Mishra, S. K., Gupta, S., & Krishan, R. (2025). Exploring ChatGPT adoption among teacher educators in higher education: An extended UTAUT2 approach. In R. González Vallejo, G. Moukhliiss, E. Schaeffer, & V. Paliktzoglou (Eds.), *International symposium on generative AI and education (Lecture Notes on Data Engineering and Communications Technologies, Vol. 262, pp. 55–65)*. Springer. https://doi.org/10.1007/978-3-031-98476-1_5
- Ramnarain, U., Ogegbo, A. A., Penn, M., Ojetunde, S., & Mdlalose, N. (2025). Pre-service science teachers' intention to use generative artificial intelligence in inquiry-based teaching. *Journal of Science Education and Technology*, 34, 1272–1285. <https://doi.org/10.1007/s10956-024-10159-z>
- Rao, V. N. V., Jeng, A., Drew, J., Kala, B., & Gongati, S. (2026). Students' Statistical Thinking When Using Generative AI: A Descriptive Case Study. *Journal of Statistics and Data Science Education*, 34(1), 4–15. <https://doi.org/10.1080/26939169.2025.2596684>
- Ray, S., Danks, N. P., & Calero Valdez, A. (2022). *seminr: Building and estimating structural equation models [Computer software]*. <https://CRAN.R-project.org/package=seminr>
- Sergeeva, O. V., Zheltukhina, M. R., Shoustikova, T., Tukhvatullina, L. R., Dobrokhoto, D. A., & Kondrashev, S. V. (2025). Understanding higher education students' adoption of generative AI technologies: An empirical investigation using UTAUT2. *Contemporary Educational Technology*, 17(2), Article ep571. <https://doi.org/10.30935/cedtech/16039>
- Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*, 53(11), 2322–2347. <https://doi.org/10.1108/EJM-02-2019-0189>
- Tang, K.-S., & Putra, G. B. S. (2026). Generative AI as a dialogic partner: Enhancing multiple perspectives, reasoning, and argumentation in science education with customized chatbots. *Journal of Science Education and Technology*, 35, 128–140. <https://doi.org/10.1007/s10956-025-10240-1>
- Tamilmani, K., Rana, N. P., Wamba, S. F., & Dwivedi, R. (2021). The extended unified theory of acceptance and use of technology (UTAUT2): A systematic literature review and theory evaluation. *International Journal of Information Management*, 57, Article 102269. <https://doi.org/10.1016/j.ijinfomgt.2020.102269>
- Tolba, E. G., & Youssef, N. H. (2022). High school science teachers' acceptance of using distance education in the light of UTAUT. *Eurasia Journal of Mathematics, Science and Technology Education*, 18(9), Article em2152. <https://doi.org/10.29333/ejmste/12365>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178. <https://doi.org/10.2307/41410412>
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for Information Systems*, 17(5), 328–376. <https://doi.org/10.17705/1jais.00428>
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–248. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>

Xue, L., Ghazali, N., & Mahat, J. (2026). A systematic review of UTAUT and UTAUT2 for AI adoption in education. *International Journal of Human–Computer Interaction*, 42(8), 5666–5690. <https://doi.org/10.1080/10447318.2025.2552867>

Xue, L., Rashid, A. M., & Ouyang, S. (2024). The unified theory of acceptance and use of technology (UTAUT) in higher education: A systematic review. *SAGE Open*, 14(1). <https://doi.org/10.1177/21582440241229570>

Appendix A. Survey Constructs and Items

All closed-ended items were measured on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Items were adapted from UTAUT2 and contextualized for generative AI use in advanced statistics coursework.

Table A1. Survey constructs and items.

Item	Wording
Performance Expectancy	
PE1	AI helps me understand complicated statistical ideas.
PE2	ChatGPT or other AI tools explain mathematical steps clearly.
PE3	AI gives examples that make hard topics easier to learn.
PE4	AI helps me understand results from software like R or Python.
PE5	AI is good at pointing out what assumptions models need.
PE6	AI helps me fix mistakes in my code.
PE7	AI is better than textbooks for explaining some statistics topics.
PE8	AI can help students and instructors work together better.
Effort Expectancy	
EE1	I find it easy to use AI tools for my statistics coursework.
EE2	AI tools are easy to include in my studying or homework routine.
Hedonic Motivation	
HM1	I enjoy using AI tools when I study statistics.
HM2	I believe AI makes learning statistics more interesting.
HM3	Using AI tools motivates me to study harder.
Self-Efficacy	
SE1	I feel confident using AI for tasks like analysis or writing code.
Facilitating Conditions	
FC1	When instructors clearly allow AI use, I feel more comfortable and confident using it.
FC2	Our university should teach us how to use AI tools properly.
FC3	AI tools should be included in course materials or lectures.
FC4	AI will be important for statistics students in the next few years.
FC5	Even with AI, it is important to fully understand the underlying statistics.
Behavioral Intention	
BI1	I plan to keep using AI tools in my statistics courses.
BI2	I would recommend AI tools to my classmates.
BI3	I think AI will help me with future coursework or my graduation project.
Use Behavior	
UB1	I often use AI tools while doing my statistics homework.
UB2	I use AI when I do not understand something from class.
UB3	I use AI to check my work before I submit it.
Ethical/Risk Concerns	
EC1	Sometimes AI gives wrong or confusing answers in statistics.
EC2	AI-generated code often needs corrections before I can use it.
EC3	AI tools are not great at very advanced or unusual statistics problems.
Academic Integrity and Policy	
POL1	AI should be allowed for homework, but not during closed-book exams.
POL2	If I use AI in my work, I should say so and cite it.
POL3	I use AI to check my work but not to do everything for me.
POL4	Students should say in their assignment if they used AI to help.

Note. Some items were not retained in the final trimmed model after measurement and structural diagnostics.