

Behavioral Intention to Use Metaverse-Based Learning Among Medical and Paramedical Students in Iran: A Cross-Sectional Study

Seyyede Fatemeh Mousavi Baigi

Mashhad University of Medical Sciences

Masoumeh Sarbaz

Mashhad University of Medical Sciences

Khalil Kimiafar

khalil.kimiafar@gmail.com

Mashhad University of Medical Sciences

Article

Keywords: Metaverse, behavioral intention, medical education, paramedical students, technology acceptance model, PLS-SEM

Posted Date: September 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10475880/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Behavioral Intention to Use Metaverse-Based Learning Among Medical and Paramedical Students in Iran: A Cross-Sectional Study

Seyyede Fatemeh Mousavi Baigi^{1,2,3,*}, Masoumeh Sarbaz¹, Khalil Kimiafar^{1*}

1. Department of Health Information Technology, School of Paramedical and Rehabilitation Sciences, Mashhad University of Medical Sciences, Mashhad, Iran
2. Student Research Committee, Mashhad University of Medical Sciences, Mashhad, Iran
3. Student Committee of Medical Education Development, Education Development Center, Mashhad University of Medical Sciences, Mashhad, Iran

*** Correspondence:**

Khalil Kimiafar, Professor, Department of Health Information Technology, School of Paramedical and Rehabilitation Sciences, Mashhad University of Medical Sciences, Mashhad, Iran. E-mail: Khalil.kimiafar@gmail.com;

Seyyede Fatemeh Mousavi Baigi, PhD Candidate, Department of Health Information Technology, School of Paramedical and Rehabilitation Sciences, Mashhad University of Medical Sciences, Mashhad, Iran. E-mail: Mousavi9713@gmail.com

ORCID ID:

Seyyede Fatemeh Mousavi Baigi: 0000-0002-2214-0077

Masoumeh Sarbaz: 0000-0001-5456-8505

Khalil Kimiafar: 0000-0003-0351-4675

Running title: Intention to Use Metaverse-Based Learning

Abstract

Background

Evidence on students' acceptance of metaverse-based learning in health professions education remains limited, particularly in middle-income and non-Western educational settings. This study examined behavioral intention to use metaverse-based learning among medical and paramedical students in Iran using an extended Technology Acceptance Model.

Methods

A cross-sectional survey was conducted at Mashhad University of Medical Sciences from June 24 to July 24, 2026. Using proportionate stratified random sampling, 185 students were recruited from the Faculties of Medicine and Paramedical Sciences. The primary analysis used a context-adapted 25-item model excluding one non-equivalent perceived-usefulness item; the 26-item model was retained for sensitivity analysis. Reflective Mode-A partial least squares structural equation modelling, 5,000-sample bootstrapping, repeated 10-fold cross-validation, and measurement-invariance testing were performed.

Results

The response rate was 96.9%. Mean behavioral intention was 3.89 (SD 0.73) on a five-point scale. The model explained 70.1% of its variance. Personal innovativeness in information technology ($\beta=0.419$, 95% CI 0.249–0.591) and perceived enjoyment ($\beta=0.388$, 95% CI 0.211–0.552) showed direct associations with intention. Perceived ease of use had no direct association but showed a positive indirect effect, mainly through perceived enjoyment ($\beta=0.225$, 95% CI 0.128–0.328). Direct effects of perceived usefulness and perceived cyber risk were unsupported. The extended model outperformed the base model in out-of-sample prediction (RMSE 0.447 vs 0.525; Q^2_{predict} 0.627 vs 0.484; permutation $p=0.002$). Paramedical students reported higher behavioral intention than medical students (mean difference 0.392, 95% CI 0.183–0.600). However, personal innovativeness and cyber-risk measures showed poor psychometric performance, and most between-faculty path comparisons lacked compositional invariance. The 26-item sensitivity analysis produced similar conclusions.

Conclusions

Students showed favorable intentions toward metaverse-based learning. Perceived enjoyment was the clearest actionable determinant, whereas ease of use operated mainly indirectly. Implementation should be staged, curriculum-aligned, usable, and equitable. Multicenter and longitudinal studies should refine the Persian measures and evaluate use, learning outcomes, accessibility, and cost.

Keywords: Metaverse; behavioral intention; medical education; paramedical students; technology acceptance model; PLS-SEM

Introduction

Digital transformation is reshaping health professions education by expanding the range of tools available for simulation, collaboration, assessment, and experiential learning. Within this evolving landscape, the metaverse has emerged as an interconnected digital environment in which users may interact through avatars and immersive technologies, including virtual, augmented, and mixed reality. In educational settings, metaverse-based platforms may extend conventional online learning by providing a stronger sense of presence, interaction, and shared participation in three-dimensional environments [1–4]. Students can potentially attend virtual classes, interact with instructors and peers, manipulate digital objects, and participate in simulated learning activities without being physically present in the same location [1,3,4].

The potential relevance of the metaverse may be particularly important in medical and health professions education. Health sciences curricula require learners to integrate theoretical knowledge with clinical reasoning, procedural skills, teamwork, and communication. However, opportunities for repeated clinical practice may be constrained by patient-safety considerations, limited access to clinical cases, geographical barriers, and the availability of instructors or simulation facilities. Metaverse-based environments may provide controlled and repeatable settings in which students can observe clinical scenarios, practice selected skills, collaborate with other learners, and receive structured feedback. Recent work has demonstrated the feasibility of teaching and assessing medical students in immersive multi-user environments, highlighting the potential of these platforms to support interactive learning and formative assessment [5]. Nevertheless, technical feasibility alone does not ensure that students will accept or consistently use such platforms.

Evidence from higher education suggests that students are often interested in the educational potential of the metaverse but may have limited previous experience with it. Talan and Kalinkara found that students generally perceived the metaverse as potentially useful for education, while also expressing reservations about its suitability for all courses [3]. Similarly, Rojas et al. reported that metaverse environments could enhance interaction and engagement in online learning, although students' perceptions were influenced by usability, technological requirements, and the quality of the learning experience [4]. Among health sciences students, Sezer and Gül identified relatively high awareness of the metaverse and positive attitudes toward digital technologies, but actual experience with metaverse applications remained limited [6]. These findings suggest that educational adoption depends not only on access to technology but also on students' beliefs, confidence, expectations, and perceived risks.

The Technology Acceptance Model provides a widely used theoretical basis for examining these factors. The original model proposes that perceived usefulness and perceived ease of use are central determinants of individuals' attitudes and intentions toward information technology [2]. Perceived usefulness refers to the extent to which a person believes that using a technology will improve performance, whereas perceived ease of use reflects the extent to which its use is expected to require minimal effort [2]. However, the novelty and immersive characteristics of metaverse-

based learning suggest that these two constructs may not fully explain students' adoption decisions. Personal innovativeness, self-efficacy, perceived enjoyment, and concerns about cyber risk may also influence whether students are willing to engage with such environments.

Al-Adwan et al. therefore extended the Technology Acceptance Model by incorporating personal, technological, and inhibiting factors into a model of students' intention to use metaverse-based learning platforms [1]. Their findings showed that perceived usefulness, personal innovativeness in information technology, and perceived enjoyment were important positive predictors of behavioral intention, whereas perceived cyber risk acted as an inhibitor. They also found that the direct effect of perceived ease of use on behavioral intention was not statistically significant, suggesting that ease of use may operate indirectly through other perceptions [1]. This extended framework is particularly relevant to health professions education, where the perceived educational value of immersive technologies must be considered alongside students' digital confidence, willingness to experiment with emerging technologies, and concerns about privacy and security.

Studies involving healthcare learners and professionals have generally reported favorable but heterogeneous perceptions of the metaverse. Health sciences students have demonstrated positive attitudes toward digital technologies and relatively high metaverse awareness [6]. Nurses have also expressed optimism regarding potential applications of the metaverse in patient education and virtual nursing, although their awareness and attitudes varied according to individual characteristics and future-oriented perspectives [7]. Among healthcare professionals, innovative work behavior has been positively associated with metaverse knowledge and awareness, indicating that openness to innovation may play an important role in the adoption of emerging digital environments [8]. However, much of the available research has focused on awareness, knowledge, or general attitudes within a single student or professional group. Comparatively less is known about how the determinants of behavioral intention operate across medical and paramedical students studying within the same educational and cultural context.

Medical and paramedical students may differ in their educational experiences, clinical exposure, digital learning requirements, and expectations regarding the practical usefulness of emerging technologies. Such differences may influence their perceptions of the usefulness, enjoyment, ease of use, security, and educational relevance of metaverse-based learning. Understanding both shared and group-specific patterns is important for universities seeking to develop inclusive digital education strategies rather than implementing a uniform technology model across heterogeneous health-professions programs.

Therefore, this study aimed to investigate behavioral intention to use metaverse-based learning among medical and paramedical students in Iran. Using the extended Technology Acceptance Model proposed by Al-Adwan et al. [1], the study examined the relationships of perceived usefulness, perceived ease of use, self-efficacy, perceived enjoyment, perceived cyber risk, and personal innovativeness in information technology with students' behavioral intention. It also

compared the observed construct scores of medical and paramedical students and explored potential between-group heterogeneity after assessing measurement invariance. By examining explanatory, predictive, and group-related aspects of metaverse acceptance, this study seeks to provide evidence that can inform the context-sensitive design and implementation of metaverse-based learning in medical and health professions education.

Methods

Study design and setting

This analytical cross-sectional study was conducted among students enrolled in the Faculty of Medicine and the Faculty of Paramedical Sciences at Mashhad University of Medical Sciences, Mashhad, Iran, from June 24 to July 24, 2026. The study was reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for cross-sectional studies [9]. Because data were collected electronically, the design and reporting of the survey were also informed by the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [10].

The Faculty of Medicine group included students enrolled in general medical education at different educational stages, including basic sciences, physiopathology, clinical clerkship, and internship, as well as students enrolled in postgraduate and basic-science programmes administered by the Faculty of Medicine. The Faculty of Paramedical Sciences group included students from radiology, social work, occupational therapy, health information management, laboratory sciences, optometry, physiotherapy, and speech therapy programmes. The study was registered at Mashhad University of Medical Sciences under project tracking code 4042054.

Participants and eligibility criteria

Students were eligible if they were actively enrolled in one of the participating faculties during the study period, were able to read Persian, were included in the corresponding faculty sampling frame, and provided electronic informed consent. Students were excluded if they were no longer actively enrolled, declined participation, did not provide informed consent, or submitted a response that was confirmed to be a duplicate submission by the same individual. Similar response patterns among different students were not regarded as duplicate data.

A questionnaire was considered complete when all mandatory items had been answered and the electronic form had been submitted. Because the study had a cross-sectional design, students who did not respond after two reminders were classified as nonresponders rather than as participants lost to follow-up.

Sample-size calculation

The minimum sample size was estimated separately for the two faculties using Cochran's formula with finite-population correction [11]:

$$n = \frac{NZ^2p(1-p)}{d^2(N-1) + Z^2p(1-p)}$$

In this formula, N represented the size of the source population, $Z = 1.96$ corresponded to a 95% confidence level, $p = 0.50$ was selected to generate the maximum required sample size, and $d = 0.10$ represented the desired absolute precision.

The source populations comprised 913 students in the Faculty of Paramedical Sciences and 2,191 students in the Faculty of Medicine. The calculated minimum sample sizes were approximately 87 students for the Faculty of Paramedical Sciences and 92 students for the Faculty of Medicine. To adopt a more conservative recruitment target, the minimum medical-faculty sample was increased from 92 to 96 students in the study protocol.

To compensate for potential nonresponse, four additional students were invited beyond each faculty-specific target. Consequently, 91 paramedical students and 100 medical students were invited. Of these, 87 paramedical students and 98 medical students completed the questionnaire, corresponding to response rates of 95.6% and 98.0%, respectively. The overall response rate was 96.9%. The final analytical sample therefore consisted of 185 students, including 87 paramedical students and 98 medical students. The two valid medical responses above the protocol target of 96 were retained because the estimated sample size represented a minimum requirement rather than a maximum allowable sample.

Sampling procedure

Proportionate stratified random sampling was used. Academic programmes in the Faculty of Paramedical Sciences and educational stages or programmes in the Faculty of Medicine were treated as sampling strata. The required sample within each stratum was determined using proportional allocation:

$$n_h = \frac{N_h}{N} \times n$$

In this equation, N_h represented the number of eligible students in stratum h , N represented the corresponding faculty population, and n represented the faculty-specific target sample. When proportional allocation produced fractional values, rounding was reconciled so that the sum of the stratum-specific allocations equalled the predefined sample target for each faculty. Within each stratum, students were selected by simple random sampling from the relevant student roster.

Selected students received an electronic invitation containing information about the study, an electronic consent statement, and a link to the questionnaire. Two reminders were sent to selected students who had not responded. Students who did not submit the questionnaire after the second

reminder were classified as nonresponders. The required sample size was achieved through electronic data collection, and supplementary paper-based recruitment was not required.

Study instrument

Behavioral intention to use metaverse-based learning was assessed using the questionnaire developed by Al-Adwan et al. within an extended Technology Acceptance Model [1]. The original instrument comprised 26 items measuring seven reflective constructs. Perceived usefulness and perceived ease of use were each measured using four items, self-efficacy was measured using three items, perceived enjoyment was measured using four items, perceived cyber risk was measured using four items, personal innovativeness in information technology was measured using four items, and behavioral intention was measured using three items.

Responses were recorded on a five-point Likert scale ranging from 1, strongly disagree, to 5, strongly agree. The questionnaire also collected demographic and academic information, including age, gender, faculty, degree level, recorded academic programme or educational stage, semester of entry, and year of entry.

During the final content audit, one perceived-usefulness item, PU4, was found to retain the source-domain wording referring to “higher tourism education.” In the original item, this phrase referred to university-level education in the field of tourism and was not contextually equivalent to medical or health professions education. A post hoc, content-driven decision was therefore made to exclude PU4 from the context-adapted primary measurement model. The remaining three perceived-usefulness indicators retained the core conceptual content of the construct. No other item was removed from the primary model solely to improve reliability, model fit, or statistical significance.

To preserve analytical transparency, the original 26-item model including PU4 was re-estimated as a sensitivity analysis and reported in the supplementary material. Accordingly, the primary context-adapted model comprised 25 items, whereas the complete source-instrument model comprised 26 items.

Translation and cultural adaptation

The English-language questionnaire was translated and culturally adapted into Persian in accordance with the World Health Organization process for translation and adaptation of instruments and established recommendations for cross-cultural adaptation [12,13]. The adaptation process involved forward translation into Persian, reconciliation of the translated wording and preparation of a consensus Persian version, independent back-translation into English, comparison of the original and back-translated versions, and multidisciplinary expert review. Semantic, idiomatic, experiential, and conceptual equivalence were considered throughout this process.

Face validity was assessed by a multidisciplinary panel with expertise in medical education, medical informatics, health information management, and biostatistics. The panel evaluated the

clarity, readability, simplicity, comprehensibility, and contextual appropriateness of the Persian items.

Content validity was assessed qualitatively by an expert panel. The experts evaluated the relevance, clarity, simplicity, comprehensibility, and contextual appropriateness of each item for medical and paramedical students. Items requiring modification were reviewed and revised through discussion until consensus was reached [14, 15]. The final English and Persian versions of the administered questionnaire, together with the scoring procedure and the analytical status of PU4, are provided in **Table S1**.

Electronic survey administration

The questionnaire was administered using an electronic survey form. All demographic variables required for analysis and all questionnaire items were configured as mandatory; consequently, respondents could not submit an incomplete questionnaire.

The first page of the survey included information about the study objectives, voluntary participation, confidentiality, and the right to decline participation without any effect on academic status or educational services. Participants could access the questionnaire only after providing electronic informed consent.

Mandatory completion prevented item-level missingness among submitted questionnaires, although it did not prevent unit nonresponse among invited students who chose not to participate. Submitted records were screened for duplicate study identifiers, and no confirmed duplicate submissions were identified.

Study variables and scoring

The primary outcome was behavioral intention to use metaverse-based learning, measured using the three BI indicators. The secondary constructs were perceived usefulness, perceived ease of use, self-efficacy, perceived enjoyment, perceived cyber risk, and personal innovativeness in information technology.

Responses were scored from 1 to 5. PCR1 and PCR2 were reverse-scored so that higher PCR scores represented greater perceived cyber risk, while PIIT3 was reverse-scored so that higher PIIT scores represented greater personal innovativeness:

$$X_{\text{reverse}} = 6 - X$$

For descriptive analyses, each construct score was calculated as the arithmetic mean of its corresponding indicators. In the primary context-adapted analysis, the perceived-usefulness score was calculated using PU1, PU2, and PU3. In the sensitivity analysis based on the complete source instrument, the perceived-usefulness score was calculated using PU1 through PU4. Higher scores represented higher levels of the corresponding construct.

No participant was excluded because of an unusual, extreme, or similar response pattern in the absence of independent evidence that the response was invalid or duplicated.

Statistical analysis

All statistical analyses were conducted in Python 3.13.5 using a reproducible analytical workflow. Data management and tabulation were performed using pandas 2.2.3 and NumPy 2.3.5. Inferential statistical procedures, probability distributions, resampling methods, and matrix computations were implemented using SciPy 1.17.0 and statsmodels 0.14.6. Repeated cross-validation and predictive modelling were performed using scikit-learn 1.8.0, and graphical outputs were generated using Matplotlib 3.10.8.

Reflective Mode-A partial least squares path modelling was implemented using reproducible Python routines based on NumPy and SciPy matrix operations. All indicators were standardized before model estimation. Statistical tests were two-sided, and a p -value below 0.05 was considered statistically significant. When families of related hypotheses were examined, Holm's sequential procedure was used to control the family-wise error rate.

Descriptive analysis

Categorical variables were summarized using frequencies and percentages. Continuous variables and construct scores were summarized using means, standard deviations, medians, interquartile ranges, minimum and maximum values, and 95% confidence intervals, as appropriate.

Because the medical and paramedical faculties differed in the sizes of their source populations, supplementary population-weighted descriptive estimates were calculated using source-population sizes of 2,191 medical students and 913 paramedical students. Faculty-specific estimates and the principal PLS-SEM models were based on the observed analytical sample.

Measurement-model assessment

The seven latent constructs were specified as reflective constructs and estimated using Mode-A PLS. The measurement model was assessed by examining indicator loadings, corrected item-total correlations, Cronbach's alpha, composite reliability, average variance extracted, indicator cross-loadings, the Fornell-Larcker criterion, and the heterotrait-monotrait ratio of correlations.

Indicator loadings of approximately 0.708 or greater, reliability coefficients between 0.70 and 0.95, and average variance extracted values of at least 0.50 were considered desirable [16]. HTMT values below 0.85 were interpreted as conservative evidence of discriminant validity, whereas values below 0.90 were considered potentially acceptable for conceptually related constructs [17].

Apart from the content-based exclusion of PU4, no indicator was automatically deleted on the basis of an isolated loading, reliability statistic, or p -value. All measurement-model limitations were reported transparently.

Structural-model assessment

The structural model comprised the 13 hypothesized relationships specified in the extended TAM proposed by Al-Adwan et al. [1]. Predictor collinearity was examined using variance inflation factors. Model explanatory performance was evaluated using R^2 , adjusted R^2 , and f^2 effect sizes.

Standardized path coefficients, specific indirect effects, total indirect effects, and total effects were estimated together with percentile 95% confidence intervals derived from 5,000 participant-level bootstrap samples. A direct or indirect association was regarded as statistically supported when its bootstrap confidence interval did not include zero.

Predictive assessment

Out-of-sample predictive performance was evaluated using repeated 10-fold cross-validation with 10 repetitions. A base Technology Acceptance Model containing perceived ease of use and perceived usefulness was compared with the extended model containing personal innovativeness in information technology, perceived cyber risk, perceived ease of use, perceived usefulness, and perceived enjoyment as predictors of behavioral intention.

Predictive performance was quantified using root mean squared error, mean absolute error, and Q^2_{predict} . Paired permutation tests of squared prediction loss were used to determine whether the extended model produced significantly lower out-of-sample prediction error than the base TAM [18].

Importance–performance map analysis

Importance–performance map analysis was conducted using behavioral intention as the target construct [19]. For each predecessor construct, importance was represented by its total effect on behavioral intention, while performance was transformed to a scale ranging from 0 to 100. Constructs with relatively high total effects and lower performance were considered potential priorities for educational or implementation interventions. Because importance estimates depend on the adequacy of both the measurement and structural models, results involving constructs with measurement limitations were interpreted cautiously.

Measurement invariance and multi-group analysis

Measurement invariance between medical and paramedical students was examined using the three-step Measurement Invariance of Composite Models procedure [20]. The procedure assessed configural invariance, compositional invariance, and equality of composite means and variances. MICOM was implemented using 5,000 permutations while preserving the observed group sizes.

Differences in the 13 structural paths were subsequently evaluated using permutation-based PLS multi-group analysis and independent-group bootstrap confidence intervals. Holm adjustment was applied across the 13 path comparisons. A between-faculty path difference was regarded as suitable for confirmatory interpretation only when compositional invariance had been established for both

the predictor and outcome constructs. Path comparisons that did not meet this prerequisite were reported as exploratory.

Faculty and subgroup comparisons

Observed construct scores were compared between medical and paramedical students using Welch's independent-samples tests because equal variances and group sizes were not assumed. Results were reported using faculty-specific means and standard deviations, mean differences, 95% confidence intervals, Holm-adjusted p-values, and Hedges' g effect sizes [21]. Exploratory programme-level differences within each faculty were assessed using Kruskal–Wallis tests, with Holm adjustment across the seven construct-level omnibus comparisons. Because several programme categories contained small numbers of participants, these analyses were interpreted as descriptive and hypothesis-generating, and heatmaps were used to visualise construct profiles.

Common-method bias diagnostics

Because all primary constructs were measured using a single self-report questionnaire, potential common-method bias was explored using Harman's single-factor test and full-collinearity variance inflation factors [22]. These procedures were interpreted as diagnostic indicators rather than as definitive evidence of the presence or absence of common-method bias.

Sensitivity analyses

The primary context-adapted model comprised 25 items and excluded PU4 on conceptual grounds. Sensitivity analyses included re-estimation of the complete 26-item source model including PU4, comparison of measurement and structural estimates with and without PU4, examination of indicators displaying inconsistent reverse-wording effects, comparison of the base TAM with the extended TAM, and calculation of population-weighted descriptive estimates.

All sensitivity analyses were identified as post hoc and were not used to conceal unfavorable measurement-model findings or to select results solely on the basis of statistical significance.

Ethical considerations

The study protocol was approved by the Research Ethics Committee of Mashhad University of Medical Sciences under approval number **IR.MUMS.FHMPM.REC.1405.101**. The institutional project tracking code was **4042054**.

Electronic informed consent was obtained from all participants before access to the questionnaire was provided. Participation was voluntary, and students were informed that declining participation would have no effect on their academic status, grades, or access to educational services. The analytical data were de-identified and handled confidentially. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Results

Participants and data completeness

A total of 191 students were invited to participate, including 91 students from the Faculty of Paramedical Sciences and 100 from the Faculty of Medicine. Completed questionnaires were obtained from 87 paramedical students and 98 medical students, corresponding to faculty-specific response rates of 95.6% and 98.0%, respectively, and an overall response rate of 96.9%. The final analytical sample comprised 185 students. The mean age was 22.45 years (SD 3.12), and 165 participants (89.2%) were women. All submitted questionnaires contained complete responses to the study items; therefore, no missing-data imputation was required. Participant characteristics overall and by faculty are presented in **Table 1**.

Table 1. Participant characteristics overall and by faculty. Note: Values are presented as (n) (%) unless otherwise indicated.

Characteristic	Category	Overall (N=185)	Paramedical sciences (n=87)	Medicine (n=98)
Age, years	Mean (SD)	22.45 (3.12)	22.72 (3.70)	22.21 (2.48)
Gender	Female	165 (89.2)	77 (88.5)	88 (89.8)
	Male	20 (10.8)	10 (11.5)	10 (10.2)
Degree	Bachelor's, continuous	68 (36.8)	68 (78.2)	0 (0.0)
	Bachelor's, discontinuous	2 (1.1)	2 (2.3)	0 (0.0)
	Master's	12 (6.5)	10 (11.5)	2 (2.0)
	Professional doctorate	100 (54.1)	4 (4.6)	96 (98.0)
	PhD	3 (1.6)	3 (3.4)	0 (0.0)
Entry semester	First semester	151 (81.6)	76 (87.4)	75 (76.5)
	Second semester	34 (18.4)	11 (12.6)	23 (23.5)
Entry year	1400	15 (8.1)	0 (0.0)	15 (15.3)
	1401	21 (11.4)	15 (17.2)	6 (6.1)
	1402	44 (23.8)	37 (42.5)	7 (7.1)
	1403	85 (45.9)	24 (27.6)	61 (62.2)
	1404	20 (10.8)	11 (12.6)	9 (9.2)

Construct scores and faculty comparisons

Overall and faculty-specific construct scores are summarised in **Table 2** and visually compared in **Figure 1**. The mean behavioral-intention score was 3.888 (SD 0.729) in the observed sample. After weighting the faculty-specific estimates to the source populations of 913 paramedical and 2,191 medical students, the estimated overall behavioral-intention score was 3.819 (95% CI 3.722–3.917).

Paramedical students reported higher behavioral intention than medical students (4.096 versus 3.704; mean difference 0.392, 95% CI 0.183–0.600; Holm-adjusted ($p=0.001$); Hedges' ($g=0.554$)). They also reported higher personal innovativeness, self-efficacy, perceived usefulness, and perceived enjoyment, together with lower perceived cyber risk. The largest between-faculty differences were observed for self-efficacy (mean difference 0.487, 95% CI 0.293–0.682; ($g=0.735$)) and perceived usefulness (mean difference 0.535, 95% CI 0.301–0.769; ($g=0.663$)). Perceived ease of use did not differ between the two faculties (mean difference 0.109, 95% CI –0.105 to 0.323; Holm-adjusted ($p=0.317$)). Because the PIIT and PCR measurement blocks did not demonstrate adequate psychometric performance, observed score differences involving these two constructs should be interpreted cautiously.

Table 2. Construct scores and comparisons between paramedical and medical students. Note: Positive mean differences indicate higher scores among paramedical students. Holm correction was applied across the seven construct comparisons.

Construct	Overall mean (SD)	Paramedical mean (SD)	Medical mean (SD)	Mean difference	95% CI	Holm-adjusted (p)	Hedges' (g)
PIIT	3.488 (0.584)	3.649 (0.549)	3.344 (0.578)	0.305	0.141 to 0.469	0.001	0.538
PCR	2.924 (0.463)	2.842 (0.482)	2.997 (0.434)	–0.155	–0.289 to –0.022	0.046	–0.338
SE	3.811 (0.703)	4.069 (0.721)	3.582 (0.602)	0.487	0.293 to 0.682	<0.001	0.735
PEU	3.686 (0.726)	3.744 (0.815)	3.635 (0.637)	0.109	–0.105 to 0.323	0.317	0.150
PU	3.905 (0.845)	4.188 (0.816)	3.653 (0.792)	0.535	0.301 to 0.769	<0.001	0.663
PE	3.982 (0.752)	4.195 (0.821)	3.793 (0.631)	0.402	0.187 to 0.617	0.001	0.551
BI	3.888 (0.729)	4.096 (0.808)	3.704 (0.598)	0.392	0.183 to 0.600	0.001	0.554

Abbreviations: BI, behavioral intention; PCR, perceived cyber risk; PE, perceived enjoyment; PEU, perceived ease of use; PIIT, personal innovativeness in information technology; PU, perceived usefulness; SE, self-efficacy.

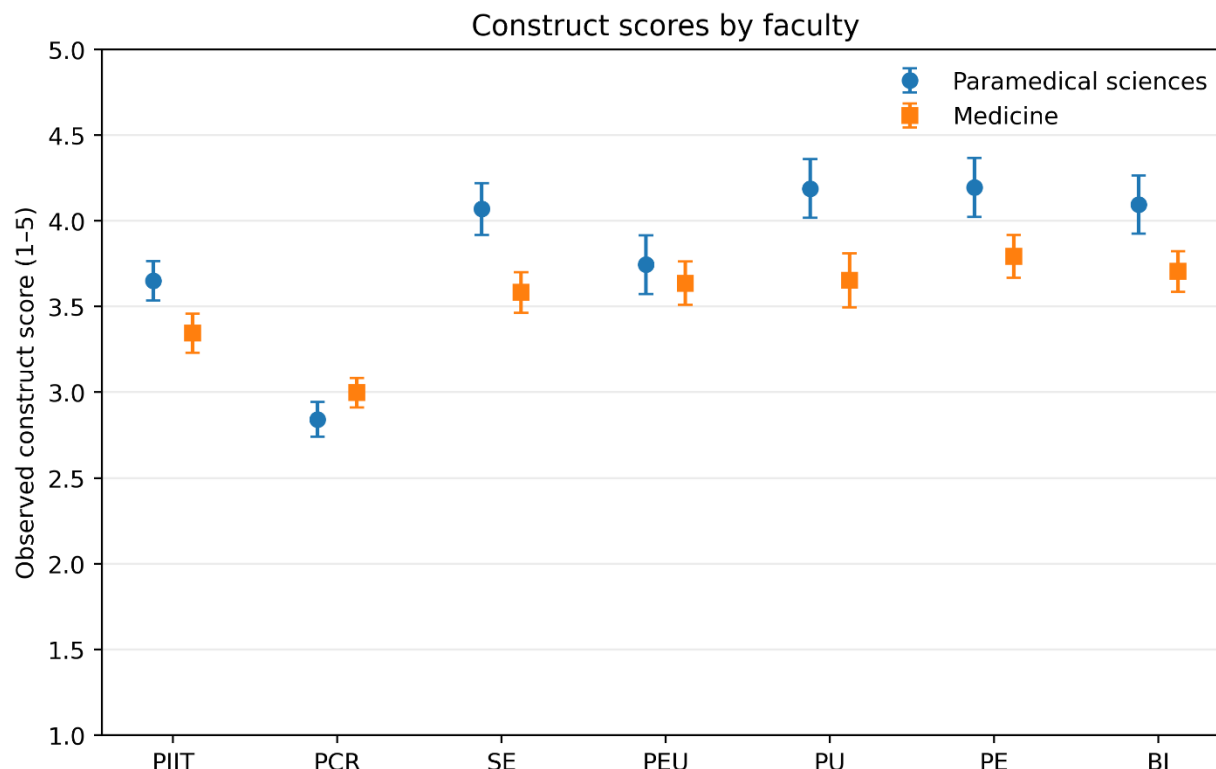


Figure 1. Observed construct scores among students from the Faculties of Paramedical Sciences and Medicine. Points represent faculty-specific means and error bars represent 95% confidence intervals. Construct-score comparisons are descriptive observed-score comparisons and should be interpreted cautiously for PIIT and PCR because of their measurement limitations.

Measurement model

The reliability and convergent-validity results for the primary 25-item model are reported in **Table 3**. Complete item-level measurement diagnostics, including descriptive statistics, corrected item-total correlations, indicator loadings with 95% bootstrap confidence intervals, cross-loadings, HTMT values, and Fornell-Larcker results, are provided in **Table S2**.

The PU, PEU, SE, PE, and BI measurement blocks demonstrated satisfactory internal consistency and convergent validity. Following the context-based exclusion of PU4, the three-item PU block showed high reliability and convergent validity, with Cronbach's ($\alpha=0.925$), composite reliability of 0.954, AVE of 0.875, and indicator loadings ranging from 0.903 to 0.952.

In contrast, the complete PIIT and PCR blocks did not meet conventional measurement criteria. PIIT had Cronbach's ($\alpha=0.493$) and composite reliability of 0.725, and its reverse-worded PIIT3 indicator had a negative loading of -0.246 . PCR had Cronbach's ($\alpha=0.035$) and composite reliability of 0.273, with negative loadings for PCR3 and PCR4 of -0.562 and -0.381 , respectively. The maximum HTMT value was 0.950 for the PIIT-BI association, while the PIIT-SE and PIIT-PEU values also exceeded the conservative threshold of 0.85. These findings indicate

that the translated measurement model did not achieve complete psychometric adequacy. Consequently, structural findings involving PIIT and PCR were interpreted cautiously and were not treated as fully confirmatory.

Table 3. Reliability and convergent validity of the primary 25-item measurement model.

Note: PU4 was excluded from the primary model because its wording referred to higher tourism education. PIIT and PCR did not meet conventional reliability criteria. AVE indicates average variance extracted.

Construct	Number of items	Cronbach's (alpha)	Composite reliability	AVE	Loading range
PIIT	4	0.493	0.725	0.528	−0.246 to 0.858
PCR	4	0.035	0.273	0.518	−0.562 to 0.903
SE	3	0.844	0.907	0.765	0.861 to 0.893
PEU	4	0.886	0.921	0.745	0.823 to 0.894
PU	3	0.925	0.954	0.875	0.903 to 0.952
PE	4	0.937	0.957	0.848	0.872 to 0.950
BI	3	0.850	0.919	0.791	0.860 to 0.926

Structural model

The complete set of standardised structural path coefficients, bootstrap confidence intervals, and hypothesis decisions is reported in **Table 4** and visualised in **Figure 2**. The primary model explained 70.1% of the variance in behavioral intention ($R^2=0.701$); adjusted ($R^2=0.692$)). Inner VIF values ranged from 1.000 to 2.834, indicating no critical predictor collinearity. The largest effect sizes were observed for PIIT→SE ($f^2 = 0.735$), PEU→PE ($f^2 = 0.430$), PIIT→PEU ($f^2 = 0.326$), PIIT→BI ($f^2 = 0.225$), and PE→BI ($f^2 = 0.177$).

Personal innovativeness had a positive direct association with behavioral intention ($\beta=0.419$, 95% CI 0.249–0.591), as did perceived enjoyment ($\beta=0.388$, 95% CI 0.211–0.552). In contrast, the direct associations of perceived cyber risk ($\beta=-0.013$, 95% CI −0.107 to 0.071), perceived ease of use ($\beta=-0.010$, 95% CI −0.130 to 0.138), and perceived usefulness ($\beta=0.130$, 95% CI −0.005 to 0.258) with behavioral intention were not statistically supported.

Personal innovativeness positively predicted self-efficacy ($\beta=0.651$), perceived ease of use ($\beta=0.536$), and perceived usefulness ($\beta=0.279$). Self-efficacy was positively associated with both perceived ease of use ($\beta=0.225$) and perceived usefulness ($\beta=0.230$). Perceived ease of use strongly predicted perceived enjoyment ($\beta=0.548$, 95% CI 0.425–0.669). Overall, H1–H6, H10, and H13 were supported, whereas H7–H9 and H11–H12 were not supported.

Table 4. Structural paths in the primary 25-item extended Technology Acceptance Model.

Note: Confidence intervals and empirical two-sided (p)-values were obtained from 5,000 participant-level bootstrap samples.

Hypothesis	Structural path	(\beta)	95% bootstrap CI	(p)	Decision
H1	PIIT→SE	0.651	0.540 to 0.744	<0.001	Supported
H2	PIIT→PU	0.279	0.112 to 0.452	<0.001	Supported
H3	PIIT→PEU	0.536	0.397 to 0.665	<0.001	Supported
H4	PIIT→BI	0.419	0.249 to 0.591	<0.001	Supported
H5	SE→PU	0.230	0.016 to 0.443	0.035	Supported
H6	SE→PEU	0.225	0.054 to 0.398	0.009	Supported
H7	PCR→PU	−0.144	−0.329 to 0.027	0.102	Not supported
H8	PCR→BI	−0.013	−0.107 to 0.071	0.747	Not supported
H9	PEU→PU	0.098	−0.081 to 0.291	0.271	Not supported
H10	PEU→PE	0.548	0.425 to 0.669	<0.001	Supported
H11	PEU→BI	−0.010	−0.130 to 0.138	0.898	Not supported
H12	PU→BI	0.130	−0.005 to 0.258	0.057	Not supported
H13	PE→BI	0.388	0.211 to 0.552	<0.001	Supported

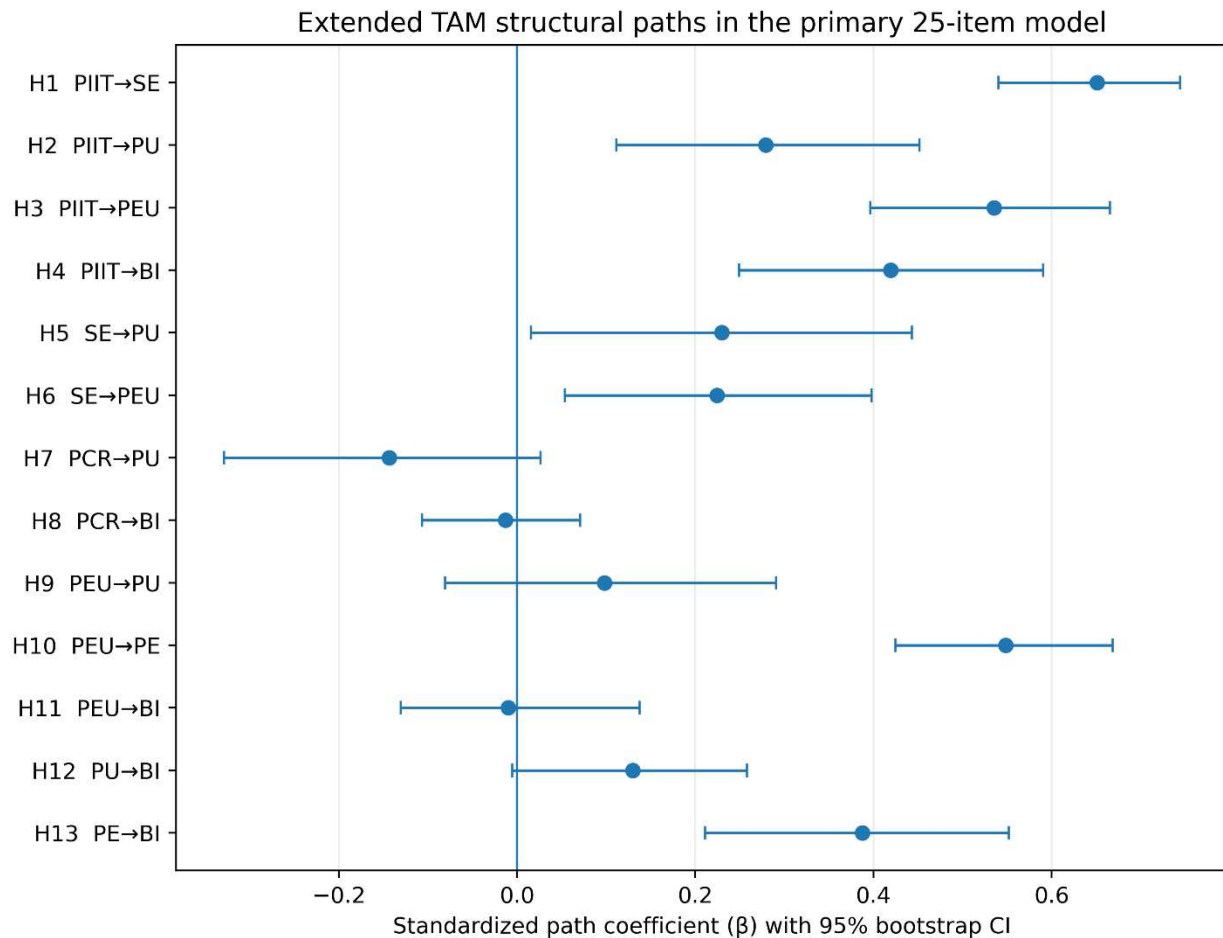


Figure 2. Standardised structural path coefficients and 95% bootstrap confidence intervals for the primary 25-item extended Technology Acceptance Model. Confidence intervals were estimated using 5,000 participant-level bootstrap samples. Intervals that do not cross zero indicate statistically supported structural associations.

Indirect and total effects

The explanatory performance of the endogenous constructs and the principal indirect and total effects on behavioral intention are summarised in **Table 5**. Although perceived ease of use had no supported direct association with behavioral intention, its total indirect effect was positive and statistically supported ($\beta=0.225$, 95% CI 0.128–0.328), primarily through perceived enjoyment.

Personal innovativeness also had a positive indirect effect on behavioral intention ($\beta=0.203$, 95% CI 0.083–0.334), resulting in a total effect of 0.622 (95% CI 0.495–0.751). Self-efficacy had a smaller but supported indirect effect on behavioral intention ($\beta=0.078$, 95% CI 0.021–0.161). The indirect and total effects of perceived cyber risk were not statistically supported.

Table 5. Explanatory, indirect, and predictive performance of the primary model

Panel A. Variance explained in endogenous constructs

Endogenous construct	Number of predictors	(R ²)	Adjusted (R ²)
SE	1	0.424	0.421
PEU	2	0.494	0.488
PU	4	0.389	0.376
PE	1	0.301	0.297
BI	5	0.701	0.692

Panel B. Principal indirect and total effects on behavioral intention

Source construct	Effect	Estimate	95% bootstrap CI	(p)
PIIT	Indirect effect	0.203	0.083 to 0.334	<0.001
PIIT	Total effect	0.622	0.495 to 0.751	<0.001
SE	Indirect and total effect	0.078	0.021 to 0.161	0.004
PEU	Indirect effect	0.225	0.128 to 0.328	<0.001
PEU	Total effect	0.216	0.054 to 0.403	0.010
PCR	Indirect effect	-0.019	-0.062 to 0.004	0.152
PCR	Total effect	-0.031	-0.136 to 0.059	0.489
PU	Total effect	0.130	-0.005 to 0.258	0.057
PE	Total effect	0.388	0.211 to 0.552	<0.001

Panel C. Repeated cross-validated prediction of the behavioral-intention composite

Model	RMSE	MAE	(Q ² _{{\mathrm{predict}}})
Base TAM	0.525	0.417	0.484
Extended TAM	0.447	0.338	0.627

The mean participant-level reduction in squared prediction loss was 0.077; paired permutation ($p=0.002$).

Out-of-sample predictive performance

The comparative predictive performance of the base and extended Technology Acceptance Models is reported in **Table 5** and visualised in **Figure 3**. For the behavioral-intention composite, repeated 10-fold cross-validation yielded an RMSE of 0.525, an MAE of 0.417, and a (Q^2_{predict}) of 0.484 for the base TAM. The corresponding values for the extended model were an RMSE of 0.447, an MAE of 0.338, and a (Q^2_{predict}) of 0.627.

The mean participant-level reduction in squared prediction loss was 0.077 and was statistically supported by the paired permutation test ($p=0.002$). The extended model also produced lower RMSE values for each of the three BI indicators. However, the participant-level reduction in

prediction loss was statistically supported for BI1, whereas the improvements for BI2 and BI3 did not reach the prespecified significance threshold. Indicator-level predictive results for the behavioral-intention composite and the three BI indicators are provided in **Table S4**. Overall, the extended model demonstrated superior out-of-sample prediction of the behavioral-intention composite compared with the base TAM.

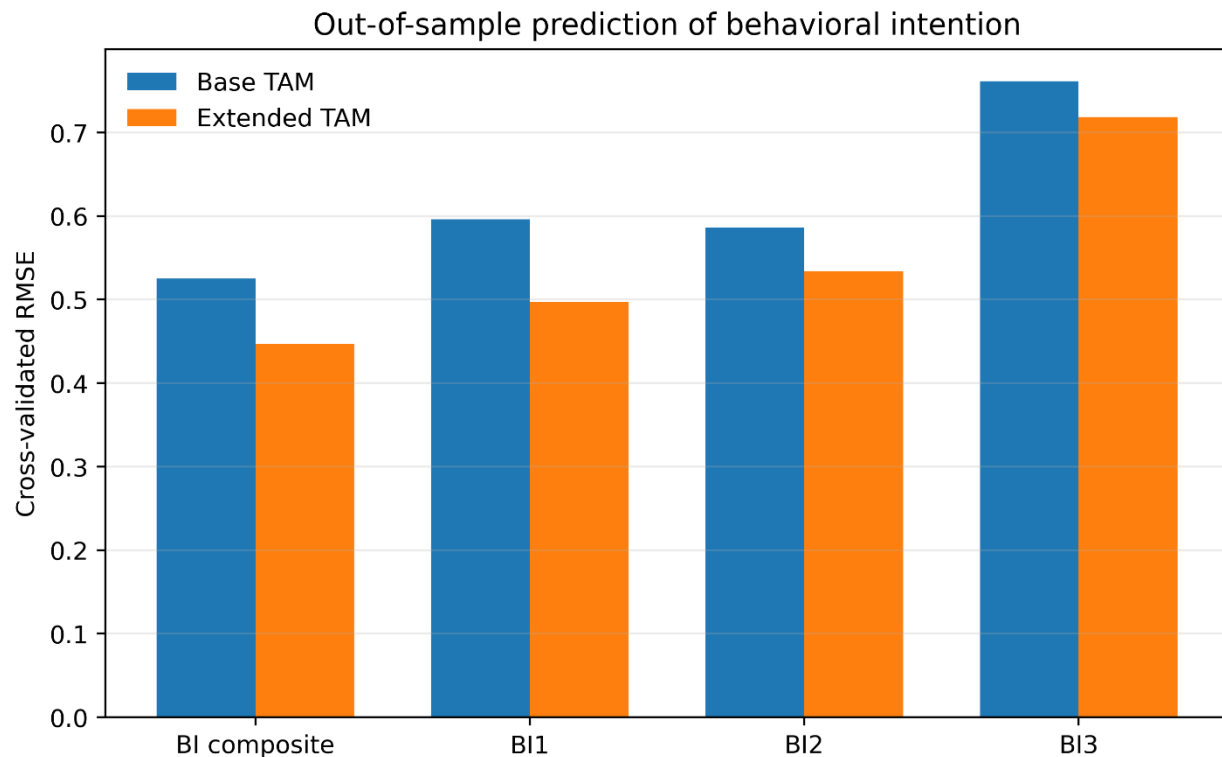


Figure 3. Repeated cross-validated prediction error for behavioral intention under the base and extended Technology Acceptance Models. Values represent root mean squared error obtained from repeated 10-fold cross-validation. Lower values indicate better out-of-sample predictive performance.

Importance–performance analysis

The importance–performance map for behavioral intention is presented in **Figure 4**. Personal innovativeness had the largest estimated total effect on behavioral intention (importance 0.622) and a performance score of 62.2. However, its implementation priority should be interpreted cautiously because of the measurement limitations of the PIIT block.

Perceived enjoyment combined a comparatively large total effect of 0.388 with a performance score of 74.6 and represented the most psychometrically robust actionable construct in the model. Perceived ease of use had a smaller total effect of 0.216 and a performance score of 67.2, consistent with its predominantly indirect contribution to behavioral intention.

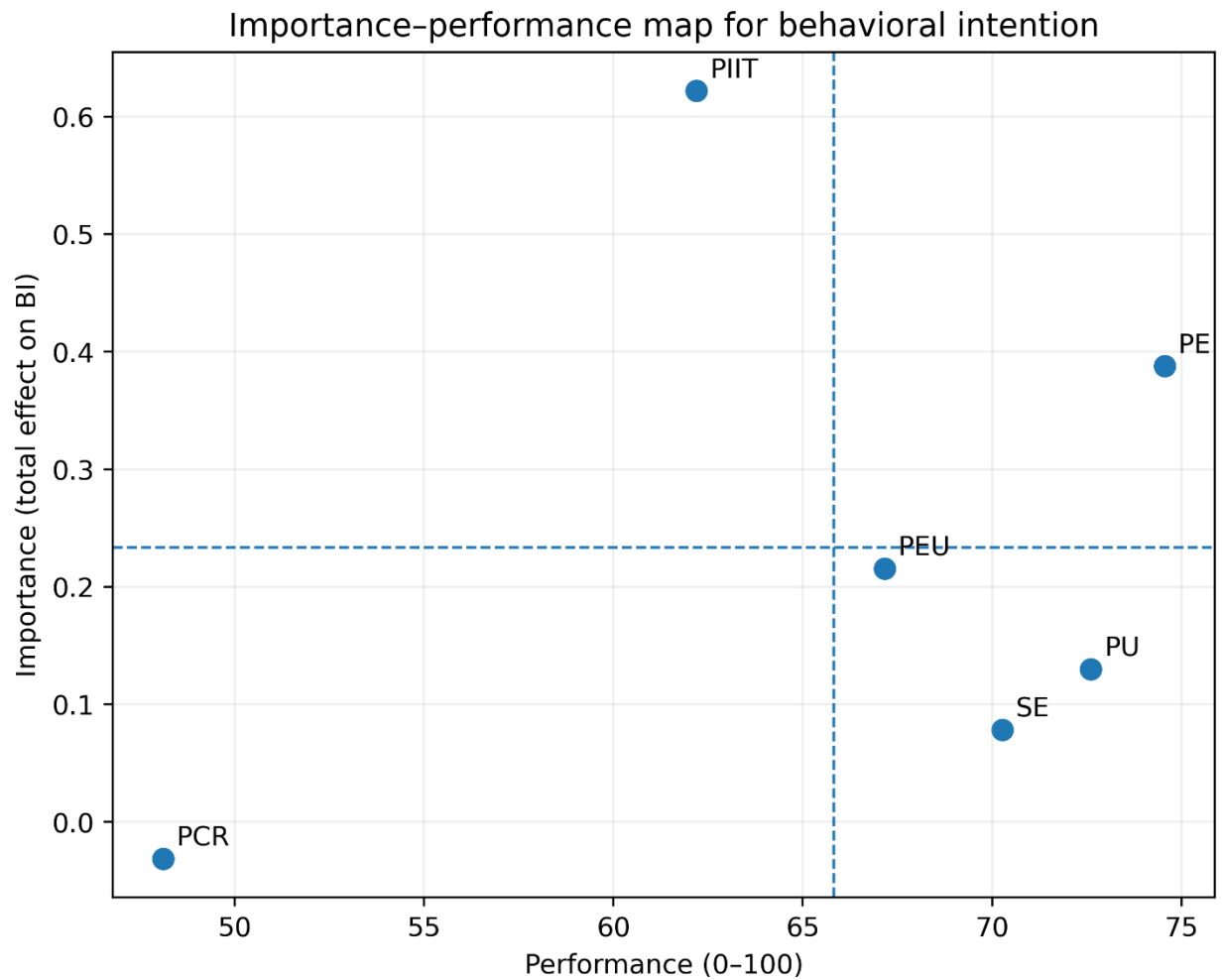


Figure 4. Importance–performance map for behavioral intention. Importance represents the total effect of each predecessor construct on behavioral intention, while performance is rescaled from 0 to 100. Interpretation of PIIT should remain cautious because the construct did not demonstrate adequate measurement performance.

Measurement invariance and faculty-specific structural differences

The complete MICOM results and permutation-based multi-group comparisons are reported in **Table S3**. Configural invariance was established for all constructs. However, compositional invariance was established only for perceived usefulness and behavioral intention. Accordingly, most faculty-specific structural-path comparisons did not satisfy the measurement-invariance prerequisite for confirmatory interpretation.

The PU→BI pathway was the only structural comparison for which both the predictor and outcome constructs achieved compositional invariance. This pathway was positive among paramedical students ($\beta=0.351$, 95% CI 0.125–0.586) but was not supported among medical students

(($\beta=-0.059$), 95% CI -0.267 to 0.125). The between-faculty difference was 0.410 , with a permutation ($p=0.002$) and a Holm-adjusted ($p=0.026$). Although this comparison met the compositional-invariance requirement, it was interpreted as exploratory because of the modest group-specific sample sizes and the measurement limitations identified elsewhere in the model. All other faculty-specific path comparisons were treated as exploratory and were not interpreted as evidence of faculty moderation.

Exploratory programme-level profiles

Programme-level construct profiles are presented as heatmaps in **Figures 5 and 6**. Among paramedical students, programme-level differences remained statistically significant after Holm adjustment for personal innovativeness ($H(7) = 22.86$, adjusted $p = 0.011$), perceived ease of use ($H(7) = 23.15$, adjusted $p = 0.011$), and behavioral intention ($H(7) = 19.63$, adjusted $p = 0.032$). The heatmap suggested comparatively higher behavioral-intention scores in physiotherapy and laboratory sciences, although some programmes contained very few participants and the visual profiles should not be interpreted as stable programme rankings.

The recorded medical-faculty categories showed less pronounced separation in their overall construct profiles, and none of the omnibus between-category comparisons remained statistically significant after Holm correction. The medical heatmap should be interpreted descriptively because the recorded basic-sciences category was heterogeneous and several categories were very small, including an internship category containing one participant. These exploratory profiles were retained because they provide a rapid visual summary of potential programme-level heterogeneity, but they were not used to support substantive between-programme conclusions.

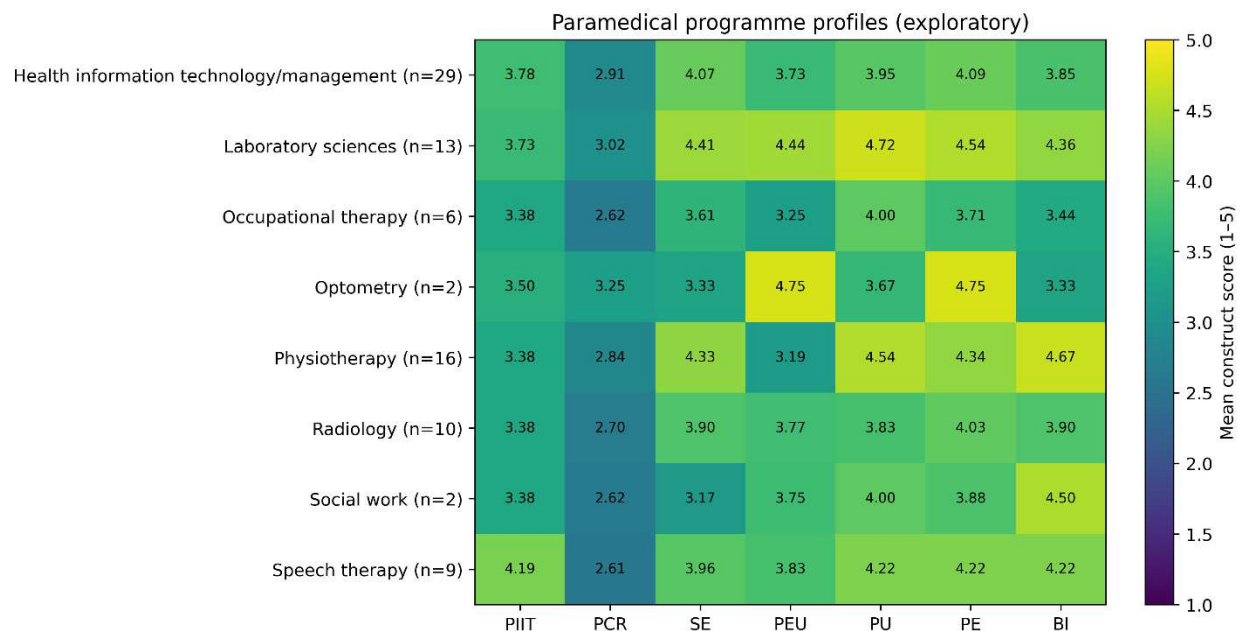


Figure 5. Mean construct profiles across paramedical programmes. Cells represent mean construct scores, and row labels report the number of students in each programme. The profiles are descriptive and exploratory because several programmes had small sample sizes.

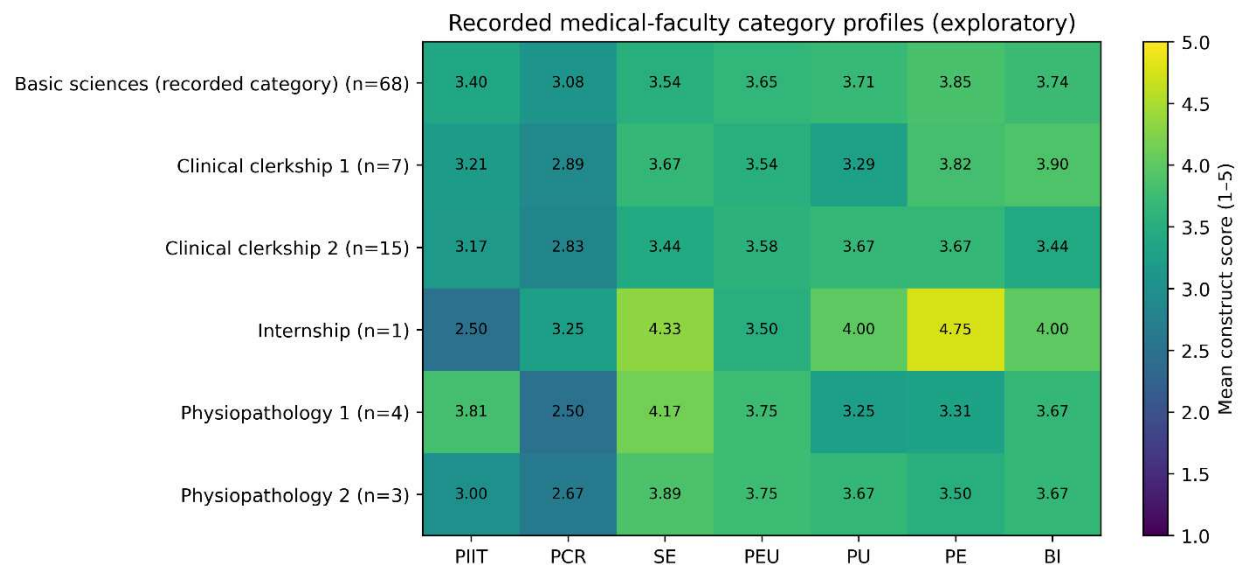


Figure 6. Mean construct profiles across recorded medical-faculty categories. Cells represent mean construct scores, and row labels report the number of students in each recorded category. The basic-sciences category was heterogeneous, and several categories were small; therefore, the figure is descriptive and exploratory.

Common-method diagnostics

Common-method and approximate model-fit diagnostics are reported in **Table S2**. Harman's single-factor analysis indicated that the first unrotated factor accounted for 46.0% of the total item variance. Full-collinearity VIFs were also examined as supplementary diagnostic indicators. Because neither procedure can definitively establish the presence or absence of common-method bias, the findings were interpreted cautiously.

Sensitivity analysis retaining PU4

A detailed comparison of the primary 25-item model and the complete 26-item model retaining PU4 is provided in **Table S5**. Re-estimation of the full administered instrument produced substantively similar findings. The complete model explained 69.8% of the variance in behavioral intention, compared with 70.1% in the context-adapted 25-item model.

The positive direct associations of personal innovativeness and perceived enjoyment with behavioral intention remained supported, whereas the direct associations of perceived cyber risk, perceived ease of use, and perceived usefulness remained unsupported. Excluding PU4 increased the PU→BI coefficient from 0.101 to 0.130, but the confidence interval continued to include zero. The measurement limitations of PIIT and PCR were unchanged.

In the complete 26-item model, compositional invariance was achieved only for behavioral intention; therefore, none of the full-model path differences met the measurement-invariance requirement for confirmatory between-faculty interpretation. Overall, the principal explanatory and predictive conclusions were robust to the treatment of PU4.

Discussion

Principal findings

This study examined medical and paramedical students' behavioral intention to use metaverse-based learning within an extended Technology Acceptance Model. Students reported a relatively high overall level of behavioral intention, with a mean score of 3.89 out of 5. The primary model explained 70.1% of the variance in behavioral intention, and the extended model demonstrated better out-of-sample predictive performance than the base TAM. Personal innovativeness in information technology and perceived enjoyment were the only constructs with statistically supported direct associations with behavioral intention. Perceived ease of use did not have a direct association with intention but contributed indirectly, primarily through perceived enjoyment. Perceived usefulness and perceived cyber risk were not directly associated with behavioral intention. Paramedical students reported higher observed scores for behavioral intention, self-efficacy, perceived usefulness, and perceived enjoyment than medical students. However, most between-faculty structural comparisons could not be interpreted confirmatorily because compositional invariance was not established for several constructs.

These findings extend the original extended TAM proposed by Al-Adwan et al. [1] in several ways. Consistent with that study, personal innovativeness and perceived enjoyment emerged as important determinants of intention, while perceived ease of use did not exert an independent direct effect. However, unlike the source study, perceived usefulness was not directly associated with behavioral intention in the present sample. The confidence interval for the PU→BI pathway only slightly crossed zero, and the corresponding (p)-value was 0.057. Therefore, this finding should be interpreted as insufficient evidence for a direct effect rather than evidence that perceived usefulness is irrelevant. The results instead suggest that the relative importance of traditional TAM constructs may change when students evaluate a novel and largely prospective educational technology rather than a system they already use routinely.

Behavioral intention and the role of perceived enjoyment

The relatively high behavioral-intention score is consistent with earlier studies showing that university and health sciences students generally perceive the metaverse as a potentially attractive educational environment, despite limited prior experience with immersive platforms [3,4,6]. Positive intention may reflect the anticipated capacity of metaverse-based environments to make learning more interactive, experiential, and socially engaging. Nevertheless, behavioral intention should not be interpreted as evidence of actual adoption. Students may express enthusiasm toward a novel technology before encountering practical constraints such as equipment availability,

connectivity requirements, cybersickness, technical complexity, or a mismatch between platform design and curriculum needs.

Perceived enjoyment was the most psychometrically robust direct predictor of behavioral intention in the present study. This finding is theoretically plausible because metaverse-based learning is not merely a utilitarian information system; it is also an immersive and experiential environment. Enjoyment may increase voluntary engagement, sustain attention, encourage exploration, and reduce the psychological distance associated with unfamiliar technologies. Previous studies have similarly suggested that engagement, interaction, and the experiential quality of virtual environments shape students' perceptions of metaverse-based learning [3,4].

The importance of enjoyment does not imply that metaverse activities should be designed primarily as entertainment. Rather, enjoyment may function as a mechanism through which students remain engaged with demanding educational tasks. This distinction is important in health professions education, where novelty and visual appeal are valuable only when they support clinical reasoning, procedural learning, teamwork, or communication. Emerging intervention studies provide evidence that immersive environments can support substantive learning outcomes. Metaverse-based teaching has been found feasible for medical-student instruction and assessment [5], while an immersive emergency-skills program improved objective structured clinical examination performance among medical students [23]. Similarly, metaverse-based problem-based learning in radiology produced educational outcomes comparable to face-to-face instruction and was positively evaluated by students [24]. These studies indicate that enjoyment and engagement can coexist with academically meaningful learning when the technology is aligned with explicit educational objectives.

Personal innovativeness, self-efficacy, and readiness for adoption

Personal innovativeness had the largest estimated total effect on behavioral intention and was positively related to self-efficacy, perceived ease of use, and perceived usefulness. At a substantive level, this pattern suggests that students who are more willing to experiment with emerging technologies may also feel more capable of using them and may recognize their educational value more readily. This interpretation is consistent with evidence that innovative orientations among healthcare professionals are associated with greater knowledge and awareness of the metaverse [8].

Digital readiness may therefore operate as a prerequisite for meaningful engagement. Among nursing students in Egypt, digital literacy was strongly associated with metaverse perspectives and partially mediated the relationship between metaverse perspectives and academic motivation [25]. In the present study, self-efficacy also contributed indirectly to behavioral intention by improving perceptions of ease of use and usefulness. Taken together, these findings imply that access to a platform alone may be insufficient. Students may also require introductory training, opportunities for low-stakes experimentation, technical support, and structured guidance before they can evaluate the educational value of immersive learning accurately.

Personal innovativeness should nevertheless be interpreted cautiously in this study because the PIIT measurement block did not demonstrate adequate psychometric performance. The reverse-worded PIIT3 item had a negative loading, internal consistency was low, and the HTMT value between PIIT and behavioral intention was high. Accordingly, the estimated importance of PIIT should be regarded as provisional. From an implementation perspective, innovative students could be engaged as early adopters or peer facilitators, but educational programs should not be designed only for technologically enthusiastic students. Overreliance on early adopters may widen participation gaps by disadvantaging students with less prior exposure, lower digital confidence, or limited access to suitable devices.

Ease of use as an indirect enabling mechanism

Perceived ease of use did not directly predict behavioral intention, but its total indirect effect was positive and statistically supported, mainly through perceived enjoyment. This finding suggests that ease of use may function as an enabling condition rather than an end in itself. A platform that is easier to navigate may reduce cognitive and technical burden, allowing students to focus on interaction, exploration, and the learning task. In turn, a smoother experience may increase enjoyment and strengthen intention to use the platform.

This mediated pattern is compatible with the source extended TAM [1] and refines the traditional TAM proposition that ease of use directly shapes acceptance [2]. For immersive educational technologies, usability may influence intention primarily by improving the quality of the experience. This interpretation is also consistent with intervention studies in which students benefited from metaverse-based simulation when the educational environment was structured and supported. For example, a metaverse-based psychiatric nursing simulation improved students' knowledge, critical thinking, and communication skills, but the authors also emphasized the need to identify technical limitations and select appropriate educational topics [26].

The practical implication is that implementation strategies should reduce unnecessary technical complexity. Orientation sessions, guided practice, clear navigation, responsive technical support, and gradual exposure may be more effective than assuming that students will independently learn to use immersive platforms. The goal should be to make the technology sufficiently transparent that attention remains on the clinical or educational task.

Perceived usefulness and the difference between anticipated and demonstrated value

Perceived usefulness was high overall but did not have a statistically supported direct association with behavioral intention in the pooled model. One explanation is that students were evaluating the potential value of a technology with which many may have had limited direct experience. Under such conditions, perceived usefulness may remain abstract and may not translate into intention until learners encounter credible, curriculum-specific applications.

Applied evidence from medical education increasingly demonstrates that usefulness depends on the educational task. Metaverse environments have been used for clinical teaching and assessment

[5], emergency-skills training [23], radiology problem-based learning [24], psychiatric nursing simulation [26], virtual collaboration in an academic health informatics laboratory [27], and surgical planning [28]. These applications differ substantially in technological requirements and pedagogical purpose. Consequently, asking students whether the metaverse is useful in general may produce less decisive judgments than allowing them to evaluate a specific application that solves an identifiable educational problem.

The absence of a direct pooled PU→BI effect therefore should not be interpreted as a reason to disregard usefulness. Instead, it suggests that perceived usefulness may need to be demonstrated through authentic use cases. Short pilot activities connected to recognizable curricular needs—such as clinical case discussion, anatomy visualization, communication practice, procedural rehearsal, or interprofessional collaboration—may allow students to form more concrete judgments about value.

Differences between medical and paramedical students

Paramedical students reported higher behavioral intention, self-efficacy, perceived usefulness, and perceived enjoyment than medical students. The difference in behavioral intention was of moderate magnitude, while the largest observed differences involved self-efficacy and perceived usefulness. These differences may reflect variation in curricular structure, prior exposure to digital tools, perceived compatibility with professional tasks, or expectations regarding educational realism.

Students in some paramedical programs may more readily associate immersive technologies with visual, spatial, rehabilitative, laboratory, or simulation-based activities. Medical students, by contrast, may evaluate new educational technologies against high expectations for clinical authenticity, diagnostic complexity, patient safety, and transfer to real clinical environments. They may therefore require stronger evidence that a virtual activity offers added value beyond existing clinical teaching, simulation centers, or bedside instruction.

The PU→BI pathway was positive among paramedical students but not supported among medical students. This was the only between-faculty path comparison for which both constructs achieved compositional invariance. The result may indicate that perceived practical value translates more directly into intention among paramedical students, whereas medical students' intentions may depend more heavily on other considerations, such as enjoyment, innovation readiness, realism, or curricular legitimacy. However, this finding should remain exploratory. The group-specific samples were modest, and measurement invariance was not established for most other constructs.

The program-level heatmaps provided additional descriptive evidence of heterogeneity. Physiotherapy and laboratory-sciences students showed comparatively high behavioral-intention profiles, while the medical-faculty categories were less clearly differentiated. These visual patterns should not be treated as stable program rankings because several subgroups were small and the recorded basic-sciences category was heterogeneous. Nevertheless, they generate useful

hypotheses for future studies examining whether the perceived relevance of the metaverse differs across procedural, diagnostic, rehabilitative, laboratory, and clinically oriented curricula.

Resource context, national income, and equity

The present findings should also be considered within the broader resource context of metaverse implementation. Cross-country comparisons must be made cautiously because studies differ in populations, instruments, educational systems, and technology exposure. Nevertheless, the available literature suggests an implementation gap between higher-resource institutions and settings in which digital infrastructure, equipment access, or technical support may be more constrained.

In several comparatively well-resourced settings, research has progressed from measuring general acceptance to testing concrete applications and educational outcomes. Examples include immersive medical-student assessment [5], emergency-skills training in China [23], radiology problem-based learning in Spain [24], and metaverse-assisted surgical planning in Italy [28]. In contrast, studies from settings that may face greater resource constraints more often emphasize awareness, digital literacy, readiness, prior experience, and access. Studies among health sciences students, nurses, and healthcare professionals in Turkey found generally positive perceptions but limited direct experience and variation according to technology-related characteristics [6–8]. Similarly, the Egyptian nursing study demonstrated that technology access, online-learning experience, and digital literacy were closely related to metaverse perspectives and academic motivation [25].

This contrast should not be interpreted as lower interest or capability in lower-resource settings. Rather, it may indicate that adoption is conditioned by institutional readiness. High behavioral intention can coexist with limited access to head-mounted displays, unstable connectivity, high procurement costs, inadequate technical support, or uncertainty about data governance. National income level is therefore better understood as a proxy for system-level conditions than as a direct explanation of individual acceptance.

Equity should be incorporated into implementation from the outset. An equitable strategy may include shared institutional laboratories, device-loan schemes, browser-accessible alternatives, low-bandwidth modes, scheduled access outside normal teaching hours, accessibility accommodations, and support for students with lower digital confidence. Frameworks for equitable virtual care similarly emphasize accessibility, interoperability, governance, humanization, and integration with existing services [29]. Without such safeguards, metaverse-based education could widen rather than reduce educational inequalities.

Perceived cyber risk and data governance

Perceived cyber risk was not directly associated with behavioral intention or perceived usefulness. However, this result should not be interpreted as evidence that privacy and security are unimportant. The PCR block had very low internal consistency and mixed loading directions,

indicating that the construct was not measured reliably in the present sample. The items combined positively worded statements about security safeguards with negatively worded concerns about unauthorized data distribution and privacy invasion. These formulations may have activated different interpretations rather than a single coherent perception of cyber risk.

Healthcare-related metaverse platforms may involve personal identifiers, behavioral traces, voice or video recordings, spatial data, and potentially sensitive educational or clinical information. Therefore, institutions should establish transparent policies regarding data collection, ownership, storage, third-party access, consent, and platform accountability before implementation. Studies among nurses and healthcare professionals have likewise emphasized ethical, privacy, and security considerations even when overall attitudes toward the metaverse were favorable [7,8].

Measurement considerations

The measurement results are central to interpreting this study. PU, PEU, SE, PE, and BI showed satisfactory reliability and convergent validity, whereas PIIT and PCR did not. The negative loading of the reverse-worded PIIT3 item and the mixed loading pattern within PCR suggest possible wording or method effects. Reverse-keyed items can reduce measurement quality when respondents process negation inconsistently, particularly in short and heterogeneous scales [30]. Such problems may become more pronounced during translation because grammatical negation, emphasis, and response direction do not always transfer equivalently across languages.

These findings reinforce the distinction between linguistic translation and construct equivalence. Cross-cultural adaptation requires evaluation of semantic, experiential, and conceptual correspondence rather than literal translation alone [13]. Future Persian-language validation should include cognitive interviews, think-aloud testing, pilot factor analysis, and explicit evaluation of method effects associated with reverse wording. Rephrasing all indicators in the same conceptual direction or modeling a wording factor may be preferable to relying on post hoc reverse scoring.

The context-based exclusion of PU4 was also justified by content rather than statistical performance because the administered item referred specifically to higher tourism education. The sensitivity analysis retaining PU4 yielded substantively similar structural and predictive findings, which supports the robustness of the main conclusions. Nevertheless, the need to remove the item illustrates the importance of conducting a detailed domain-specific review before administration.

Practical implications for health professions education

The results support a staged rather than technology-driven implementation strategy. Universities should begin with a clearly defined educational problem and determine whether a metaverse-based solution offers advantages over existing approaches. Pilot activities should be curriculum-aligned, time-limited, and evaluated using both learner outcomes and implementation indicators.

Perceived enjoyment emerged as the most defensible actionable construct, while ease of use contributed indirectly. Accordingly, platforms should provide meaningful interaction, immediate feedback, opportunities for exploration, and a sense of social presence without adding unnecessary technical burden. Enjoyment should be designed around educational engagement rather than novelty alone.

Students with higher innovation readiness may facilitate early implementation, but orientation and technical support should be available to all learners. Faculty development is equally important because the educational value of the metaverse depends on instructional design, facilitation, assessment, and debriefing. Institutions should also evaluate device availability, connectivity, accessibility, privacy, staff workload, and long-term maintenance before scaling.

The superior predictive performance of the extended model is practically relevant. Compared with the base TAM, the extended model reduced RMSE for the behavioral-intention composite by approximately 15% and increased (Q^2_{predict}) from 0.484 to 0.627. This finding suggests that implementation planning based only on usefulness and ease of use may overlook important motivational and personal determinants. However, because PIIT and PCR had measurement limitations, perceived enjoyment currently provides the clearest basis for actionable interpretation.

Strengths and limitations

This study had several strengths. Proportionate stratified random sampling was used across two faculties, the response rate was high, and all 185 valid questionnaires were retained. Mandatory electronic responses eliminated item-level missingness. The analysis evaluated the measurement and structural models, indirect effects, out-of-sample prediction, importance–performance relationships, measurement invariance, faculty differences, and sensitivity to the treatment of PU4. The comparison between the base and extended TAM also moved beyond explanatory fit by evaluating predictive performance [18]. In addition, unfavorable measurement findings were reported rather than concealed through extensive item deletion.

Several limitations should be considered. First, the cross-sectional design precludes causal inference. The reported paths represent statistical associations within the proposed theoretical model and should not be interpreted as evidence that changing one construct would necessarily cause a change in behavioral intention.

Second, behavioral intention was self-reported and did not represent actual platform use, persistence, learning achievement, or clinical performance. Novelty may have increased favorable responses among students with little direct experience. Future studies should assess intention before exposure, actual use behavior during implementation, and sustained adoption after the novelty period.

Third, the study was conducted at a single university, and most participants were women. The findings may not generalize to other Iranian universities, private institutions, other health

professions, or different national contexts. The medical-faculty group also included heterogeneous educational stages and postgraduate or basic-science programs, while several program-level subgroups were small.

Fourth, the PIIT and PCR blocks showed poor psychometric performance, discriminant validity was incomplete, and the approximate SRMR indicated that global model approximation was not optimal. Therefore, findings involving these constructs must be considered exploratory. The inability to establish compositional invariance for most constructs also restricted confirmatory comparison of structural paths between faculties [20].

Fifth, all constructs were assessed within the same electronic questionnaire, creating the possibility of common-method bias. Harman's single-factor result and full-collinearity VIFs were used only as diagnostics and cannot rule out common-method variance [22]. Mandatory responses may also have encouraged some respondents to provide rapid or less considered answers.

Sixth, the study did not directly measure prior metaverse experience, device ownership, internet quality, digital literacy, disability-related accessibility needs, or institutional technology access. These variables may help explain acceptance, especially when comparing resource contexts. Finally, although the Python analytical workflow was reproducible, independent replication using another established PLS-SEM implementation would further strengthen computational verification.

Future research

Future studies should first refine and validate the Persian PIIT and PCR measures using cognitive interviewing and independent samples. Longitudinal and experimental designs should then examine whether behavioral intention predicts actual adoption and whether perceived enjoyment, ease of use, and self-efficacy change after direct exposure.

Multi-center research is needed to compare medical and paramedical programs across universities with different levels of digital infrastructure. Such studies should measure resource availability directly rather than relying on country-income categories. They should also examine whether low-cost, desktop-based, mobile, or shared-device approaches produce comparable acceptance and learning outcomes to headset-dependent systems.

Finally, future evaluations should combine acceptance outcomes with objective educational measures, including knowledge retention, skills performance, transfer to clinical practice, engagement analytics, adverse effects, cost, accessibility, and equity. Qualitative interviews may also clarify why usefulness translated into intention among paramedical students but not among medical students. Future research should also move beyond acceptance surveys to the co-design, development, and evaluation of metaverse education delivered through structured courses and user-centred educational applications. Evidence from adjacent fields of emerging digital health education indicates that favorable attitudes among healthcare students may coexist with gaps in

conceptual knowledge and practical skills, while students value interactive, offline-capable, and personalized educational tools designed to address these needs [31–33].

Conclusions

Medical and paramedical students demonstrated relatively high behavioral intention to use metaverse-based learning. Perceived enjoyment had a robust direct association with intention, while perceived ease of use contributed primarily through indirect pathways. Personal innovativeness showed a large estimated contribution but requires cautious interpretation because of measurement limitations. The extended TAM predicted behavioral intention more accurately than the base model, indicating that motivational and personal factors add value beyond usefulness and ease of use alone.

Paramedical students reported higher intention and more favorable observed scores on several constructs, although most structural differences between faculties remained exploratory because measurement invariance was incomplete. The findings support staged, curriculum-aligned, enjoyable, usable, and equitable implementation rather than technology adoption for its own sake. Before large-scale deployment, the Persian measurement model should be refined, and the proposed relationships should be examined using actual metaverse exposure, objective educational outcomes, and multi-center samples.

Declarations

Ethics approval and consent to participate: The study was approved by the Research Ethics Committee of Mashhad University of Medical Sciences (IR.MUMS.FHMPM.REC.1405.101) and was conducted in accordance with the Declaration of Helsinki. All participants received electronic information about the study and provided electronic informed consent before accessing the questionnaire.

Clinical trial number: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: The data supporting the results of this study are available from the corresponding author upon reasonable request.

Code availability: No custom software, algorithm, or computational tool was developed in this study. The analyses were conducted using standard statistical procedures. The analytical outputs and detailed results required to understand and evaluate the findings are provided in the Supplementary Information.

Competing interests: The authors declare that they have no competing interests.

Funding: This research received no specific external grant. Institutional administrative support was provided by Mashhad University of Medical Sciences under project No. 4042054.

Authors' contributions: Seyyedeh Fatemeh Mousavi Baigi: Conceptualization; Methodology; Validation; Formal analysis; Data curation; Supervision; Writing - review & editing; Writing - original draft; Visualization; Software. Masoumeh Sarbaz: Writing - review & editing; Methodology; Validation; Investigation. Khalil Kimiafar: Conceptualization; Project administration; Supervision; Writing - review & editing; Validation; Methodology. All authors read and approved the final manuscript.

Acknowledgements: This study is part of project No. 4042054 conducted by the Student Research Committee, Mashhad University of Medical Sciences, Mashhad, Iran. We extend our sincere gratitude to the “Student Research Committee” and the “Research & Technology Chancellor” of Mashhad University of Medical Sciences for their support of this study. We also wish to thank the healthcare students of Mashhad University of Medical Sciences who participated in this study and contributed to the completion of the research.

AI Use Statement: AI-assisted technology was used in the preparation of this manuscript. Specifically, ChatGPT by OpenAI was used solely for English language editing, improving readability, and refining the structure of selected sections of the manuscript. It was not used to generate research ideas, design the study, collect or analyze data, interpret findings, or formulate scientific conclusions. All AI-assisted content was critically reviewed, revised, and approved by the authors. The authors take full responsibility for the accuracy, integrity, originality, and ethical compliance of the manuscript.

References

1. Al-Adwan AS, Li N, Al-Adwan A, Abbasi GA, Albelbisi NA, Habibi A. Extending the technology acceptance model (TAM) to predict university students' intentions to use metaverse-based learning platforms. *Educ Inf Technol*. 2023;28(11):15381–15413. doi: 10.1007/s10639-023-11816-3.

2. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 1989;13(3):319–340. doi: 10.2307/249008.
3. Talan T, Kalinkara Y. Students' opinions about the educational use of the metaverse. *Int J Technol Educ Sci.* 2022;6(2):333–346. Available from: <https://eric.ed.gov/?id=EJ1345311>.
4. Rojas E, Hülsmann X, Estriegana R, Rückert F, Garcia-Esteban S. Students' perception of metaverses for online learning in higher education: hype or hope? *Electronics.* 2023;12(8):1867. doi: 10.3390/electronics12081867.
5. Rodriguez-Florido MÁ, Reyes-Cabrera JJ, Melián A, Hernández-Flores CN, Ruiz-Azola J, Maynar M. Feasibility of teaching and assessing medical students in the metaverse: design and features for its learning efficiency. *J New Approaches Educ Res.* 2024;13(1):9. doi: 10.1007/s44322-024-00009-6.
6. Sezer H, Gül A. Health sciences students' perspectives on metaverse and digital technology use: cross-sectional descriptive study. *BMC Nurs.* 2024;23:656. doi: 10.1186/s12912-024-02309-w.
7. Ergin E, Yalcinkaya T, Cinar Yucel S. Nurses' knowledge of, attitudes towards and awareness of the metaverse, and their future time perspectives: a cross-sectional study. *BMC Nurs.* 2024;23:414. doi: 10.1186/s12912-024-02048-y.
8. Duzcu T. The relationship between health professionals' perceptions of innovative work behavior and their metaverse knowledge and awareness levels. *BMC Health Serv Res.* 2025;25:329. doi: 10.1186/s12913-025-12492-4.
9. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology statement: guidelines for reporting observational studies. *PLoS Med.* 2007;4(10):e296. doi: 10.1371/journal.pmed.0040296.
10. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys. *J Med Internet Res.* 2004;6(3):e34. doi: 10.2196/jmir.6.3.e34.
11. Cochran WG. Sampling techniques. 3rd ed. New York: John Wiley & Sons; 1977.
12. World Health Organization. Process of translation and adaptation of instruments. Geneva: World Health Organization. Available from: https://www.who.int/substance_abuse/research_tools/translation/en/.
13. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000;25(24):3186–3191. doi: 10.1097/00007632-200012150-00014.

14. Lawshe CH. A quantitative approach to content validity. *Pers Psychol.* 1975;28(4):563–575. doi: 10.1111/j.1744-6570.1975.tb01393.x.
15. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health.* 2007;30(4):459–467. doi: 10.1002/nur.20199.
16. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. *Eur Bus Rev.* 2019;31(1):2–24. doi: 10.1108/EBR-11-2018-0203.
17. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J Acad Mark Sci.* 2015;43:115–135. doi: 10.1007/s11747-014-0403-8.
18. Shmueli G, Sarstedt M, Hair JF, Cheah JH, Ting H, Vaithilingam S, et al. Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *Eur J Mark.* 2019;53(11):2322–2347. doi: 10.1108/EJM-02-2019-0189.
19. Ringle CM, Sarstedt M. Gain more insight from your PLS-SEM results: the importance–performance map analysis. *Ind Manag Data Syst.* 2016;116(9):1865–1886. doi: 10.1108/IMDS-10-2015-0449.
20. Henseler J, Ringle CM, Sarstedt M. Testing measurement invariance of composites using partial least squares. *Int Mark Rev.* 2016;33(3):405–431. doi: 10.1108/IMR-09-2014-0304.
21. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat.* 1979;6(2):65–70. Available from: <https://www.jstor.org/stable/4615733>.
22. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol.* 2003;88(5):879–903. doi: 10.1037/0021-9010.88.5.879.
23. Huang H, Yin J, Lv F, Lin Y, Zou J. A study on the impact of open source metaverse immersive teaching method on emergency skills training for medical undergraduate students. *BMC Med Educ.* 2024;24(1):859. doi: 10.1186/s12909-024-05862-9.
24. Rudolphi-Solero T, Bajos-Ariza F, Lorenzo-Álvarez R, Domínguez-Pinos D, Ruiz-Gómez MJ, Sendra-Portero F. Problem-based learning in radiology achieves similar results in classroom and metaverse settings. *Insights into Imaging.* 2025 Jun 12;16(1):121.
25. Amin SM, Atta MH, Abd Elmoaty AE, Ismail HA, Kotp MH, Basyouny HA, Aly MA, El-Sayed AA. Bridging virtual and real learning: the role of digital literacy and metaverse perspectives in enhancing academic motivation in nursing education. *Bmc Nursing.* 2025 Aug 25;24(1):1113.

26. Yang SY, Kang MK. Efficacy testing of a multi-access metaverse-based early onset schizophrenia nursing simulation program: a quasi-experimental study. *Int J Environ Res Public Health*. 2023;20(1):449. doi: 10.3390/ijerph20010449.
27. Zhu S, Vennemeyer S, Xu C, Wu DTY. Adopting a metaverse-based workspace to support research team collaboration: a pilot study from an academic health informatics laboratory. *JAMIA Open*. 2023;6(1):ooad010. doi: 10.1093/jamiaopen/ooad010.
28. Volpi G, Gatti C, Quarà A, Piramide F, Amparore D, Alessio P, De Cillis S, Colombo M, Busacca G, Sica M, Verri P. Metaverse surgical planning for robotic surgery: preliminary experience and users' perception. *Therapeutic Advances in Urology*. 2024 Dec;16:17562872241297524.
29. Veras M, Labbé DR, Furlano J, Zakus D, Rutherford D, Pendergast B, Kairy D. A framework for equitable virtual rehabilitation in the metaverse era: challenges and opportunities. *Frontiers in rehabilitation sciences*. 2023 Aug 24;4:1241020.
30. Menold N. How do reverse-keyed items in inventories affect measurement quality and information processing? *Field Methods*. 2020. doi: 10.1177/1525822X19890827.
31. Mousavi Baigi SF, Sarbaz M, Ghaddaripouri K, Ghaddaripouri M, Mousavi AS, Kimiafar K. Attitudes, knowledge, and skills towards artificial intelligence among healthcare students: a systematic review. *Health Sci Rep*. 2023;6(3):e1138. doi: 10.1002/hsr2.1138.
32. Mousavi Baigi SF, Sarbaz M, Darroudi A, Kimiafar K. Medical and paramedical students' perspectives on artificial intelligence in a low- and middle-income country: a cross-sectional study. *Health Sci Rep*. 2025;8:e71046. doi: 10.1002/hsr2.71046.
33. Mousavi Baigi SF, Norouzi Aval R, Sarbaz M, Tabatabaei SM, Kimiafar K. Content and structural needs assessment for an artificial intelligence education mobile app in healthcare: a mixed methods study. *BMC Med Educ*. 2025;25:1717. doi: 10.1186/s12909-025-08328-8.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile1.docx](#)
- [SupplementaryFile2.pdf](#)