BMC Nursing

https://doi.org/10.1186/s12912-026-05054-4

Article in Press

Development of the nursing artificial intelligence readiness scale for nursing students: a validity and reliability study

Received: 14 March 2026

Accepted: 13 July 2026

Published online: 16 July 2026

Cite this article as: Akçoban S., Koca G. & Berşe S. Development of the nursing artificial intelligence readiness scale for nursing students: a validity and reliability study. ***BMC Nurs*** **(2026). https://doi.org/10.1186/s12912-026-05054-4**

Sümeyye Akçoban, Gülay Koca & Soner Berşe

Development of the Nursing Artificial Intelligence Readiness Scale for Nursing Students: A Validity and Reliability Study

Sümeyye Akçoban[1], Gülay Koca[2] and Soner Berşe[2*]

1 Department of Health Services, Kırıkhan Vocational School, Hatay Mustafa Kemal University, Hatay, Turkey

2 Department of Nursing, Faculty of Health Sciences, Gaziantep University, Gaziantep, Turkey

Corresponding author: Soner Berşe, E-mail: sonerberse@gantep.edu.tr

Author Information

Author 1

Sümeyye Akçoban, Assist. Prof. Dr, PhD

Affiliation: Department of Health Services, Kırıkhan Vocational School, Hatay Mustafa Kemal University, Hatay, Turkey

E-mail: sumeyyea2016@gmail.com

ORCID: https://orcid.org/0000-0002-7533-8140

Author 2

Gülay Koca (BSN Student)

Affiliation: Department of Nursing, Faculty of Health Sciences, Gaziantep University, Gaziantep, Turkey

E-mail: gulaykocaa27@gmail.com

ORCID: https://orcid.org/0009-0006-2730-5789

Author 3 (Corresponding Author)

Soner Berşe, Assoc. Prof. Dr, PhD

Affiliation: Department of Nursing, Faculty of Health Sciences, Gaziantep University, Gaziantep, Turkey

E-mail: sonerberse@gantep.edu.tr

ORCID: https://orcid.org/0000-0001-9108-3216

Phone: +90 342 360 12 00/4962

Address: Gaziantep University, Faculty of Health Sciences, Department of Nursing, 27310 Şehitkamil, Gaziantep, Turkey

Abstract

Background: The rapid integration of artificial intelligence (AI) into healthcare has amplified the need for nurses who can engage with AI-supported systems safely and effectively. Assessing nursing students' AI readiness through a psychometrically sound instrument is essential for guiding curriculum design and targeted educational interventions. This study aimed to develop and psychometrically evaluate the Nursing Artificial Intelligence Readiness Scale (NAIRS) for nursing students.

Methods: This methodological scale development study was conducted with 416 nursing students enrolled during the 2025–2026 academic year at a public university in Türkiye. To avoid conducting exploratory and confirmatory factor analyses on the same dataset, the full sample was randomly split into two independent subsamples (EFA: n = 200; CFA: n = 216). The two subsamples were comparable in key demographic and AI-related characteristics. An initial 40-item draft was generated from the literature and reviewed by 10 faculty experts using a content validity evaluation; the mean CVR was .88. Following expert review, five items were removed, reducing the pool to 35 items. The revised draft was then pilot tested with 50 nursing students to assess item clarity, readability, comprehensibility, and administration flow, and minor revisions were made prior to the main application. The final 20-item scale was structured across four theoretical domains: Knowledge/Awareness, Willingness to Use AI, Self-efficacy, and Ethical Awareness. Construct validity was examined using EFA (minres extraction; Promax rotation) and first-order CFA (ML estimation) with multiple model fit indices. Convergent and discriminant validity were assessed using CR/AVE and the Fornell–Larcker criterion, respectively. Internal consistency was evaluated with Cronbach's alpha in both subsamples, and temporal stability was tested via a 2-week test–retest application in a subgroup (n = 45) using ICC (two-way random, single measures).

Results: The final scale demonstrated a four-factor structure explaining 52.08% of the total variance, with factor loadings ranging from .351 to .909. CFA supported the proposed model with good fit (χ^2/df = 1.426, CFI = .971, TLI = .967, GFI = .910, RMSEA = .044, SRMR = .049). Internal consistency was high: subscale alphas ranged .801–.828 (EFA sample) and .858–.902 (CFA sample), and the total scale alpha was .911. Test–retest reliability indicated strong stability, with ICC values ranging .896–.963 across subscales and .952 for the total scale.

Conclusions: The findings provide initial evidence that the NAIRS is a valid and reliable instrument for assessing nursing students' readiness for artificial intelligence across knowledge/awareness, willingness to use AI, self-efficacy, and ethical awareness domains. The scale may be useful for educational needs assessment and curriculum planning in nursing education.

Clinical trial number: not applicable

Keywords: Artificial intelligence; Nursing students; Readiness; Scale development; Validity; Reliability.

Contributions to the Literature

- This study provides a nursing-context–specific, psychometrically tested instrument to assess nursing students' readiness for AI technologies using a rigorous EFA–CFA cross-validation approach (EFA n = 200; CFA n = 216).
- The scale conceptualizes AI readiness as a multidimensional construct comprising knowledge/awareness, willingness to use AI, self-efficacy, and ethical awareness.
- The findings provide psychometric evidence supporting the use of the instrument in educational needs assessment, curriculum evaluation, and the planning of targeted AI-related training in nursing education.

INTRODUCTION

In the twenty-first century, the widespread adoption of computers and the internet has transformed the technology and information sectors into innovation-driven fields. Subsequently, artificial intelligence applications have rapidly become an indispensable part of daily life [1]. This transformation process has necessitated that individuals and institutions continuously adapt and embrace innovation as a strategic requirement [2]. In particular, digitalization and data-driven decision-making approaches in clinical processes related to healthcare delivery have gained momentum globally. AI-supported applications have become increasingly prominent in diagnosis, monitoring, risk classification, and clinical decision support [3].

The increasing use of artificial intelligence in healthcare systems directly affects nursing practice. Nursing is a critical stakeholder in the design, implementation, and evaluation of AI-based systems in areas such as clinical workflows, care planning, patient safety, and continuity of care [4]. Indeed, the literature on AI applications in nursing indicates that AI-based technologies are rapidly diversifying in nursing practice; however, significant needs persist regarding education, acceptance, workflow integration, and ethical/professional role ambiguity in the transition to practice. In this context, nurses' knowledge, skills, and attitudes toward AI are considered prerequisites for safe and effective use in clinical practice [5, 6].

For sustainable management of this transformation, it is important to assess not only AI literacy and technological competencies but also students' attitudes toward AI in nursing education. A study conducted with nursing students reported that students generally exhibited positive attitudes toward AI; however, these attitudes should be considered together with educational needs, intention to use, and anxiety levels [7]. Similarly, another study indicated that students' readiness and acceptance levels for AI-based technologies were at a moderate level and that this needed to be supported through structured education [8].

Readiness is defined as a multidimensional construct encompassing an individual's tendency to adopt and use new technology, perceived competence, and openness to change. In the technology readiness literature, this concept has been addressed within a theoretical framework measuring individuals' propensity to embrace new technologies [2]. In the context of AI, readiness extends beyond the willingness to use and encompasses understanding clinical roles, benefit–risk assessment, ethical sensitivity, and self-efficacy. The MAIRS-MS, developed in this direction, is one of the instruments providing validity and reliability evidence for measuring readiness for medical AI [9]. Current studies show that readiness for AI among nursing students is associated with anxiety, literacy, and attitude variables, and that the combined assessment of these constructs is important for educational planning [10, 11].

However, when the literature is examined, two key limitations emerge: (i) existing measures of AI among nursing students often focus on a single construct (e.g., readiness or acceptance alone) or rely on adaptations of instruments developed for other professional groups; (ii) the number of nursing-specific instruments that comprehensively assess multiple dimensions of AI readiness within the same psychometric framework remains limited [4, 8, 9]. Therefore, there is a need for a valid and reliable assessment tool that comprehensively assesses multiple dimensions of AI readiness aligned with nursing education outcomes. The aim of this study was to develop and psychometrically evaluate the Nursing Artificial Intelligence Readiness Scale (NAIRS) for nursing students.

METHODS

Study Design

This methodological study was conducted to develop and psychometrically evaluate a valid and reliable instrument for assessing nursing students' readiness for artificial intelligence (AI) technologies. The study followed established scale development procedures, including item generation, expert review for content validity, pilot testing, item analysis, exploratory factor analysis, confirmatory factor analysis, and reliability testing. The research was conducted following the scale development steps recommended by DeVellis (2017) and Tavşancıl (2018) [12, 13].

Study Population and Sample

The study population comprised all students enrolled in the nursing department of a public university during the 2025–2026 academic year. A total of 416 nursing students who voluntarily agreed to participate and completed the data collection forms in full were included. In scale development studies, a sample size of at least 5–10 times the number of items is recommended [13, 14]. Because the initial draft scale consisted of 40 items, a minimum of 200–400 participants was targeted, and this criterion was met.

Considering that conducting exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) on the same sample is methodologically inappropriate [15, 16], the full dataset was randomly divided into two subgroups: 200 participants for EFA and 216 participants for CFA. After data collection was completed, participants were assigned a random number using a computer-generated randomization procedure and sorted accordingly. The first 200 cases were allocated to the EFA subsample and the remaining 216 cases to the CFA subsample, ensuring that each participant had an equal probability of being assigned to either group. This split allowed the structure to be verified in an independent sample through cross-validation. The comparability of the two subsamples was examined using an independent-samples t-test for age and chi-square tests for sex, academic year, previous AI training, and frequency of AI use; no statistically significant differences were found (all $p > .05$). Additionally, to evaluate test–retest reliability, the scale was re-administered after a two-week interval to 45 randomly selected participants from the EFA subsample. For the test–retest procedure, participants from the EFA subsample were assigned random numbers using the same computer-generated randomization approach, and 45 students were selected for re-assessment. All selected participants completed the second administration after the two-week interval.

Data Collection Tools

Item generation, content validity, and pilot testing. An initial item pool was generated based on a review of the literature on artificial intelligence readiness, technology readiness, digital health literacy, and the use of AI in nursing and healthcare education. The preliminary draft consisted of 40 items representing four theoretically defined domains: During item generation, particular attention was given to ensuring that the items reflected nursing-specific educational and clinical contexts rather than general technology acceptance. The initial pool intentionally included overlapping and conceptually similar items to allow subsequent refinement during expert evaluation and psychometric testing. Examples of the initial items addressed topics such as understanding the basic principles of AI, confidence in using AI-supported systems, willingness to incorporate AI into future nursing practice, and awareness of ethical issues related to AI-assisted decision-making. To assess content validity, the draft items were reviewed by 10 faculty experts specializing in nursing sciences and measurement/evaluation. The experts evaluated each item in terms of relevance, clarity, comprehensibility, and representativeness using a three-point rating scale. Based on these evaluations, the mean content validity ratio (CVR) was .88 [17]. Following expert review, five items were removed from the initial pool. These items were excluded because they were judged to be conceptually redundant with other items, insufficiently representative of the target construct, or ambiguously worded. Several remaining items were revised to improve clarity, linguistic precision, and consistency with the theoretical definitions of the four domains. Following expert review, the revised draft scale was pilot tested with 50 nursing students to assess item clarity, comprehensibility, readability, and the overall administration process. The pilot participants were also invited to provide qualitative feedback regarding item wording, interpretation, and ease of completion. Although no items were removed during the pilot phase, minor wording and formatting revisions were made to improve readability and reduce potential misunderstandings. Based on feedback obtained during the pilot application, minor wording and formatting revisions were made before the main data collection.

Personal Information Form. This form was prepared by the researchers to collect information regarding participants' age, sex, academic year, prior AI-related training, and frequency of AI use.

Nursing Artificial Intelligence Readiness Scale (NAIRS). The draft scale was developed as a total of 40 items across four theoretical dimensions through a comprehensive review of the relevant literature, evaluation of existing scales, and expert consultation: (a) Knowledge/Awareness (10 items), (b) Willingness to Use AI (10 items), (c) Self-efficacy (10 items), and (d) Ethical Awareness (10 items). Items were rated on a 5-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Undecided, 4 = Agree, 5 = Strongly Agree). Total scores on the final 20-item scale range from 20 to 100, with higher scores indicating higher AI readiness.

Data Collection

Data were collected online during the 2025–2026 fall semester. Participants were provided with detailed information about the purpose and procedures of the study, and voluntary informed consent was obtained. Only students who provided informed consent were included. Ethical approval was obtained from the relevant university's ethics committee. The final version used in the main application was established after the completion of expert review and pilot testing procedures.

Data Analysis

Data were analyzed using descriptive statistics, item analysis, exploratory factor analysis, confirmatory factor analysis, convergent and discriminant validity analyses, and reliability analyses. Statistical analyses were performed in Python 3.10 using the factor_analyzer, semopy, scipy, and pandas libraries.

Item analysis. Item discrimination was evaluated using three criteria: (1) corrected item–total correlation ($r \geq .30$ cutoff [18]), (2) differences between the lower 27% and upper 27% groups (independent samples t-test), and (3) changes in Cronbach's alpha if an item was deleted. Items with corrected item–total correlations below .30, those showing no significant difference in upper–lower group comparisons, or those whose removal increased the alpha value were eliminated.

Construct validity. Before EFA, the suitability of the data for factor analysis was examined using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. The number of factors was determined using parallel analysis [19], the Kaiser criterion (eigenvalue > 1), and inspection of the scree plot. The minimum residual (minres/ordinary least squares) extraction method was used, and Promax (oblique) rotation was selected. Promax rotation allows the correlations between factors that are theoretically expected to be related to be freely estimated [20]. Items were primarily evaluated using loadings of .40 or higher on their primary factor and absence of cross-loadings above .32 on more than one factor; however, items with loadings between .35 and .39 were considered for retention when they were theoretically relevant, did not show problematic cross-loadings, and were supported by acceptable corrected item–total correlations and satisfactory CFA loadings [21].

To confirm the identified structure, first-order CFA was performed on the independent CFA subsample ($n = 216$) using maximum likelihood (ML) estimation. Prior to CFA, the dataset was screened for missing data and distributional assumptions. Only questionnaires that were completed in full were included in the analyses; therefore, no missing data were present in the final dataset and no imputation procedures were required. The distributional characteristics of the observed variables were examined using skewness and kurtosis statistics. Skewness values ranged from –0.94 to 0.12 and kurtosis values ranged from –0.82 to 0.29, indicating no severe deviation from normality [22] and the results supported the use of maximum likelihood estimation. In addition, item distributions were examined for extreme asymmetry and potential floor or ceiling effects, and no problematic patterns were identified. These findings further supported the appropriateness of maximum likelihood estimation for the CFA analyses. Model fit was evaluated using multiple fit indices: χ^2/df ratio, Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Goodness-of-Fit Index (GFI), Adjusted Goodness-of-Fit Index (AGFI), Normed Fit Index (NFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Reference values from the relevant literature were used as acceptable and good fit criteria [23, 24, 25].

Convergent and discriminant validity. Convergent validity was assessed using Composite Reliability (CR ≥ .70) and Average Variance Extracted (AVE ≥ .50). Discriminant validity was evaluated using the Fornell–Larcker criterion, requiring each subscale's √AVE to exceed its correlations with other subscales [26, 27].

Reliability. Internal consistency reliability was calculated using Cronbach's alpha coefficients separately in both subsamples (EFA and CFA). Temporal stability was assessed in the 45-participant subgroup using the Intraclass Correlation Coefficient (ICC; two-way random, single measures) and Pearson correlation coefficients from two measurements taken two weeks apart. Paired-samples t-tests were used to compare the means between the two measurements. The classification proposed by Koo and Li (2016) was used for interpreting ICC values: <.50 poor, .50–.75 moderate, .75–.90 good, >.90 excellent [28].

Ethics approval and consent to participate

Ethical approval was obtained from the Non-Interventional Research Ethics Committee of Gaziantep University (Session Date: 28/05/2025; Decision No: 2025/179). The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. All participants provided informed consent prior to participation.

RESULTS

Content validity and pilot testing. Before the main application, the draft items were reviewed by 10 faculty experts specializing in nursing sciences and measurement/evaluation. Using Lawshe's (1975) three-point format, experts rated each item as essential, useful but not essential, or not necessary. Item-level content validity ratios were calculated and averaged across items; the mean CVR was .88 [17]. In addition, the experts provided qualitative feedback regarding clarity, comprehensibility, and representativeness, and wording revisions were made accordingly.

Following revisions based on expert feedback, the scale was pilot tested with 50 nursing students, and minor wording and formatting changes were made prior to the main administration. The 50 students who participated in the pilot testing were not included in the main psychometric analyses.

Participant Characteristics

A total of 416 nursing students participated in the study. The mean age of the participants was 20.95 ± 3.59 (min = 18, max = 48). Of the participants, 53.4% (n = 222) were female and 46.6% (n = 194) were male. Regarding academic year distribution, first-year students constituted the largest group (53.8%, n = 224). Of the participants, 32.5% (n = 135) reported having received AI-related training, while 67.5% (n = 281) had not. Regarding frequency of AI use, 34.6% (n = 144) reported "sometimes," 26.9% (n = 112) "often," and 15.6% (n = 65) "very often" (Table 1).

Table 1 presents the sociodemographic characteristics of the participants. The EFA and CFA subsamples did not differ significantly in terms of age ($t = -0.516$, $p = .606$), sex ($\chi^2 = 0.121$, $p = .728$), academic year ($\chi^2 = 2.127$, $p = .546$), previous AI training ($\chi^2 = 0.112$, $p = .738$), or frequency of AI use ($\chi^2 = 4.609$, $p = .330$).

Item Analysis

Starting from an initial pool of 40 items, 5 items were removed during the expert review phase based on content validity evaluations, reducing the pool to 35 items. The 35-item version was then pilot tested with 50 nursing students; no items were removed at this stage, but minor wording adjustments were made. Subsequently, item reduction procedures were conducted using the EFA subsample (n = 200). Items were evaluated according to multiple criteria, including corrected item–total correlations, factor loadings, cross-loadings, communality values, and theoretical consistency with the intended construct. Fifteen items were removed because they demonstrated corrected item–total correlations below the recommended threshold, substantial cross-loadings on multiple factors, low communality values, or insufficient conceptual fit with the emerging factor structure, resulting in the final 20-item scale. The corrected item–total correlations of the 20 items in the final scale ranged from .361 to .614, all exceeding the .30 cutoff [18]. The differences between the lower 27% and upper 27% groups were statistically significant for all items ($p < .001$). Removal of any single item did not increase the total Cronbach's alpha coefficient (.880). These findings indicate that all items in the final scale have adequate discriminative power and make meaningful contributions to the scale.

Exploratory Factor Analysis

Before EFA, the suitability of the data for factor analysis was assessed. The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was .858, indicating that the sample size was "good" for factor analysis [29]. Bartlett's test of sphericity was statistically significant ($\chi^2(190) = 1742.04$, $p < .001$), confirming that the correlation matrix was not an identity matrix and that factor analysis was applicable.

When parallel analysis and the Kaiser criterion (eigenvalue > 1) were evaluated together, a four-factor structure was identified as appropriate. The eigenvalues of the first four factors were 6.313, 2.417, 1.757, and 1.595, respectively, while the fifth factor's eigenvalue (0.937) was below 1. According to the pattern matrix obtained with Promax rotation, 20 items clustered under four factors, explaining 52.08% of the total variance. This value meets the 40–60% range expected in the social sciences [13]. Primary factor loadings ranged from .351 to .909. Although a small number of items showed relatively modest loadings, these items were retained because they were conceptually relevant, did not exhibit problematic cross-loadings, and demonstrated acceptable corrected item–total correlations together with satisfactory standardized loadings in the CFA model (Table 4). In particular, items M30 and M38 showed relatively lower factor loadings (.351 and .386, respectively) and lower communality values compared with the other items. However, both items were retained because they did not exhibit problematic cross-loadings, demonstrated acceptable corrected item–total correlations (M30 = .461; M38 = .361), and loaded satisfactorily on their respective latent factors in the CFA model (M30 = .691; M38 = .798). Furthermore, these items represented conceptually important aspects of the self-efficacy and ethical awareness domains that were not fully captured by the remaining items. Removing either item did not result in a meaningful improvement in internal consistency reliability or overall model fit. Therefore, both statistical evidence and theoretical considerations supported their retention in the final scale.

Table 2 shows that the four-factor structure is fully consistent with the theoretical framework. The five items under Factor 1 (Knowledge/Awareness; M1, M4, M6, M8, M9) measure knowledge of AI fundamentals and healthcare applications. The items under Factor 2 (Willingness to Use AI; M11, M12, M13, M15, M17) reflect willingness and attitudes toward using AI technologies. The Factor 3 (Self-efficacy) items (M21, M26, M28, M29, M30) measure self-confidence and perceived competence in using AI systems. The Factor 4 (Ethical Awareness) items (M31, M35, M38, M39, M40) assess awareness of ethical values in AI use.

The similar contribution of each factor to the total variance (12.76%–13.25%) indicates a balanced distribution of subscale weights within the scale. The total explained variance of 52.08% demonstrates that the scale adequately represents the construct. The moderate inter-factor correlations (.22–.51) obtained after Promax rotation support that the dimensions are related but distinguishable constructs.

<u>Insert Table 2</u>

Confirmatory Factor Analysis

The four-factor structure obtained from EFA was tested through first-order CFA on the independent CFA subsample ($n = 216$). Model fit indices are presented in Table 3.

Table 3 shows that the model demonstrated good fit to the data. The χ^2/df ratio (1.426) being below 2 indicates excellent model fit [20]. CFI (.971) and TLI (.967) exceeded .95, reflecting good fit. GFI (.910) was above the .90 threshold. AGFI (.896) was at an acceptable level. NFI (.910) exceeded the acceptable cutoff. Both RMSEA (.044) and SRMR (.049) were below .05, indicating good fit [23]. The simultaneous achievement of acceptable or good fit across all indices strongly supports the confirmation of the four-factor structure identified in EFA.

Standardized factor loadings ranged from .626 to .838 for Knowledge/Awareness, .738 to .832 for Willingness to Use AI, .691 to .810 for Self-efficacy, and .758 to .833 for Ethical Awareness. All standardized factor loadings being above .50 and statistically significant ($p < .001$) indicate that each item adequately represents its latent variable [27]. The CFA path diagram is presented in Figure 1.

<u>Insert Table 3 and Figure 1 here</u>

Item Statistics, Reliability, and Convergent Validity

Item-level statistics, internal consistency coefficients, composite reliability (CR), and average variance extracted (AVE) values for the final 20-item scale are presented in Table 4.

Table 4 shows that the corrected item–total correlations of all items ranged from .361 to .614, all exceeding the .30 cutoff. This indicates that each item is consistent with the measured construct [18].The finding that removal of any item did not increase the total Cronbach's alpha (.880) confirms that all items make meaningful contributions to the scale.

Standardized factor loadings (λ) from CFA ranged from .626 to .838, all above the .50 threshold. The highest factor loadings were observed for M8 (.838) and M9 (.797) in Knowledge/Awareness, M13 (.832) and M11 (.826) in Willingness to Use AI, M21 (.810) and M28 (.788) in Self-efficacy, and M40 (.833) and M35 (.826) in Ethical Awareness.

Regarding internal consistency reliability, subscale alpha values ranged from .801 to .828 in the EFA subsample (n = 200) and from .858 to .902 in the CFA subsample (n = 216), all exceeding the .70 threshold [30]. Composite reliability (CR) values ranged from .860 to .904, exceeding the .70 threshold. All AVE values (.555–.652) were above the .50 threshold, demonstrating that convergent validity was established [27].

<u>Insert Table 4</u>

Discriminant Validity

Discriminant validity was evaluated using the Fornell–Larcker criterion. According to this criterion, each subscale's √AVE must exceed its correlations with other subscales [26]. Results are presented in Table 5.

Table 5 shows that each subscale's √AVE (diagonal values: .745–.808) exceeded its correlations with other subscales (.261–.689). The √AVE of Knowledge/Awareness (.745) exceeded its highest correlation with Self-efficacy (.575). Similarly, the √AVE of Willingness to Use AI (.794) exceeded its highest correlation with Self-efficacy (.689). The Fornell–Larcker criterion being met across all comparisons confirms that the four subscales form distinguishable constructs and that the scale's discriminant validity is supported [26].

Examination of inter-factor correlations revealed that the highest correlation was between Willingness to Use AI and Self-efficacy (.689). This finding may be interpreted as students who are willing to use AI also feeling competent in using it. The moderate-to-high correlation between Knowledge/Awareness and Self-efficacy (.575) indicates that AI knowledge is associated with self-efficacy perception. The lowest correlation being observed between Ethical Awareness and

Knowledge/Awareness (.261) suggests that ethical sensitivity is a relatively independent construct from knowledge level.

Insert Table 5

Test–Retest Reliability

To evaluate the temporal stability of the scale, it was re-administered to 45 randomly selected participants from the EFA subsample after a two-week interval. Results are presented in Table 6.

Table 6 shows that ICC values for all subscales and the total scale ranged from .896 to .963. According to the classification proposed by Koo and Li (2016), the ICC value of Knowledge/Awareness (.896) corresponds to "good" reliability, while Willingness to Use AI (.910), Self-efficacy (.963), Ethical Awareness (.902), and the total scale (.952) correspond to "excellent" reliability [28].

Pearson correlation coefficients ranged from .901 to .962, all statistically significant ($p < .001$), indicating strong consistency between the two measurements. Paired-samples t-test results showed no statistically significant difference between the first and second measurement means for any subscale or total score (all $p > .05$). The closeness of test and retest means (e.g., total scale: 72.27 vs. 71.78) confirms that the scale produced consistent measurements over the two-week period.

Taken together, these findings indicate that the NAIRS provides temporally stable and reliable measurements. The consistency of the scale when administered at different time points constitutes important evidence of measurement reliability.

Insert Table 6

Summary of Validity and Reliability Findings

When all findings regarding the psychometric properties of the Nursing Artificial Intelligence Readiness Scale (NAIRS) are considered together: (1) In terms of construct validity, the four-factor structure identified through EFA was confirmed by CFA, and all fit indices met acceptable or good fit criteria; (2) In terms of convergent validity, all CR (.860–.904) and AVE (.555–.652) values exceeded the threshold values; (3) In terms of discriminant validity, the

Fornell–Larcker criterion was met across all comparisons; (4) In terms of internal consistency reliability, Cronbach's alpha values exceeded acceptable levels in both the EFA and CFA subsamples; (5) In terms of test–retest reliability, ICC values indicated good to excellent stability. These findings demonstrate that the NAIRS is a valid and reliable instrument for assessing nursing students' AI readiness levels.

DISCUSSION

In this study, the psychometric properties of the scale developed to assess nursing students' readiness for artificial intelligence were examined, and the findings demonstrated that the instrument is valid and reliable. The increasing adoption of AI-based technologies in healthcare has made it increasingly important to evaluate healthcare professionals' readiness, competencies, and perceptions regarding these technologies. AI-supported clinical decision support systems and data-driven care models are transforming healthcare delivery and increasing nurses' interaction with these technologies [4, 31]. Therefore, determining nursing students' readiness for artificial intelligence is important for developing nursing education to adapt to digital health environments [32].

The exploratory factor analysis conducted in this study revealed that the scale exhibited a four-factor structure explaining 52.08% of the total variance. Factor loadings ranging from .351 to .909 indicate that the items demonstrated adequate associations with their respective factors. In scale development studies, a total explained variance of 40–60% is generally considered acceptable [27]. In this regard, the explained variance ratio obtained in this study can be considered adequate. Moreover, the identification of multidimensional factor structures in scale development studies allows for a more comprehensive assessment of different dimensions of the measured construct [12]. Similar multidimensional factor structures have also been reported in scale development studies examining technology acceptance and digital health competencies in healthcare [33, 34]. From this perspective, the four-factor structure obtained in this study supports AI readiness being a multidimensional construct encompassing different dimensions such as knowledge/awareness, willingness to use AI, self-efficacy, and ethical awareness. The literature also reports that instruments developed for AI similarly exhibit multidimensional structures. For example, the Medical Artificial Intelligence Readiness Scale for Medical Students (MAIRS-MS) developed by Karaca et al. (2021) defined medical AI readiness as a multidimensional construct comprising cognitive, affective, and skill-based dimensions [9]. These findings indicate that the factor structure obtained in this study is consistent with the conceptual framework described in the literature.

The confirmation of the structure identified through EFA by CFA provides important evidence for the construct validity of the scale. In scale development studies, confirming the factor structure obtained through EFA with CFA is considered one of the fundamental methodological approaches for evaluating whether the theoretical structure of the instrument holds across different samples [16, 24]. The CFA results obtained in this study showed that the model fit indices were within acceptable limits. In particular, the CFI value of .971 and the RMSEA value of .044 demonstrate good fit between the proposed model and the data. In structural equation modeling studies, CFI and TLI values above .90 and RMSEA values below .08 and SRMR values below .10 are considered acceptable for model fit [23, 25].

When reliability analyses are examined, the internal consistency coefficients of the scale are high. Subscale Cronbach's alpha coefficients ranged from .801 to .828 in the EFA subsample and from .858 to .902 in the CFA subsample, with the total scale alpha being .911. In the literature, a Cronbach's alpha above .70 is considered adequate and above .80 is considered high internal consistency [27, 30]. In this regard, the obtained alpha values indicate that the scale has strong internal consistency. Furthermore, ICC values from test–retest analysis ranging from .896 to .963, with .952 for the total scale, demonstrate that the scale produces temporally stable measurements [28].

The four-factor structure observed in the present study is broadly consistent with prior work suggesting that readiness for medical AI is a multidimensional construct. In particular, the MAIRS-MS emphasized cognitive, skill-related, and attitudinal dimensions among medical students. The present study extends this line of work by proposing a nursing-specific structure that also explicitly incorporates ethical awareness, which is especially relevant in nursing because AI-supported decision-making in care settings raises questions related to patient safety, responsibility, confidentiality, and professional judgment.

The inclusion of an ethical awareness dimension is a notable contribution of the scale. In nursing education, readiness for AI should not be conceptualized solely as willingness to adopt technology or confidence in using it. Students are also expected to recognize the ethical implications of AI-assisted recommendations, data use, accountability, and the limits of automated systems in patient care.

Although a small number of items (e.g., M30 and M38) showed relatively lower factor loadings in the EFA solution, they were retained because they demonstrated acceptable corrected item–total correlations, satisfactory standardized CFA loadings, and meaningful contributions to content coverage within their respective domains. Their retention was guided not only by statistical criteria but also by theoretical considerations, as the removal of these items would have reduced the breadth of the self-efficacy and ethical awareness constructs represented in the scale. Nevertheless, future studies should continue to evaluate the performance of these items in different samples and consider potential refinement if consistently lower factor loadings are observed.

The inclusion of expert review and pilot testing during the development phase strengthened the content relevance, clarity, and comprehensibility of the items before psychometric evaluation.

The findings of this study have important implications for nursing education. The NAIRS may be used by nursing educators to identify strengths and gaps in students' readiness for AI across the domains of knowledge, willingness to use AI, self-efficacy, and ethical awareness. The scale may also support curriculum development by helping educators evaluate the effectiveness of AI-related educational interventions and monitor changes in students' readiness over time. As AI becomes increasingly integrated into clinical decision support systems, documentation processes, and digital health applications, nursing curricula should incorporate structured learning opportunities that address not only technical competencies but also ethical and professional aspects of AI use. Recent evidence on AI-supported applications, chatbots, and simulation in nursing education further supports the need for pedagogically structured, critically oriented AI learning experiences [4, 35, 36]. Accordingly, the assessment of nursing students' AI readiness may provide an important evaluative framework for educational planning and research examining the integration of digital health technologies into nursing education. Unlike instruments adapted from other disciplines, the NAIRS was developed specifically for the nursing context and may serve as a practical tool for evaluating AI readiness within nursing education programs.

Limitations

This study has several limitations that should be considered when interpreting the findings. First, the data were collected using self-report measures, which may be subject to response bias and social desirability effects. Second, the study employed a cross-sectional design, which does not allow for the evaluation of changes in AI readiness over time. Third, external criterion-related validity was not examined using an established comparison instrument, which limits the evidence for convergent validity with related constructs. Fourth, measurement invariance across sex, academic year, or prior AI-related training was not tested; therefore, the equivalence of the scale across subgroups remains to be established. Finally, the study was conducted at a single public university in Türkiye, which may limit the generalizability of the findings to nursing students from other institutions and educational settings. In addition, more than half of the participants were first-year students. Their limited clinical experience may have influenced their perceptions and understanding of the clinical applications of artificial intelligence. Future studies should evaluate the scale in more diverse samples with a more balanced distribution across academic years and educational settings.

Conclusion and Future Directions

Taken together, these findings provide initial evidence that the NAIRS is a valid and reliable instrument for assessing nursing students' readiness for artificial intelligence across four related but distinguishable domains: knowledge/awareness, willingness to use AI, self-efficacy, and ethical awareness. The scale may be useful for identifying educational needs, monitoring readiness profiles, and informing curriculum development in nursing education. Future studies should examine the scale in multicenter samples, test its measurement invariance across student subgroups, and explore its associations with related constructs such as AI literacy, AI anxiety, and technology acceptance.

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from the Non-Interventional Research Ethics Committee of Gaziantep University (Session Date: 28/05/2025; Decision No: 2025/179). The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. All participants provided informed consent prior to participation.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no known competing interests.

Funding

No external funding was received for this study.

Authors' Contributions

SA: Conceptualization, Methodology, Supervision, Writing – Review & Editing

SB: Conceptualization, Methodology, Formal Analysis, Data Curation, Writing – Original Draft

GK: Conceptualization, Data Collection, Investigation, Writing – Original Draft

Acknowledgements

The authors would like to thank all nursing students who voluntarily participated in this study.

References

1. Torous J, Linardon J, Goldberg SB, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. *World Psychiatry*. 2025;24(2):156-174. doi:10.1002/wps.21299

2. Parasuraman A. Technology Readiness Index (Tri). *J Serv Res*. 2000;2(4):307-320. doi:10.1177/109467050024001

3. Çevik Aktura S, Dertli S. Readiness of Nurses and Nursing Students to Use Artificial Intelligence in Healthcare and Influencing Factors. *İzmir Katip Çelebi Üniversitesi Sağlık Bilim Fakültesi Derg*. 2026;11(1):105-112. doi:10.61399/ikcusbfd.1640227

4. Ronquillo CE, Peltonen LM, Pruinelli L, et al. Artificial intelligence in nursing: Priorities and

opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative. *J Adv Nurs*. 2021;77(9):3707-3717. doi:10.1111/jan.14855

5. von Gerich H, Moen H, Block LJ, et al. Artificial Intelligence-based technologies in nursing: A scoping literature review of the evidence. *Int J Nurs Stud*. 2022;127. doi:10.1016/j.ijnurstu.2021.104153

6. Van Bulck L, Couturier R, Moons P. Applications of artificial intelligence for nursing: has a new era arrived? *Eur J Cardiovasc Nurs*. 2023;22(3):E19-E20. doi:10.1093/eurjcn/zvac097

7. Salameh B, Qaddumi J, Hammad B, et al. Nursing Students' Attitudes Toward Artificial Intelligence: Palestinian Perspectives. *SAGE Open Nurs*. 2025;11. doi:10.1177/23779608251343297

8. Al-Olaimat K, Salameh B, Abdulhalim Alqadi R, et al. Readiness and Acceptance of Nursing Students Regarding AI-Based Health Care Technology on the Training of Nursing Skills in Saudi Arabia: Cross-Sectional Study. *JMIR Nurs*. 2025;8(1):e71653. doi:10.2196/71653

9. Karaca O, Çalışkan SA, Demir K. Medical artificial intelligence readiness scale for medical students (MAIRS-MS) - development, validity and reliability study. *BMC Med Educ*. 2021;21(1). doi:10.1186/s12909-021-02546-6

10. Burucu R, Türkben Polat H. Nursing students' readiness and anxiety regarding medical artificial intelligence: A mixed-methods study. *Nurse Educ Pract*. 2025;88. doi:10.1016/j.nepr.2025.104538

11. Boztepe H, Akdeniz Kudubeş A, Semerci Şahin R, Durmuş Sarıkahya S, Çınar Özbay S. Examining the relationship between nursing students' readiness, literacy and attitudes toward medical artificial intelligence. *Nurse Educ Pract*. 2025;88. doi:10.1016/j.nepr.2025.104568

12. DeVellis RF. *Scale Development: Theory and Applications*. 3rd ed. SAGE Publications;

2017.

13. Tavşancıl E. *Tutumların Ölçülmesi ve SPSS Ile Veri Analizi*. 6th ed. Nobel yayıncılık; 2018.

14. Comrey AL, Lee HB. *A First Course in Factor Analysis*. Vol 2nd ed. Lawrence Erlbaum Associates; 1992. doi:10.4324/9781315827506

15. Worthington RL, Whittaker TA. Scale Development Research. *Couns Psychol*. 2006;34(6):806-838. doi:10.1177/0011000006288127

16. Fabrigar LR, Wegener DT, Maccallum RC, Strahan EJ. Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychol Methods*. 1999;4(3):272-299.

17. Lawshe CH. A quantitative approach to content validity. *Pers Psychol*. 1975;28(4):563-575. doi:10.1111/j.1744-6570.1975.tb01393.x

18. Buyukozturk S. Handbook of Data Analysis for Social Sciences: Statistics, Research Design, SPSS Applications, and Interpretation. 27th ed. Pegem Akademi; 2018.

19. Horn JL. A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika*. 1965;30(2):179-185. doi:10.1007/BF02289447

20. Tabachnick BG, Fidell LS. *Using Multivariate Statistics (6th ed.)*. Pearson; 2013.

21. Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pract Assessment, Res Eval*. 2005;10(1). doi:10.7275/JYJ1-4868

22. Curran PJ, West SG, Finch JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychol Methods*. 1996;1(1):16-29. doi:10.1037/1082-989X.1.1.16

23. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6(1):1-55. doi:10.1080/10705519909540118

24. Kline RB. *Principles and Practice of Structural Equation Modeling*. 4th ed. Guilford Press.;

2016.

25. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods Psychol Res*. 2003;8(2):23-74. Accessed March 3, 2026. https://www.stats.ox.ac.uk/~snijders/mpr_Schermelleh.pdf

26. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J Mark Res*. 1981;18(1):39-50. doi:10.1177/002224378101800104

27. Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate Data Analysis*. 7th ed. Pearson; 2010.

28. Koo TK, Li MY, Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

29. Kaiser HF. An index of factorial simplicity. *Psychometrika*. 1974;39(1):31-36. doi:10.1007/BF02291575

30. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. MacGraw-Hill; 1994.

31. Maleki Varnosfaderani S, Forouzanfar M. The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century. *Bioeng 2024, Vol 11,*. 2024;11(4). doi:10.3390/bioengineering11040337

32. Buchanan C, Howitt ML, Wilson R, Booth RG, Risling T, Bamford M. Predicted Influences of Artificial Intelligence on Nursing Education: Scoping Review. *JMIR Nurs*. 2021;4(1):e23933. doi:10.2196/23933

33. Van Der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. *J Med Internet Res*. 2017;19(1):e27. doi:10.2196/jmir.6709

34. Chun H, Park EJ, Choi SK, Yoon H, Okan O, Dadaczynski K. Validating the Digital Health

Literacy Instrument in Relation to COVID-19 Information (COVID-DHL-K) among South Korean Undergraduates. *Int J Environ Res Public Health*. 2022;19(6):3437. doi:10.3390/ijerph19063437

35. Ozdemir S, Akcoban S, Berse S, Tosun B. The effect of artificial intelligence-supported applications on critical thinking motivation and clinical decision making in nursing students: path analysis. BMC Med Educ. 2026;26:602. doi:10.1186/s12909-026-08829-0

36. Cakir B, Berk A, Berse S, Gov P, Gungor MI. Chatbot-centered integration of simulation and artificial intelligence in nursing education: a narrative review. Glob Nurs Insights. 2025;12(3):290-300. doi:10.31125/globnursinsights.1762516

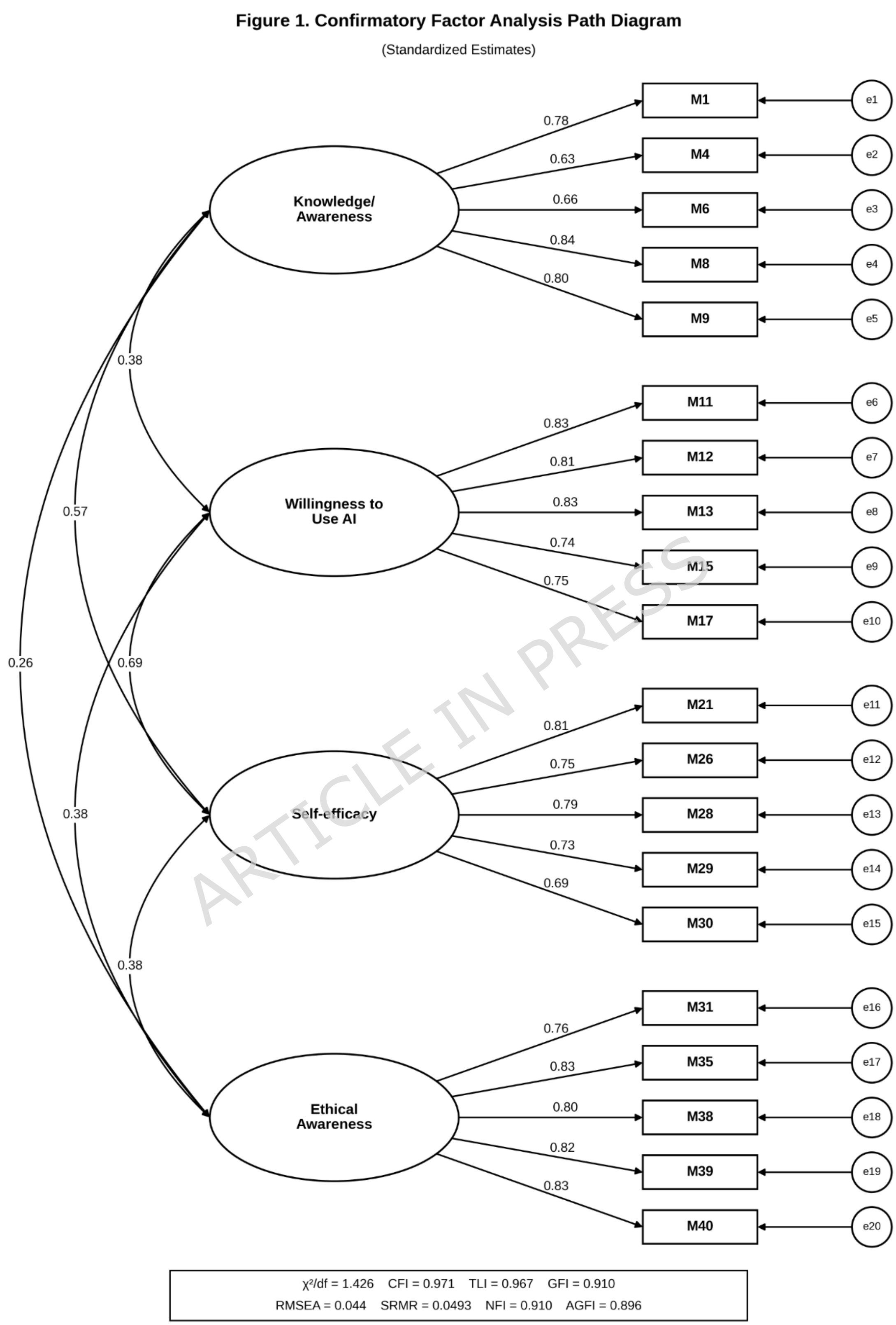

Figure 1. Confirmatory Factor Analysis Path Diagram (Standardized Estimates)

Table 1. *Sociodemographic Characteristics of the Participants (n = 416)*

Variable	**Category**	**n**	**%**	**Mean ± SD**
Sex	Female	222	53.4	
	Male	194	46.6	
Academic Year	1st year	224	53.8	
	2nd year	59	14.2	
	3rd year	73	17.5	
	4th year	60	14.4	
AI Training	Yes	135	32.5	
	No	281	67.5	
Frequency of AI Use	Never	23	5.5	
	Rarely	72	17.3	
	Sometimes	144	34.6	
	Often	112	26.9	
	Very often	65	15.6	
Age				20.95±3.59

Table 2. *Pattern Matrix from Exploratory Factor Analysis with Promax Rotation (n = 200)*

Item	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**h²**
	(Knowledge)	*(Willingness to Use AI)*	*(Self-efficacy)*	*(Ethics)*	
M1	**.812**	.020	-.061	.134	.681
M4	**.584**	.026	.025	-.049	.345
M6	**.460**	.008	.221	-.057	.263
M8	**.681**	.105	.069	-.017	.480
M9	**.849**	-.067	-.089	.052	.736
M11	-.065	**.805**	.111	-.068	.668
M12	.144	**.703**	-.027	-.016	.516
M13	.051	**.909**	-.121	-.104	.854

M15	-.223	**.530**	.135	.175	.379
M17	.173	**.485**	-.017	-.068	.270
M21	-.043	.128	**.773**	-.018	.615
M26	-.015	.030	**.827**	-.040	.686
M28	-.024	-.039	**.871**	-.105	.772
M29	.064	-.085	**.567**	.061	.337
M30	.094	.032	**.351**	.159	.159
M31	-.103	.104	.049	**.565**	.343
M35	.014	-.192	.003	**.823**	.715
M38	.040	.237	-.125	**.386**	.222
M39	-.006	-.131	.068	**.853**	.750
M40	.143	-.036	-.059	**.774**	.624
Eigenvalue	**6.313**	**2.417**	**1.757**	**1.595**	
Variance %	**13.25**	**13.11**	**12.96**	**12.76**	
Total variance explained (%)	**52.08**				

Note. Factor loadings ≥.35 are shown in bold. Extraction method: Minres; Rotation: Promax. KMO = .858; Bartlett $\chi^2(190) = 1742.04$, $p < .001$.

Table 3. *Confirmatory Factor Analysis Fit Indices (n = 216)*

Fit Index	**Value**	**Acceptable Fit**	**Good Fit**
χ^2/sd	1.426	2-5	<2
CFI	.971	≥.90	≥.95
TLI	.967	≥.90	≥.95
GFI	.910	≥.85	≥.90
AGFI	.896	≥.80	≥.90
NFI	.910	≥.90	≥.95
RMSEA	.044	≤.08	≤.05
SRMR	.049	≤.10	≤.05

Note. Reference values: Hu & Bentler (1999), Kline (2016), Schermelleh-Engel et al. (2003).

Table 4. *Item Analysis, Reliability, Construct Reliability, and Convergent Validity Results*

Factor/Item	**X̄**	**SD**	**rjx**	**α if deleted**	**λ (CFA)**	**α EFA**	**α CFA**	**CR**	**AVE**
Knowledge/Awareness						***.828***	***.858***	***.860***	***.555***
M1	3.52	0.91	.608	.871	.779				
M4	3.39	1.06	.400	.877	.626				
M6	3.48	0.97	.452	.875	.662				
M8	3.42	1.05	.583	.871	.838				
M9	3.53	1.06	.478	.874	.797				
Willingness to Use AI						***.814***	***.893***	***.895***	***.630***
M11	3.48	1.00	.585	.871	.826				
M12	3.44	0.99	.588	.871	.814				
M13	3.56	0.99	.535	.873	.832				
M15	3.66	1.03	.444	.876	.738				
M17	3.62	1.12	.425	.876	.753				
Self-efficacy						***.818***	***.868***	***.869***	***.572***
M21	3.42	1.06	.614	.870	.810				
M26	3.55	1.03	.581	.871	.751				
M28	3.69	1.00	.506	.874	.788				
M29	3.38	1.08	.437	.876	.735				
M30	3.35	1.02	.461	.875	.691				
Ethical Awareness						***.801***	***.902***	***.904***	***.652***
M31	3.69	1.25	.386	.878	.758				
M35	3.99	1.03	.362	.878	.826				
M38	3.92	1.06	.361	.878	.798				
M39	3.94	1.07	.456	.875	.822				
M40	3.97	1.10	.498	.874	.833				
Scale Total						**.880**	**.911**		

Note. Means and standard deviations were calculated using the full sample (N = 416). Corrected item–total correlations (rjx) and Cronbach's alpha if item deleted were computed from the EFA subsample (n = 200). Standardized factor loadings (λ), composite reliability (CR), and average variance extracted (AVE) were derived from the CFA subsample (n = 216). α EFA and α CFA denote subscale internal consistency coefficients calculated in the EFA and CFA subsamples, respectively.

Table 5. *Inter-Factor Correlations and Fornell–Larcker Discriminant Validity Criterion*

	Knowledge/Awareness	**Willingness to Use AI**	**Self-efficacy**	**Ethical Awareness**
Knowledge/Awareness	**.745**			
Willingness to Use AI	.377	**.794**		
Self-efficacy	.575	.689	**.756**	
Ethical Awareness	.261	.376	.385	**.808**

Note. Diagonal values (bold): √AVE; below-diagonal values: inter-factor correlations from CFA.

Table 6. *Test–Retest Reliability Results (n = 45, 2-Week Interval)*

Subscale	**Test Mean ± SD**	**Retest Mean ± SD**	**r**	**ICC**	**t**	**p**	**Interpretation**
Knowledge/Awareness	17.33±4.29	17.27±3.79	.901	.896	0.240	.811	Good
Willingness to Use AI	17.84±3.77	17.64±3.90	.910	.910	0.822	.415	Excellent
Self-efficacy	17.49±4.54	17.44±4.53	.962	.963	0.240	.811	Excellent
Ethical Awareness	19.60±4.63	19.42±4.24	.904	.902	0.602	.550	Excellent
Total Scale	**72.27±12.69**	**71.78±11.53**	**.956**	**.952**	**0.873**	**.388**	**Excellent**

Note. ICC: Intraclass correlation coefficient (two-way random, single measures). Interpretation criteria (Koo & Li, 2016): <.50 poor, .50–.75 moderate, .75–.90 good, >.90 excellent.