Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Measuring AI Readiness in Health Professions Education: Development and Validation of the AI Competence and Preparedness Evaluation Resource (AICPER)

Divya Kewalramani

dk1362@rwjms.rutgers.edu

Rutgers, The State University of New Jersey

Oluwaseun J. Adeyemi

Rutgers, The State University of New Jersey

Stuti Shah

Rutgers, The State University of New Jersey

Justin L. Benton

Rutgers, The State University of New Jersey

Dylan R. Nieman

Rutgers, The State University of New Jersey

Advaith Bongu

Rutgers, The State University of New Jersey

Nell Maloney Patel

Rutgers, The State University of New Jersey

Archana Pradhan

Rutgers, The State University of New Jersey

Laura Sclafani

Rutgers, The State University of New Jersey

Joseph A. Barone

Rutgers, The State University of New Jersey

Christine Repsha

Rutgers, The State University of New Jersey

Steven Glogocheski

Rutgers, The State University of New Jersey

Angela R. Starkweather

Rutgers, The State University of New Jersey

Laura E. Liang

Rutgers, The State University of New Jersey

Claire Brown
Rutgers, The State University of New Jersey
Perry N. Halkitis
Rutgers, The State University of New Jersey
Amy P. Murtha
Rutgers, The State University of New Jersey
Mayur Narayan
Rutgers, The State University of New Jersey

Research Article

Keywords: artificial intelligence, health professions education, psychometrics, instrument validation, nursing students, pharmacy students, public health students, factor analysis

Posted Date: July 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9955555/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background: Few instruments measuring health professions trainees' perceptions of artificial intelligence (AI) have undergone rigorous psychometric validation in U.S. populations. Existing instruments were not designed to capture an interprofessional perspective and neglect the multidimensional impact of AI on clinical practice, education, and research domains. To address this gap, we developed and validated the Artificial Intelligence in Competence and Preparedness Evaluation Resource (AICPER).

Methods: A cross-sectional online survey of students enrolled in nursing, pharmacy, and public health programs was administered at a U.S. university to evaluate the psychometric properties of AICPER. Content validity was assessed by subject-matter experts using a four-point relevance scale. Item Content Validity Index, Scale Content Validity Index (I-CVI, S-CVI), and modified kappa were calculated to quantify item relevance and expert agreement beyond chance. Internal consistency reliability was estimated using Cronbach's alpha, ordinal alpha, and McDonald's omega. Structural validity was examined via Exploratory and Confirmatory Factor Analysis (EFA and CFA), supplemented by a method-effects model for reverse-scored items, configural-vs-metric measurement invariance across programs, bootstrap 95% confidence intervals on factor loadings, and listwise-deletion sensitivity analysis.

Results: Seven instrument experts and 193 students completed the survey. The 20-item scale demonstrated acceptable content validity. EFA recovered a three-factor structure: Functional Benefits, Ethics and Trust, and Educational Engagement. Item and factor analysis resulted in an 18-item instrument which demonstrated good internal consistency at the total-scale level (α=0.86). Metric invariance was supported across programs (ΔCFI = -0.001). All 18 standardized loadings had bootstrap 95% CIs excluding zero. Method effects are present and account for a non-trivial share of variance in reverse-coded items, but the substantive three-factor structure persists when these effects are modeled.

Conclusions: AICPER provides health-professions educators and researchers with a rigorously validated, multidimensional instrument measuring trainee perceptions across functional, ethical, and engagement domains. Its three-domain structure supports targeted curriculum needs assessment, intervention evaluation, and cross-disciplinary benchmarking.

BACKGROUND

Artificial intelligence (AI) is rapidly transforming healthcare delivery, medical education, and biomedical research [1-5]. Advances in machine learning, natural language processing, computer vision, and generative AI have expanded AI use across clinical decision-making, diagnostic imaging, workflow optimization, predictive analytics, training, and scientific writing. Generative AI tools have accelerated trainees’ exposure to AI-assisted technologies in educational and clinical environments [6-8].

Health profession trainees increasingly require competencies that extend beyond basic AI awareness. Understanding how trainees perceive AI, its clinical utility, educational relevance, ethical implications,

and research applications, is a prerequisite to designing needs-aligned training and evaluating whether it produces the intended attitudinal and behavioral changes. Several validated instruments address these perceptions. The Medical AI Readiness Scale for Medical Students (MAIRS-MS) [9] measures four AI-readiness domains (cognition, ability, vision, ethics) and has been cross-culturally revalidated [10-13]. In nursing, two validated Turkish scales for hospital nurses [14] and nursing students [15] have emerged but have not been validated in U.S. samples. A 2025 systematic review of 37 nursing AI studies explicitly identifies this gap [16]. In pharmacy and public health, no validated population-specific AI perception instrument gas been published despite growing curricular interest. Outside medicine, the General Attitudes towards AI Scale (GAAIS) [17] measures general AI attitudes but is not healthcare specific; adjacent constructs such as AI anxiety [18] and AI literacy [19] measure related but distinct dimensions sometimes treated interchangeably with attitudes.

Three distinct gaps remain in measuring health professions learners' perceptions of AI. First, no instrument has been validated in an interdisciplinary sample of students (nursing, pharmacy, public health) where AI is rapidly entering curricula in parallel; smaller surveys in these populations lack formal psychometric validation in U.S. [20,21]. Second, no existing instrument distinguishes perceptions across clinical practice, education, and research as distinct sub-constructs; a trainee may welcome AI clinically yet distrust it in scholarly writing, and without sub-scores, curricula cannot match interventions to concerns. Third, most predate the widespread availability of generative AI, leaving the questions most relevant to current trainees (academic integrity, AI-assisted writing, trust in chatbot outputs) underrepresented.

To address these gaps, we developed **Artificial Intelligence in Competence and Preparedness Evaluation Resource (AICPER)**, a multidomain instrument assessing health professions trainees' perceptions of AI across clinical practice, education, and research. This study assessed its psychometric properties among U.S. health profession trainees, ultimately seeking to enhance AI integration in education.

METHODS

Study Design

An online cross-sectional survey of students enrolled across three health-professions schools at a U.S. University: The School of Nursing, Pharmacy, and Public Health, was conducted. The study was approved by the Institutional Review Board (IRB #Pro2025001398). Reporting follows the COSMIN Reporting Guideline for measurement property studies (Additional File 1) [22].

Study Participants

Eligible participants were students aged 18 or older currently enrolled in nursing, pharmacy, or public health programs, willing to complete an English-language online questionnaire and provide informed

consent. Content experts required an advanced degree (Master's, PhD, or MD) and prior experience in healthcare education, clinical research, or innovation.

Survey Administration

The survey was administered electronically using REDCap (Research Electronic Data Capture), a secure web-based platform [23]. The link was distributed via the official student listserv at each school and remained open for four weeks, with a single reminder email sent during the open period. Participation was voluntary, anonymous, and uncompensated. No incentive was offered for participation. Respondents could skip any item or exit the survey at any time without penalty.

Instrument Development

AICPER items were developed through a targeted literature review of existing instruments and AI competency frameworks in health professions education and refined iteratively by an institutional panel comprising leadership from Schools of Nursing, Pharmacy, and Public Health. The preliminary 20-item instrument (Additional File 2; 10 clinical-practice items A1-A10, six research items A11-A16, four education items A17-A20) was piloted with 10 trainees (not included in the final analytic sample) to assess clarity, response burden, and face validity. Pilot feedback was reviewed asynchronously with the panel and informed final item wording.

Each item was scored on a four-point Likert scale ranging from strongly disagree (1) to strongly agree (4), with an additional "I don't know enough to say (IDK)" option to distinguish non-opinion responses from neutral responses, given the heterogeneity of AI exposure. Negatively worded items (A7, A8, A9, A12, A20) were reverse-scored so that higher scores reflected more favorable perceptions and engagement with AI-related concepts.

Missing Data and Imputation

Two sources of item-level missingness were distinguished: (1) item non-response (i.e. a blank item); and (2) the selection of IDK, which we treated as missing for psychometric analyses because the option was offered to distinguish non-opinion from neutrality. Both were combined into a single missingness indicator for imputation because both are non-substantive with respect to the latent attitude being measured. To characterize self-reported AI exposure heterogeneity separately, we derived an ‘AI Uncertainty Index’ per respondent, defined as the count of IDK responses across the 20 items.

To examine the validity of treating these responses as missing rather than as low-confidence answers, we tested whether (a) Uncertainty Index counts were predicted by self-rated AI understanding, prior AI exposure, training year, and program; and (b) whether Uncertainty Index counts predicted mean AICPER scores after adjusting for the same covariates. Per-item IDK rates were also examined to identify item-content patterns associated with respondent uncertainty.

Missingness was assessed using Little's Missing Completely at Random test and found consistent with data missing completely at random [24]. We addressed missingness using Multiple Imputation by Chained Equations (MICE) [25,26] with Bayesian-ridge regression, generating M=20 imputed datasets. Scalar estimates including Cronbach's α (with Bonett standard errors [27]), confirmatory factor analysis (CFA) standardized factor loadings, factor variances and covariances, and item intercepts, were pooled across imputations using Rubin's rules [28]. CFA fit indices are reported as means and ranges across imputations rather than via formal Rubin's-rules pooling, because pooling them requires procedures outside standard Rubin's application (Procrustes alignment, the Meng-Rubin D2 statistic) [29,30].

Content Validity Analysis (CVA)

Content experts evaluated each AICPER item for relevance on a four-point ordinal scale (1 = not relevant; 4 = highly relevant). Ratings of 3 or 4 were coded relevant (1), and ratings of 1 or 2 not relevant (0). The Item Content Validity Index (I-CVI) was the proportion of experts rating each item as relevant. Inter-rater agreement was estimated using the Polit, Beck and Owen (2007) modified kappa [31]. Acceptability bands were defined as $\kappa \geq 0.74$ excellent, 0.60-0.74 good, 0.40-0.59 fair. Items with $\kappa < 0.40$ were flagged for revision. The Scale Content Validity Index averaged (S-CVI/Ave) was the mean I-CVI across items [32].

Item Analysis and Internal Consistency

Item performance was assessed using floor and ceiling effects (proportions endorsing extreme scale values), item mean (difficulty), SD (variability), and corrected item-total correlation on the MICE-imputed analytic dataset (n=193). Floor and ceiling effects below 20% were considered ideal [33]. Item-total correlations ≥0.30 were considered adequate [34,35]. Internal consistency was assessed using: Cronbach's α [36], ordinal α computed on the polychoric correlation matrix [37], and McDonald's ω-total [38]. Values of 0.70-0.79 were considered adequate, 0.80-0.89 good, and ≥0.90 excellent.

Exploratory Factor Analysis (EFA)

EFA used maximum likelihood extraction with oblique (oblimin) rotation. Factor retention was guided by multiple criteria: Horn's parallel analysis on polychoric eigenvalues (95th percentile over 500 iterations), the scree bend, eigenvalues >1, cumulative variance ≥50%, and conceptual interpretability. To address ordinal scaling of Likert items, a sensitivity EFA used unweighted least squares (ULS) extraction on the polychoric correlation matrix with oblimin rotation. Consistent with prior studies [39-42], we treated loadings of $|<0.32|$ as salient and report loadings $|>0.25|$ as suppressed. Internal consistency for each EFA subscale was assessed using Cronbach's α.

Confirmatory Factor Analysis (CFA)

CFA used diagonally weighted least squares with robust mean- and variance-adjusted test statistics (DWLS/WLSMV) [43]. The baseline three-factor model followed the EFA-recovered structure, with item loading on its EFA-assigned factor and no correlated error terms. We pre-identified four item pairs as candidates for correlated-error specification on substantive content grounds (A4↔A5, workflow burden reduction; A7↔A8, AI failure-mode concerns; A13↔A14, AI in scholarly publishing; A17↔A18, formal AI credentialing) and evaluated each empirically using modification indices (MI) and standardized expected parameter changes (SEPC).

Model fit was evaluated using Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Following Hu and Bentler [44], CFI/TLI ≥ 0.95 and RMSEA ≤ 0.06 indicate good fit; CFI/ TLI 0.90-0.94 and RMSEA 0.06-0.08 indicate acceptable fit. Bootstrap percentile 95% confidence intervals (CI) on each standardized factor loading used 200 resamples of the imputed dataset [45].

Method-Effects Analysis

Five items were reverse-scored (A7, A8, A9, A12, A20). To evaluate whether the factor defined predominantly by reverse-scored items was partly an artifact of reverse-scoring direction [46], an alternative model was fit with a single orthogonal method factor loading the four reverse-scored items retained in the final scale (A7, A8, A9, A12). The three-factor model and the substantive-plus-method model were compared on CFI, TLI, RMSEA, and BIC.

Measurement Invariance

Measurement invariance (i.e. the degree to which a latent construct is measured equivalently across subgroups) was tested across the three programs by comparing two nested models. The configural model freely estimated all parameters within each program; the metric model constrained factor loadings to be equal across programs. Acceptance criteria for metric invariance were ΔCFI ≤0.010 and ΔRMSEA ≤0.015 between models [47,48]. Per-program Cronbach's α was also computed.

Sample Size

Our sample size is adequate for EFA based on recommendations of 5-10 participants per item [49]. For CFA, conventional minima recommend approximately 200 for stable estimation [50]; CFA on our same sample was therefore interpreted as preliminary. CFA evidence was supplemented by bootstrap 95% CI on factor loadings (2000 nonparametric resamples) and a listwise-deletion sensitivity analysis (n=74).

Analyses were conducted using Python 3.10 (Python Software Foundation, Wilmington, DE, USA) with the pandas, NumPy, SciPy, scikit-learn, factor_analyzer, and semopy packages. Key estimates were cross-verified using R version 4.4.2 (R Foundation for Statistical Computing, Vienna, Austria) with the psych, lavaan, and mice packages.

RESULTS

Sociodemographic Characteristics

Seven instrument experts completed content validation. Mean (SD) age was 50.4 (17.7) years; 57.1% male, 57.1% White, and 100.0% non-Hispanic. The panel spanned nursing (n=3, 42.9%), pharmacy (n=1, 14.3%), and medicine (n=3, 42.9%) with median (Q1, Q3) research experience of15.0 (6.5, 20.0) years.

A total of 193 students completed the survey (Table 1). Most were aged 20-30 years (78.8%), identified as female (75.6%), and enrolled in doctoral-level programs (62.7%). Pharmacy students represented 49.7% of the sample, followed by nursing (30.1%) and public health (20.2%). Most reported prior practical exposure to AI (64.2%); 57.5% reported no programming experience.

Missing Data and AI Uncertainty Index

Item non-response was minimal, ranging from 0% to 1.0% per item. Across the 20 items, respondents selected IDK on a mean of 2.05 items (median 1, IQR 0–3, SD 2.44, range 0–10), with no significant differences across programs (Nursing 2.26, Pharmacy 2.02, Public Health 2.05; $F[2, 190]=0.21$, $p=0.81$) (Additional File 3, Supplementary Table S1). The distribution was right-skewed (skewness=1.20; Supplementary Figure S1): 38.9% selected IDK on no items and 19.2% on five or more, with a single respondent reaching the maximum of 10. Per-item IDK rates ranged from 0.5% (A7, AI errors) to 24.4% (A10, organizational investment in AI, Additional File 3, Supplementary Table S2), with the highest rates on items requiring respondents to project AI's effects into specialized domains they do not directly control.

The Uncertainty Index was independently predicted by self-rated AI conceptual understanding ($\beta= -0.92$, $p=0.004$, MICE-pooled) but was not associated with mean AICPER score ($\beta= -0.012$, $p=0.20$), indicating respondent uncertainty was statistically independent of attitude valence, supporting treatment of IDK responses as missing for imputation. Full models (complete-case and MI-pooled) are in Supplementary Tables S3 and S4.

Content Validity Assessment

CVA demonstrated acceptable agreement across the seven reviewers (Table 2). Mean I-CVI and S-CVI were both 0.86, with item I-CVI ranging from 0.71 to 1.00 and modified kappa from 0.66 to 1.00; six items

achieved perfect agreement (A3, A6, A7, A8, A16, A18). All reached at least 'good' agreement (κ ≥0.60) and were retained.

Item Analysis

Across the 20 AICPER items (Table 3), five (A6, A7, A8, A9, A12) demonstrated floor effects exceeding 20%, all reverse-keyed items (after reverse-scoring, low scale values reflect strong concern). Five items (A1, A2, A4, A5, A18) demonstrated ceiling effects above 20%. Item means ranged from 1.39 to 3.55, with SDs 0.55 to 0.94. After removing A16 (very high ceiling, 57.0%) and A20 (poor item-total correlation), the final 18-item scale demonstrated good internal consistency: pooled Cronbach's α=0.863; ordinal α=0.878; McDonald's ω=0.877.

Exploratory Factor Analysis: A priori vs. Emergent Structure

The instrument was generated according to an *a priori* three-domain structure separating perceptions of AI in clinical practice, research, and education. Horn's parallel analysis retained three factors (polychoric eigenvalues exceeded the 95th percentile of random eigenvalues for factors 1–3: 6.06, 2.03, and 1.48 vs. 1.68, 1.54, and 1.44, respectively); the scree plot showed a clear bend after the third factor (Additional File 3, Supplementary Figure S2, Table S5); cumulative variance reached 53.2% at three factors (above the 50% benchmark); and the three-factor solution was conceptually interpretable. EFA recovered three factors: Functional Benefits (FB), Ethics and Trust (ET), and Educational Engagement (EE), rather than the a priori clinical-practice/research/education structure. Research items distributed across FB (A10) and ET (A11–A15), and two items (A16, A20) were dropped during item analysis.

Under maximum likelihood extraction with oblimin rotation, FB comprised six items (A1-A5, A10) with loadings 0.58-0.82 (pooled α=0.840; ordinal α=0.856; ω=0.857). ET comprised nine items (A6, A7, A8, A9, A11, A12, A13, A14, A15) with loadings 0.29-0.69 (pooled α=0.759; ordinal α=0.782; ω=0.782). EE comprised three items (A17-A19) with loadings 0.60-0.86 (pooled α=0.764; ordinal α=0.788; ω=0.789). Inter-factor correlations were moderate (0.42-0.61). All items except A14 (maximum loading 0.228) met the salient-loading threshold of |0.32|. A14 was retained because its first-order communality and CFA pooled loading (0.40) were acceptable and its content (AI in peer-review) is relevant to the ET. The full EFA pattern matrix is presented in Table 4.

A sensitivity EFA using ULS extraction on the polychoric matrix (oblimin) recovered the same three-factor solution (Additional File 3, Supplementary Table S6). All six FB items (A1–A5, A10) and all three EE items (A17–A19) loaded on the corresponding factors under both approaches, and the ET core (A7, A8, A9, A12) was reproduced across methods.

Confirmatory Factor Analysis

CFA fit indices for the baseline three-factor model were excellent and stable across all imputations (Figure 1). Mean fit indices were CFI=0.987, TLI=0.984, RMSEA=0.032 (range across imputations 0.018-0.038), $\chi^2(132) = 158.3$, AIC=74.7, and BIC=202.0. MI and SEPC for the four candidate correlated error pairs are reported in Additional File 3 (Supplementary Table S7). One pair (A7↔A8) exceeded both thresholds (MI=13.8; SEPC=0.33) and three pairs did not. Pooled standardized loadings with Rubin's-rules SE are reported in Table 5.

Method-Effects Analysis and Measurement Invariance

Adding an orthogonal method factor on the four retained reverse-coded items yielded ΔCFI= +0.019 and ΔRMSEA= -0.025. The method factor accounted for between 5.9%-39.2% of variance (mean 24.8%) in the four reverse-scored items (Additional File 3, Supplementary Table S8). The substantive three-factor structure was retained under the method-effects specification.

Configural-to-metric invariance models showed ΔCFI= -0.001 and ΔRMSEA= -0.001 (Additional File 3, Supplementary Table S9), within the thresholds of |0.010| and |0.015|. This three-group test is interpreted as exploratory because the Public Health subgroup (n=39) is below the conventional 50/group minimum, leaving the ΔCFI test underpowered to detect non-invariance in that subgroup. Per-program Cronbach's α for the full 18-item AICPER was stable across Nursing (α=0.86) and Pharmacy (α=0.86), and higher in Public Health (α=0.92, Additional File 3, Supplementary Table S10).

Bootstrap Robustness and Sensitivity Analyses

All 18 CFA standardized loadings had bootstrap percentile 95% CIs excluding zero (Additional File 3: Supplementary Table S11), indicating stable recovery across 2000 resamples. Under listwise deletion (n=74), internal consistency estimates differed by <0.03 from imputed estimates (FB: 0.86 vs. 0.88; ET: 0.78 vs. 0.80; EE: 0.79 vs. 0.80; AICPER 18-item: 0.87 vs 0.89), and CFA fit indices were equivalent (Additional File 3, Supplementary Table S12). Substantive conclusions did not depend on imputation choice.

DISCUSSION

We developed and validated AICPER to assess perceptions of AI among interdisciplinary health professions trainees. The final 18-item instrument demonstrated acceptable content validity, good internal consistency, and a stable three-factor structure (FB, ET, EE). The structure was supported by excellent CFA fit, robust under a method-effects specification, and demonstrated measurement invariance across the three programs. These findings establish AICPER as the first validated AI-perception instrument for health professions trainees' enabling needs assessment, curriculum development, intervention evaluation, and cross-program benchmarking.

Distinctive Features of AICPER

AICPER is distinguished from existing instruments in three ways. First, treating FB, ET, and EE as separable sub-constructs enables differentiated interpretation that single-factor or two-factor scales preclude [9,10,17]. This granularity supports targeted curricular intervention that contemporary competency and education frameworks call for but unitary tools cannot inform [51-53]. Second, generative-AI-era items (academic integrity, AI-assisted writing, peer review, trust in AI-generated content) addresses content largely absent from pre-2023 instruments, aligning with the contemporary trainee experience. Third, exploratory three-group invariance testing detected no loading-level non-invariance across nursing, pharmacy, and public health, providing preliminary support for cross-disciplinary uses (e.g., interprofessional team training, multi-program evaluation, benchmarking) that single-population profession-specific scales cannot support.

A priori vs. Emergent Factor Structure

The scale was developed with items grouped a priori into clinical practice, research, and education domains, but EFA recovered a different organization (FB, ET, EE) that crossed the original domain boundaries. This is common in instrument validation: respondents' underlying psychological organization of items often reflects valence (utility vs. concern) and behavioral intent more strongly than topical context. The emergent structure mirrors broader technology acceptance literature, where Functional Value and Ethical Concern recur [54,55].

The three empirical domains have intuitive content. FB captures perceptions of efficiency, workflow optimization, decision support, and utility, aligning with literature demonstrating optimism about AI's productivity [56]. ET captures concerns about errors, bias, ethics, integrity, and trust, reflecting recognition that AI adoption is not viewed solely through a technological lens [57,58]. EE captures willingness to pursue AI-related learning, consistent with the view that AI literacy may become an essential competency in healthcare education [59]. Scoring each domain independently lets program directors target intervention where needed, rather than relying on a composite score that obscures differentiated needs.

Importantly, several items relating to FB, had ceiling effects exceeding 20%, while several reverse-coded items had floor effects exceeding 20%. These patterns likely reflect strong consensus among respondents: endorsement of AI's utility on the one hand, and shared ethics and trust concerns on the other. The coexistence of enthusiasm and concern is consistent with prior reports [57,58] and underscores the value of multidimensional assessment that AICPER provides.

Subscale Heterogeneity

The ET subscale showed acceptable but lower reliability than the FB and EE. Mean inter-item r=0.29 is within the range for homogeneous subscales tapping a moderately broad construct, consistent with its broader content (AI errors, bias, accountability, academic-integrity concerns, and trust in AI

recommendations). Future iterations may evaluate whether ET is best modeled as one subscale or two more homogeneous sub-domains (clinical risk and research-integrity concerns).

Practical Implications and Intended Users

AICPER addresses an immediate and growing need in health-professions-education for an evidence-based measure of trainee AI perceptions (Figure 2). For curriculum committees and program directors, AICPER supports pre-program needs assessment to identify which subscale shows the greatest deficit in incoming cohorts and to design targeted modules. The instrument's brevity (18 items) and three-subscale structure make it suitable for pre/post evaluation of AI training, consistent with outcome-assessment recommendations [60-64]. The same applies to educators, supporting faculty development needs assessment under emerging competency frameworks [60]. Given the rapid evolution of AI capabilities, AICPER also supports repeated administration across curriculum cycles to track shifting learner needs, knowledge gaps, and concerns as the underlying technology and its applications continue to develop.

The instrument also maps to the workforce-readiness domain in international AI-readiness frameworks [65,66], supporting national-level reporting. Accrediting bodies in medicine, nursing, pharmacy, and public health (e.g., Liaison Committee on Medical Education, Commission on Collegiate Nursing Education, Accreditation Council for Pharmacy Education, Council on Education for Public Health) increasingly request evidence of AI/informatics coverage for which AICPER offers an evidence-based learner-outcome measure for program self-study. Finally, the separable ET subscale can quantify ethics-driven non-adoption, addressing the documented attitude-adoption gap in which favorable AI attitudes coexist with low tool adoption [67].

Limitations

First, the sample was recruited from a single U.S. institution and did not include medical students, limiting generalizability across settings and disciplines. Second, the cross-sectional design precluded assessment of test-retest reliability and temporal stability. Third, the content expert panel was unbalanced relative to the respondent sample, with no public health representation and only one pharmacy expert; future rounds should recruit a panel matched to respondent disciplines. Fourth, although the substantive three-factor structure persisted under a method-effects specification, the method factor accounted for a non-trivial share of variance in reverse-coded items (mean 24.8%); future iterations should rephrase these items to reduce method variance. Fifth, EFA and CFA used the same sample. We addressed this through bootstrap CIs on factor loadings, exploratory invariance testing, and method-effects analysis, but independent sample replication remains the most important next step. Sixth, the three-group invariance test was exploratory: the Public Health subgroup (n=39) was below the conventional minimum, its higher Cronbach's α (0.92) likely reflects sampling variability at that smaller size rather than greater reliability in that population, and scalar invariance was not tested.

CONCLUSION

AICPER is the first psychometrically validated, multidimensional instrument for measuring U.S. health professions trainees' perceptions of AI across functional, ethical, and engagement domains. Three key findings carry implications beyond this study. First, the factor structure organized around attitudinal dimensions rather than the application domains (clinical practice, research, education), suggests that how trainees feel about AI is shaped more by underlying beliefs than by specific use cases. This should reframe AI readiness, not as domain-specific knowledge gaps, but as attitudinal profiles requiring differentiated pedagogical strategies. Second, preliminary evidence of loading invariance across nursing, pharmacy, and public health student populations is consistent with the use of AICPER as a common framework for cross-disciplinary comparison and benchmarking as accrediting bodies across health professions independently mandate AI competencies but lack standardized tools to assess baseline readiness. Third, the subscale architecture enables actionable diagnostics: a cohort that endorses AI's clinical utility but distrusts its governance (high FB, low ET) requires a different curricular response than one that is ethically comfortable but educationally disengaged (high ET, low EE). AICPER provides the measurement infrastructure to ensure that reform is evidence-informed rather than assumption-driven. Future directions include replication in independent samples, including medical students and convergent validity against MAIRS-MS and GAAIS.

Abbreviations

AI: artificial intelligence

AICPER: Artificial Intelligence in Clinical Practice, Education, and Research (Scale)

BIC: Bayesian Information Criterion

CFA: confirmatory factor analysis

CFI: Comparative Fit Index

COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments

CVI: Content Validity Index

DWLS: diagonally weighted least squares

EE: Educational Engagement (subscale)

EFA: exploratory factor analysis

ET: Ethics and Trust (subscale)

FB: Functional Benefits (subscale)

GAAIS: General Attitudes towards AI Scale

I-CVI: Item Content Validity Index

MAIRS-MS: Medical AI Readiness Scale for Medical Students

MCAR: missing completely at random

MICE: Multiple Imputation by Chained Equations

RMSEA: Root Mean Square Error of Approximation

S-CVI: Scale Content Validity Index

SD: standard deviation

TLI: Tucker-Lewis Index

WLSMV: weighted least squares mean and variance adjusted

Declarations

Ethics Approval and Consent to Participate: The study was reviewed and approved by the Rutgers Robert Wood Johnson Medical School Institutional Review Board (IRB #Pro2025001398; approved July 25, 2025). All participants provided electronic informed consent prior to study participation. All procedures were conducted in accordance with the Declaration of Helsinki and applicable U.S. federal regulations governing human-subjects research.

Consent For Publication: Not applicable. No individual personal data or identifying information is reported.

Availability of Data and Materials: Deidentified data supporting the findings of this study are openly available in FigShare at https://doi.org/10.6084/m9.figshare.32308497. The reporting guideline checklists and supplementary tables are provided as Additional Files 1-3.

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

Funding: This study received no external funding.

Authors' contributions: DK and MN conceived and designed the study and led the instrument development. OA, SS, and JB contributed to item generation, data collection, and preliminary analyses. MN provided supervision and contributed to instrument design and content validation. All authors contributed to the psychometric analysis plan. OA and DK drafted the manuscript; All authors provided critical revision. All authors read and approved the final manuscript.

Acknowledgements: We acknowledge the seven content experts who participated in the validation phase of instrument development, and the 193 nursing, pharmacy, and public health students who

completed the survey.

References

1. Angus DC, Khera R, Lieu T, Liu V, Ahmad FS, Anderson B, et al. AI, health, and health care today and tomorrow: the JAMA Summit report on artificial intelligence. JAMA. 2025;334(18):1650-1664. doi:10.1001/jama.2025.18490.
2. Kewalramani D, Jawa R, Cuschieri J, Kaafarani H, Buchman TG, Narayan M. Navigating the artificial intelligence landscape in trauma, critical care and emergency general surgery: insights from the American Association for the Surgery of Trauma (AAST) 2025 Annual Meeting Panel Discussion. Trauma Surg Acute Care Open. 2026;11(1):e002205. doi:10.1136/tsaco-2025-002205.
3. Kewalramani D, Chattopadhyay K, Benton J, Hua J, Cheruvu S, Ali H, et al. A scoping review of artificial intelligence in acute care surgery: promise, pitfalls, and a path forward. Int J Surg. 2026;112(2):2214-2224. doi:10.1097/JS9.0000000000003794.
4. Castillo-Medina A, Calleja-Zardain R, Kewalramani D, Narayan M, Mayol J. Inteligencia artificial como herramienta de la cirugía global en América Latina. Rev Colomb Cir. 2025;40(1):25-32. doi:10.30944/20117582.2622.
5. Kewalramani D, Roman DS, Lagos SA, Rammsy F, Villagran I, Escalona G, et al. Quality matters: artificial intelligence-based assessment of feedback quality predicts technical skill improvement. Surgery. 2025;187:109625. doi:10.1016/j.surg.2025.109625.
6. Alsakaker AA, Alfayez JS, Alsalamah JA, Alzughibi LS, Anaam MS, Dixon DL, et al. ChatGPT: pharmacy students' perceptions, current use trends, ethical awareness, standards of ethics, prospects and recommendations for future use. Curr Pharm Teach Learn. 2025;17(10):102382. doi:10.1016/j.cptl.2025.102382.
7. Gonzalez-Garcia A, Bermejo-Martinez D, Lopez-Alonso AI, Trevisson-Redondo B, Martín-Vázquez C, Perez-Gonzalez S. Impact of ChatGPT usage on nursing students education: a cross-sectional study. Heliyon. 2025;11(1):e41559. doi:10.1016/j.heliyon.2024.e41559.
8. Zhang JS, Yoon C, Williams DKA, Pinkas A. Exploring the usage of ChatGPT among medical students in the United States. J Med Educ Curric Dev. 2024;11:23821205241264695. doi:10.1177/23821205241264695.
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):112. doi:10.1186/s12909-021-02546-6.
10. 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.
11. Chen X, Chen Y, Xie Y, Cheng L. Translation and psychometric validation of the Medical Artificial Intelligence Readiness Scale (MAIRS-MS) for Chinese medical students. BMC Nurs. 2025;24(1):1210. doi:10.1186/s12912-025-03852-w.

12. Hamad M, Qtaishat F, Mhairat E, Al-Qunbar A, Jaradat M, Mousa A, et al. Artificial intelligence readiness among Jordanian medical students: using the Medical Artificial Intelligence Readiness Scale for Medical Students (MAIRS-MS). J Med Educ Curric Dev. 2024;11:23821205241281648. doi:10.1177/23821205241281648.
13. Moodi Ghalibaf A, Moghadasin M, Emadzadeh A, Mastour H. Psychometric properties of the Persian version of the Medical Artificial Intelligence Readiness Scale for Medical Students (MAIRS-MS). BMC Med Educ. 2023;23(1):577. doi:10.1186/s12909-023-04553-1.
14. Yılmaz D, Uzelli D, Dikmen Y. Psychometrics of the attitude scale towards the use of artificial intelligence technologies in nursing. BMC Nurs. 2025;24(1):151. doi:10.1186/s12912-025-02732-7.
15. Turan M, Cengiz Z. Nursing students' general attitudes towards artificial intelligence scale (NGAAIS): a Turkish validity and reliability study. Nurse Educ Pract. 2025;88:104574. doi:10.1016/j.nepr.2025.104574.
16. El Arab RA, Alshakihs AH, Alabdulwahab SH, Almubarak YS, Alkhalifah SS, Abdrbo A, et al. Artificial intelligence in nursing: a systematic review of attitudes, literacy, readiness, and adoption intentions among nursing students and practicing nurses. Front Digit Health. 2025;7:1666005. doi:10.3389/fdgth.2025.1666005.
17. Schepman A, Rodway P. Initial validation of the general attitudes towards artificial intelligence scale. Comput Hum Behav Rep. 2020;1:100014. doi:10.1016/j.chbr.2020.100014.
18. Wang YY, Wang YS. Development and validation of an artificial intelligence anxiety scale: an initial application in predicting motivated learning behavior. Interact Learn Environ. 2022;30(4):619-634. doi:10.1080/10494820.2019.1674887.
19. Wang B, Rau PLP, Yuan T. Measuring user competence in using artificial intelligence: validation of an artificial intelligence literacy scale. Behav Inf Technol. 2023;42(9):1324-1337. doi:10.1080/0144929X.2022.2072768.
20. Gustafson KA, Berman S, Gavaza P, Mohamed I, Devraj R, Abdel Aziz MH, et al. Pharmacy faculty and students perceptions of artificial intelligence: a national survey. Curr Pharm Teach Learn. 2025;17(6):102344. doi:10.1016/j.cptl.2025.102344.
21. Shin H, De Gagne JC, Kim SS, Hong M. The impact of artificial intelligence-assisted learning on nursing students' ethical decision-making and clinical reasoning in pediatric care: a quasi-experimental study. Comput Inform Nurs. 2024;42(10):704-711. doi:10.1097/CIN.0000000000001177.
22. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res. 2021;30(8):2197-2218. doi:10.1007/s11136-021-02822-4.
23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010.

24. Little RJA. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc. 1988;83(404):1198-1202. doi:10.1080/01621459.1988.10478722.
25. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1-67. doi:10.18637/jss.v045.i03.
26. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011;20(1):40-49. doi:10.1002/mpr.329.
27. Bonett DG. Sample size requirements for testing and estimating coefficient alpha. J Educ Behav Stat. 2002;27(4):335-340. doi:10.3102/10769986027004335.
28. Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons; 1987.
29. Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y. semTools: useful tools for structural equation modeling. R package version 0.5-6; 2022. Available from: https://CRAN.R-project.org/package=semTools.
30. Enders CK. Applied missing data analysis. New York: Guilford Press; 2010.
31. 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.
32. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health. 2006;29(5):489-497. doi:10.1002/nur.20147.
33. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42. doi:10.1016/j.jclinepi.2006.03.012.
34. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.
35. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. 5th ed. Oxford: Oxford University Press; 2014.
36. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16(3):297-334. doi:10.1007/BF02310555.
37. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide. Pract Assess Res Eval. 2012;17(3):1-13. doi:10.7275/n560-j767.
38. McDonald RP. Test theory: a unified treatment. Mahwah (NJ): Lawrence Erlbaum Associates; 1999.
39. Tabachnick BG, Fidell LS. Using multivariate statistics. 7th ed. Boston: Pearson; 2019.
40. Adeyemi OJ. Mobile phone use while driving: development and validation of knowledge, attitude, and practice survey instruments. J Safety Res. 2021;77:30-39. doi:10.1016/j.jsr.2021.01.004.
41. Adeyemi OJ, Bouillon-Minois JB, Siman N, Cuthel AM, Goldfeld KS, Grudzen CR. Knowledge and attitudes toward hospice and palliative care: instrument validation among emergency providers. Am J Hosp Palliat Care. 2023;40(3):280-290. doi:10.1177/10499091221098664.
42. Yusufov M, Adeyemi O, Flannery M, Bouillon-Minois JB, Van Allen K, Cuthel AM, et al. Psychometric properties of the functional assessment of cancer therapy-general for evaluating quality of life in

patients with life-limiting illness in the emergency department. J Palliat Med. 2024;27(1):63-74. doi:10.1089/jpm.2022.0270.

43. Beauducel A, Herzberg PY. On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. Struct Equ Modeling. 2006;13(2):186-203. doi:10.1207/s15328007sem1302_2.
44. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6(1):1-55. doi:10.1080/10705519909540118.
45. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge: Cambridge University Press; 1997.
46. DiStefano C, Motl RW. Further investigating method effects associated with negatively worded items on self-report surveys. Struct Equ Modeling. 2006;13(3):440-464. doi:10.1207/s15328007sem1303_6.
47. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Modeling. 2002;9(2):233-255. doi:10.1207/S15328007SEM0902_5.
48. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Modeling. 2007;14(3):464-504. doi:10.1080/10705510701301834.
49. Worthington RL, Whittaker TA. Scale development research: a content analysis and recommendations for best practices. Couns Psychol. 2006;34(6):806-838. doi:10.1177/0011000006288127.
50. Kline RB. Principles and practice of structural equation modeling. 4th ed. New York: Guilford Press; 2016.
51. Kewalramani D, Jawa RS, Martin CA, Gosain A, Wan D, Varghese TK, et al. Position statement from the Society of University Surgeons, Surgical Education Committee: artificial intelligence in surgical training for medical students, residents, and fellows. Surgery. 2026;190:109849. doi:10.1016/j.surg.2025.109849.
52. Josiah Macy Jr. Foundation. Conference on artificial intelligence in medical education: proceedings and recommendations. Acad Med. 2025;100(Suppl_1):S4-S14.
53. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025;388:e081554. doi:10.1136/bmj-2024-081554.
54. 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.
55. Yu T, Tian Y, Huang Q, Cheng Z, Zhang R. Explaining adoption of AI tools in education: a dual-path model of ethical concern and functional value. Front Psychol. 2026;16:1735913. doi:10.3389/fpsyg.2025.1735913.
56. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. Implement Sci. 2024;19(1):27. doi:10.1186/s13012-024-01357-9.

57. Tung T, Hasnaeen SMN, Zhao X. Ethical and practical challenges of generative AI in healthcare and proposed solutions: a survey. Front Digit Health. 2025;7:1692517. doi:10.3389/fdgth.2025.1692517.
58. Qadri S, Kewalramani D, Aboutanos MB, Narayan M, Quintana MT. Artificial intelligence in surgical care: a point-counterpoint analysis. Trauma Surg Acute Care Open. 2026;11(2):e002093. doi:10.1136/tsaco-2025-002093.
59. Ang CS. Developing AI literacy in healthcare education: bridging the gap in competency assessment. Discov Educ. 2025;4(1):372. doi:10.1007/s44217-025-00754-6.
60. Association of American Medical Colleges. Artificial intelligence competencies across the learning continuum and AI competencies for medical educators. Washington (DC): AAMC; 2025. Available from: https://www.aamc.org/about-us/medical-education/ai-competencies.
61. American Medical Association. AMA policy on advancing AI literacy in medical education. Chicago (IL): American Medical Association; 2025.
62. American Association of Colleges of Nursing. The Essentials: core competencies for professional nursing education. Washington (DC): AACN; 2021.
63. National League for Nursing. Artificial intelligence (AI) in nursing education: NLN vision statement. Washington (DC): NLN; 2023.
64. Association of Schools and Programs of Public Health. Task force for the responsible and ethical use of AI in public health research, practice, and education. Washington (DC): ASPPH; 2024.
65. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021.
66. World Health Organization Regional Office for Europe. Artificial intelligence is reshaping health systems: state of readiness across the WHO European Region. Copenhagen: WHO Regional Office for Europe; 2025.
67. Pruski M, Woolley KE, Withers KL. Barriers and facilitators to health care AI adoption among those living in Wales and working in health care in Wales: online survey. J Med Internet Res. 2025;27:e81543. doi:10.2196/81543.

Tables

Table 1. Sociodemographic characteristics of survey respondents (N = 193).

Characteristic	Category	n	%
Age (years)	20-30	152	78.8
	31-40	14	7.3
	41-50	4	2.1
	50 or older	4	2.1
	Prefer not to say	19	9.8
Sex / Gender	Woman	146	75.6
	Man	44	22.8
	Prefer not to say	1	0.5
	Prefer to self-describe	2	1.0
Program	Nursing	58	30.1
	Pharmacy	96	49.7
	Public Health	39	20.2
Degree level	Bachelor's (BSN, Accel BSN)	42	21.8
	Master's (MPH, MS)	24	12.4
	Doctoral (DNP, PhD, PharmD, DrPH)	121	62.7
	Certificate / Other	6	3.1
Year of training	Year 1	63	32.6
	Year 2	50	25.9
	Year 3 or higher	80	41.5
Prior AI experience	Formal education	15	7.8
	Practical exposure	124	64.2
	Basic general awareness	23	11.9
	None / Not reported	31	16.1
Prior programming	None / Not reported	111	57.5
	Basic	61	31.6
	Moderate	17	8.8
	Advanced	4	2.1

BSN = Bachelor of Science in Nursing; DNP = Doctor of Nursing Practice; PharmD = Doctor of Pharmacy; MPH = Master of Public Health; DrPH = Doctor of Public Health.

Table 2. Content validity assessment of the 20-item AICPER Scale (N = 7 experts).

Item	Item description	n relevant (of 7)	I-CVI	κ (PBO 2007)	κ band
A1	AI will play a significant role in [profession]	6	0.857	0.849	excellent
A2	AI technologies have potential to improve outcomes	6	0.857	0.849	excellent
A3	AI-driven analyses help professionals make better decisions	7	1	1	excellent
A4	AI can reduce workload	6	0.857	0.849	excellent
A5	AI can improve efficiency	5	0.714	0.658	good
A6	I trust AI tools in providing support / rationale	7	1	1	excellent
A7	I worry AI may make errors [REVERSE]	7	1	1	excellent
A8	I worry AI may make biased decisions [REVERSE]	7	1	1	excellent
A9	I have ethical/legal concerns about AI [REVERSE]	6	0.857	0.849	excellent
A10	Agencies/organizations should invest in AI	6	0.857	0.849	excellent
A11	AI tools can make research more efficient	6	0.857	0.849	excellent
A12	AI-generated content compromises academic integrity [REVERSE]	5	0.714	0.658	good
A13	Using AI for writing/editing academic papers is acceptable	6	0.857	0.849	excellent
A14	AI could be helpful in peer-review process	6	0.857	0.849	excellent
A15	I would have confidence in research with AI assistance	5	0.714	0.658	good
A16	Researchers should disclose when AI tools used (REMOVED)	7	1	1	excellent
A17	Likelihood to take AI Elective if offered	6	0.857	0.849	excellent
A18	Likelihood to pursue AI Certificate/badge	7	1	1	excellent
A19	Likelihood to seek additional AI training	5	0.714	0.658	good
A20	I am concerned AI might replace jobs (REMOVED) [REVERSE]	5	0.714	0.658	good

I-CVI = Item Content Validity Index; κ = modified kappa per Polit, Beck and Owen: κ = (I-CVI − Pc)/(1 − Pc), Pc = C(N,A) × 0.5⁻⁴.*

Table 3. Item analysis statistics for the 20-item AICPER Scale (N = 193, MICE-imputed).

Item	Mean	SD	% Floor	% Ceiling
A1	2.96	0.73	2.1	22.8
A2	3.1	0.68	2.1	26.4
A3	2.93	0.71	3.1	18.7
A4	3.17	0.62	0.5	28.5
A5	3.12	0.62	1	24.9
A6	1.91	0.64	24.4	1
A7	1.39	0.58	64.8	1
A8	1.78	0.87	47.2	4.1
A9	1.59	0.66	48.7	1.6
A10	2.79	0.66	3.1	10.4
A11	2.97	0.6	3.1	13.5
A12	1.62	0.59	43	0.5
A13	2.45	0.78	13	5.2
A14	2.92	0.63	4.1	11.9
A15	2.17	0.66	12.4	2.6
A16	3.55	0.55	0.5	57
A17	2.58	0.89	14	13.5
A18	2.78	0.94	12.4	23.3
A19	2.38	0.89	17.1	10.4
A20	2.09	0.87	27.5	6.2

All items reverse-coded so that higher values reflect more favorable perceptions. Items A16 and A20 removed during item analysis. Floor = % of respondents endorsing the lowest scale value (1); Ceiling = % of respondents endorsing the highest scale value (4). Floor effects ≥ 20%: A6, A7, A8, A9, A12. Ceiling effects ≥ 20%: A1, A2, A4, A5, A18.

Table 4. EFA pattern matrix (maximum likelihood extraction, oblimin rotation).

Item	Functional Benefits	Ethics & Trust	Educational Engagement
A1	0.730		
A2	0.817		
A3	0.732		
A4	0.579		
A5	0.757		
A10	0.586		
A6	0.402	0.343	
A7		0.482	
A8		0.472	
A9		0.750	
A11	0.537		
A12		0.582	
A13		0.389	0.333
A14	—		
A15	0.424	0.291	
A17			0.812
A18			0.737
A19			0.631

Loadings |≤ 0.25| suppressed for ease of interpretation. A14's maximum loading was 0.228, below the |0.25| suppression cutoff. Items below the Tabachnick & Fidell [50] |0.32| salient-loading threshold are flagged in the Results text.

Table 5. Pooled CFA standardized loadings (Rubin's-rules pooling, M = 20 imputations, DWLS estimator).

Item	Factor	Pooled std loading	SE	95% CI lower	95% CI upper
A1	FB	0.669	0.104	0.466	0.872
A2	FB	0.75	0.121	0.514	0.987
A3	FB	0.751	0.128	0.5	1.002
A4	FB	0.515	0.104	0.312	0.719
A5	FB	0.628	0.109	0.415	0.842
A10	FB	0.767	0.124	0.524	1.011
A6	ET	0.658	0.101	0.46	0.857
A7	ET	0.425	0.112	0.206	0.645
A8	ET	0.351	0.172	0.014	0.688
A9	ET	0.354	0.13	0.099	0.608
A11	ET	0.612	0.137	0.344	0.881
A12	ET	0.532	0.122	0.293	0.771
A13	ET	0.578	0.164	0.256	0.9
A14	ET	0.392	0.13	0.137	0.647
A15	ET	0.602	0.147	0.314	0.89
A17	EE	0.68	0.102	0.481	0.88
A18	EE	0.635	0.135	0.371	0.898
A19	EE	0.83	0.15	0.537	1.123

Pooled standard errors computed per Rubin's-rules. Total variance = within-imputation + (1 + 1/M) × between-imputation variance

Figures

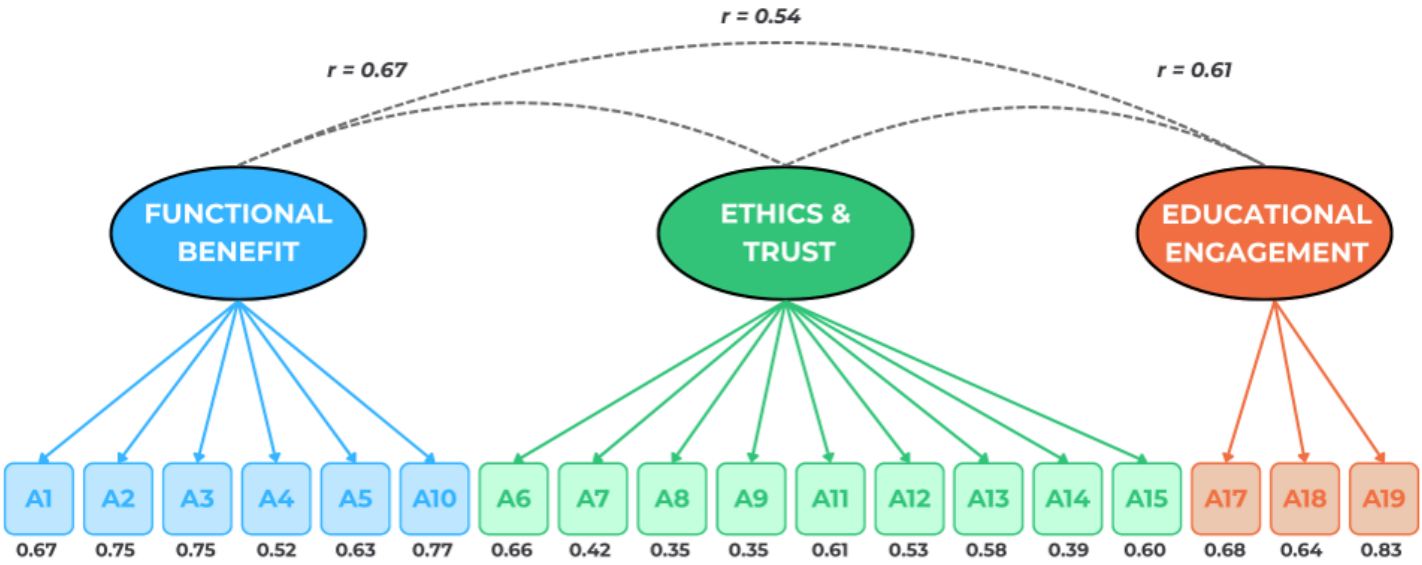


Figure 1

Baseline three-factor CFA model of the AICPER Scale with pool standardized factor loadings.

FB = Functional Benefits; ET = Ethics and Trust; EE = Educational Engagement. Each item loads on its EFA-assigned factor; no correlated error terms are specified in this baseline model. Pooled standardized factor loadings (Rubin's-rules pooling, M = 20 imputations) are shown next to each path. Inter-factor correlations are shown as curved dashed double-headed arrows with r values.

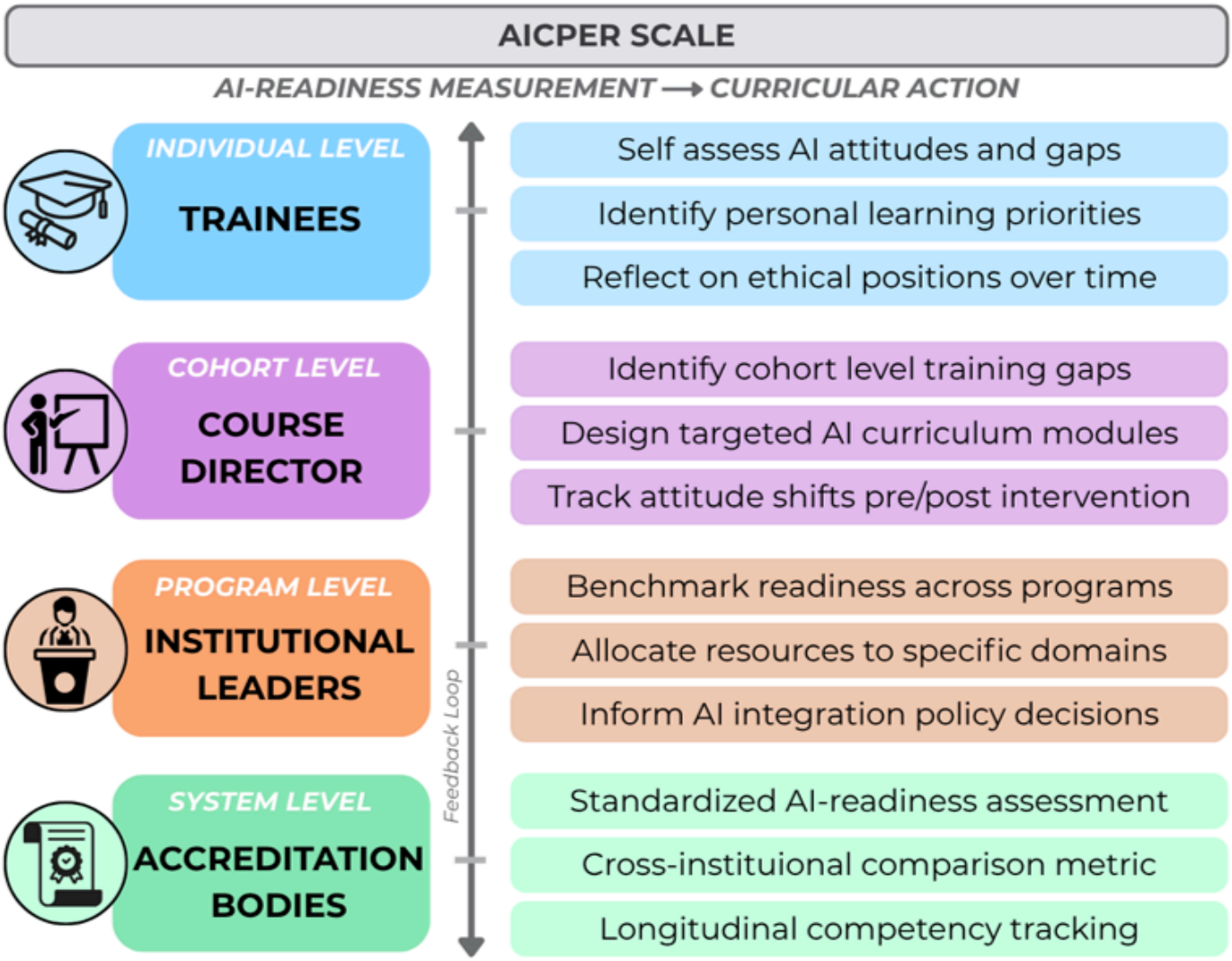


Figure 2

Practical applications of the AICPER Scale across stakeholder levels.

The AICPER Scale supports AI-readiness measurement and curricular action at four levels: individual trainees (self-assessment and reflection), course directors (cohort-level needs assessment and curriculum design), institutional leaders (cross-program benchmarking and resource allocation), and accreditation bodies (standardized assessment and longitudinal tracking). The vertical feedback loop indicates that system-level benchmarking data informs individual-level interventions, and individual-level response patterns aggregate upward to shape institutional and accreditation decisions. AICPER = Artificial Intelligence in Clinical Practice, Education, and Research.

Supplementary Files

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

- AF2AICPERScale.docx
- AF1COSMINChecklist.docx
- AF3SupplementaryTablesandFigures.docx