Assessing the Psychometric Properties of an AI and Big Data Analytics Adoption Readiness Instrument in Primary Health Care

Abstract: *Primary health care (PHC) centres in Saudi Arabia are expected to adopt artificial intelligence (AI) and big data analytics (BDA) for chronic disease management, yet no validated instrument measures readiness for these technologies among practising PHC professionals. This study assessed the content validity, face validity, feasibility and internal consistency of a 67-item AI and BDA Adoption Readiness Instrument developed for PHC centres in Al Madinah. Items were adapted from published scales and organised into eight subscales grounded in the Technology Acceptance Model (perceived usefulness, perceived ease of use) and the Technology-Organisation-Environment framework (organisational readiness, organisational culture and climate, technological readiness, environmental readiness), together with AI-related barriers and overall adoption readiness. Six experts rated each item for relevance and clarity, and a pilot study was conducted with 30 professionals (8 physicians, 8 nurses, 7 public health specialists and 7 chronic disease coordinators) from five PHC centres. All 67 items achieved an item-level content validity index (I-CVI) of at least .83 (modified kappa ≥ .82) on both relevance and clarity. The scale-level index (S-CVI/Ave) was .93 for relevance and .92 for clarity; clarity fell slightly below .90 for the barriers (.89) and adoption readiness (.86) subscales. Mean completion time was 15.0 minutes, and participants rated overall clarity at 4.20 and relevance at 4.40 on a 5-point scale. Cronbach's alpha ranged from .740 (environmental readiness) to .907 (organisational readiness), although the lower 95% confidence limits for organisational culture (.689) and environmental readiness (.564) fell below .70. Three items had corrected item-total correlations below .30, and four items (PEOU10, BR3, BR12, ER2) were revised and re-rated by the expert panel with an I-CVI of 1.00. Perceived usefulness and ease of use correlated with each other (r = .69) and with adoption readiness (r = .51 and .55); organisational readiness and culture overlapped strongly (r = .74), and the barriers subscale was unrelated to the other scales. A one-factor solution for the adoption readiness subscale explained 49.9% of variance. The instrument shows strong content validity, acceptable feasibility and adequate preliminary reliability. Its factor structure, discriminant validity between organisational subscales and test-retest stability should be established in the main survey before it is used to guide AI and BDA adoption in Saudi health clusters.*

Keywords: artificial intelligence; big data analytics; primary health care; adoption readiness; psychometric properties; content validity; reliability; technology acceptance model; technology-organisation-environment framework; Saudi Arabia.

1. Introduction

Non-communicable diseases (NCDs) are the dominant cause of death and disability in Saudi Arabia, accounting for an estimated 73% of all deaths (World Health Organization [WHO], 2018) and 67.4% of disability-adjusted life-years lost in 2019 (Boettiger et al., 2023). Detection and control gaps remain large: in 2013, 57.8% of Saudis with diabetes were undiagnosed and only 16.6% were treated and controlled (El Bcheraoui et al., 2014). In a study of 692 adults with diabetes attending 15 primary care centres in Al Madinah, only 15.7% achieved an HbA1c below 7% (Alfadhli et al., 2025). Primary health care (PHC) centres carry most of the responsibility for screening, risk stratification and long-term follow-up of this population, and under Vision 2030 they have been reorganised into integrated

health clusters supported by national digital platforms (Health Sector Transformation Program, 2023). Artificial intelligence (AI) and big data analytics (BDA) are increasingly proposed as tools for closing these gaps. Autonomous AI for diabetic retinopathy screening has met prespecified accuracy endpoints in primary care (Abràmoff et al., 2018) and increased screening completion in a randomised trial (Wolf et al., 2024), and BDA has shown moderate-to-high accuracy for predicting diabetes complications (Borges do Nascimento et al., 2021). In this study, AI refers to computational systems that perform tasks associated with human judgement, including machine-learning prediction, image interpretation, natural language processing and generative tools. BDA refers to the integration and analysis of large, heterogeneous health datasets to generate population- and patient-level insight. Because professionals rarely distinguish the two in practice, the instrument measures them jointly.

The binding constraint for PHC is less the existence of capable algorithms than the readiness of organisations and professionals to adopt, integrate and govern them. Saudi evidence shows that this readiness is uneven. Only 150 of 2,259 PHC centres had an implemented electronic health record (EHR) system in a recent national study, and lack of training (83.8%) and inadequate infrastructure (78.5%) were the most frequently cited barriers (Alzghaibi & Hutchings, 2025). Among Saudi PHC professionals, 47.2% reported being unfamiliar with AI tools and 31.6% reported using them (Alsudairy et al., 2025). Saudi health professionals also report high levels of worry about data quality and professional consequences of AI (Alsaedi et al., 2024). Most AI-readiness research has been conducted in hospitals rather than PHC (Kueper et al., 2020). Two theoretical traditions inform readiness measurement. The Technology Acceptance Model (TAM) proposes that perceived usefulness (PU) and perceived ease of use (PEOU) shape intention to use a technology (Davis, 1989). TAM has been widely applied and extended in health informatics, but it explains intention better than actual use and needs context-specific additions (Holden & Karsh, 2010; Rahimi et al., 2018; Turner et al., 2010). The Technology–Organisation–Environment (TOE) framework locates adoption in the organisation: characteristics of the technology, the organisation's structure, resources and leadership, and the external regulatory environment jointly shape adoption (Tornatzky & Fleischer, 1990). TOE has been applied to big data adoption readiness among health employees (Ghaleb et al., 2021) and to organisational e-health readiness in PHC (Kruszyńska-Fischbach et al., 2022). Organisational readiness theory adds the shared commitment and efficacy that link individual appraisals to organisational change (Weiner, 2009; Shea et al., 2014).

Existing instruments measure these levels separately. The Medical Artificial Intelligence Readiness Scale for Medical Students (MAIRS-MS) assesses individual readiness but was developed with students (Karaca et al., 2021). The Organizational Readiness for Implementing Change (ORIC) measure was validated outside health care and is not technology-specific (Shea et al., 2014). The e-health readiness model of Kruszyńska-Fischbach et al. (2022) targets facility managers and is not specific to AI. An extended TAM including organisational culture and trust has been tested in Indian hospitals (Rathore et al., 2026). No validated instrument was identified that measures AI and BDA readiness across TAM and TOE dimensions among practising PHC professionals, and none was identified for the Saudi or Al Madinah context. This study therefore developed and assessed a 67-item AI and BDA Adoption Readiness Instrument for PHC. It addresses four questions: (a) does the instrument show adequate content validity according to an expert panel; (b) is it feasible and understandable for frontline PHC professionals (face validity); (c) do its subscales show adequate internal consistency and item performance in a pilot sample; and (d) do preliminary relationships between subscales and the structure of the overall readiness subscale agree with theoretical expectations.

2. Methods

2.1 Study Design and Setting

A two-stage instrument validation design was used. Stage 1 was an expert content validation of the adapted item pool. Stage 2 was a pilot study that assessed feasibility, face validity, internal consistency, item performance and a first structural check. The study was conducted in PHC centres of the Al Madinah Health Cluster, Saudi Arabia, as the preparatory phase of a larger cross-sectional survey of AI and BDA readiness.

2.2 Instrument Development

The instrument has two parts. Section A collects demographic and professional data. Sections B–H contain 67 items rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The items form eight subscales. Two are derived from TAM: perceived usefulness (PU) and perceived ease of use (PEOU). Four are aligned with TOE: organisational readiness (OR), organisational culture and climate (OC), technological readiness (TR) and environmental readiness (ER). Two further subscales measure AI-related challenges and barriers (BR) and overall AI/BDA adoption readiness (AR), which is the intended outcome variable. Items were drawn from published instruments and reworded from generic technology to AI and BDA tools, PHC roles and the Saudi context (Table 1). Wording changes were logged item by item so that departures from the validated source wording could be traced.

Table 1: Subscales, number of items, sources and adaptation

Subscale	Items	Sources	Adaptation
Perceived usefulness (PU)	10	Davis (1989), PU1–5; Alsaedi et al. (2024), PU6–7; Rathore et al. (2026), PU8–10	Reworded from generic technology to AI/BDA tools and PHC roles
Perceived ease of use (PEOU)	10	Davis (1989), PEOU1–5; Rathore et al. (2026), PEOU6–10	Reworded to AI/BDA tools
Organisational readiness (OR)	9	Shea et al. (2014), OR1–4; Tornatzky & Fleischer (1990) and Kruszyńska-Fischbach et al. (2022), OR5–6; Kruszyńska-Fischbach et al. (2022), OR7–9	Reworded to AI/BDA adoption in PHC
Organisational culture and climate (OC)	5	Rathore et al. (2026), OC1–4; Tornatzky & Fleischer (1990) and Kruszyńska-Fischbach et al. (2022), OC5	Reworded to AI adoption; OC5 added from the organisational block
Technological readiness (TR)	8	Kruszyńska-Fischbach et al. (2022); Alam et al. (2024); Tornatzky & Fleischer (1990)	Reworded to AI/BDA; data quantity and quality split into TR5 and TR6
Environmental readiness (ER)	6	Alorfi & Alsaadi (2025); Tornatzky & Fleischer (1990)	Reworded to the Saudi PHC context (e.g., Vision 2030, Ministry of Health)
Challenges and barriers (BR)	12	Alsaedi et al. (2024)	Reworded to the PHC context
Overall adoption readiness (AR)	7	Davis (1989); Tornatzky & Fleischer (1990); Rathore et al. (2026)	Combines PHC-level readiness (AR1, AR4) with personal intention (AR2, AR3, AR5–AR7)

2.3 Expert Content Validation

A panel of six experts was invited, the minimum recommended for content validation (Lynn, 1986). Eligible experts had at least five years of relevant experience in PHC or chronic disease management and represented healthcare administration, digital health, public health and clinical practice

(physicians and chronic disease coordinators). Experts received the 67-item questionnaire with definitions of the eight constructs. They rated each item on a 4-point scale for relevance (1 = not relevant, 4 = highly relevant) and clarity (1 = not clear, 4 = very clear), wrote a comment on each item, and stated whether any important content was missing. The item-level content validity index (I-CVI) was calculated as the proportion of experts rating an item 3 or 4. With six experts, an I-CVI of at least .78 was required for retention, which in practice meant .83 (five experts) or 1.00 (Lynn, 1986; Polit & Beck, 2006). The scale-level index based on the average method (S-CVI/Ave) was calculated as the mean I-CVI of a subscale's items, with .90 as the acceptance criterion. The universal agreement index (S-CVI/UA) was calculated as the proportion of items rated 3 or 4 by every expert. Each I-CVI was adjusted for chance agreement using the modified kappa (Polit et al., 2007):

$$\kappa^{*} = \frac{\text{I-CVI} - p_c}{1 - p_c}, \qquad p_c = \frac{N!}{A!\,(N - A)!} \times 0.5^{N}$$

where N is the number of experts and A the number agreeing that the item is relevant (or clear). Values above .74 were considered excellent (Cicchetti & Sparrow, 1981). Agreement across the panel on the dichotomised agree/disagree ratings was also expressed as the mean pairwise proportion of agreement and as Fleiss' kappa (Fleiss, 1971).

2.4 Pilot Study

A quota sample of 30 professionals was recruited from five PHC centres: 8 physicians, 8 nurses, 7 public health specialists and 7 chronic disease coordinators. A sample of 30 gives a 96% probability of detecting a comprehension problem that affects at least 10% of respondents (Viechtbauer et al., 2015) and is considered adequate for initial scale piloting (Johanson & Brooks, 2010). Participants completed all 67 items, recorded their completion time, rated the overall clarity and relevance of the questionnaire on a 5-point scale, and marked any item they found unclear.

2.5 Statistical Analysis

Data were analysed in SPSS version 24. Responses were screened for out-of-range values, missing data and straight-lining (identical answers across items). Feasibility and face validity were summarised as means, standard deviations and counts; an item marked unclear by at least three participants (10%) was reviewed. Internal consistency was estimated with Cronbach's alpha (Cronbach, 1951), with 95% confidence intervals derived from the F distribution (Feldt et al., 1987); alpha of .70 or above was considered acceptable. Corrected item–total correlations below .30 were flagged, and alpha-if-item-deleted was examined. Subscale scores were computed as item means. Normality was assessed with skewness, kurtosis and the Shapiro–Wilk test, and exploratory relationships between subscales were examined with Pearson correlations. Because a full factor analysis was not feasible with n = 30, the procedure planned for the main survey (Costello & Osborne, 2005) was rehearsed with a principal axis factor analysis of the seven AR items. Reliability statistics were recalculated independently and matched the SPSS output. Agreement between expert and pilot evidence was tested with Fisher's exact test.

2.6 Ethical Considerations

[Ethics approval body, approval number and date to be added.] Participation was voluntary, responses were anonymous, and no patient data were collected.

3. Results

3.1 Expert Panel and Content Validity

All six experts returned complete forms within four days (response rate 100%). The panel comprised one healthcare administrator, two digital health experts, one public health specialist, one physician and one chronic disease coordinator, with 5 to 11 years of PHC experience (mean 7.3, SD 2.1) (Table 2).

Table 2: Expert panel profile and rating patterns

Expert	Category	Years of experience	Ratings of 2 (relevance)	Ratings of 2 (clarity)	Mean relevance	Mean clarity
E1	Healthcare administrator	6	6	4	3.40	3.51
E2	Digital health expert	8	4	5	3.31	3.34
E3	Public health specialist	5	5	4	3.43	3.51
E4	Physician	7	5	5	3.49	3.49
E5	Digital health expert	7	4	6	3.66	3.51
E6	Chronic disease coordinator	11	3	10	3.54	3.36

Every item met the retention criterion on both relevance and clarity; the lowest I-CVI was .83. On relevance, 40 of 67 items (59.7%) had an I-CVI of 1.00 and 27 had .83. On clarity, 33 items (49.3%) had 1.00 and 34 had .83. Twelve items scored .83 on both criteria: PU2, PU4, PEOU1, PEOU3, PEOU10, OR8, TR8, BR3, BR4, BR10, AR2 and AR5. The modified kappa was .82 for every item with an I-CVI of .83 and 1.00 for every item with an I-CVI of 1.00, so all items were rated excellent (> .74). Across the instrument, S-CVI/Ave was .93 for relevance and .92 for clarity, and S-CVI/UA was .60 and .49, respectively (Table 3). Relevance S-CVI/Ave reached .90 in all eight subscales. Clarity S-CVI/Ave reached .90 in six subscales but fell slightly below in BR (.89) and AR (.86). The shortfall reflected diffuse rather than concentrated disagreement: 8 of the 12 BR items and 6 of the 7 AR items each received one clarity rating of 2.

Table 3: Scale-level content validity indices (six experts)

Subscale	Items	S-CVI/Ave relevance	S-CVI/Ave clarity	S-CVI/UA relevance	S-CVI/UA clarity
PU	10	.933	.917	.60	.50
PEOU	10	.917	.917	.50	.50
OR	9	.944	.926	.67	.56
OC	5	.900	.967	.40	.80
TR	8	.958	.938	.75	.62
ER	6	.972	.944	.83	.67
BR	12	.917	.889	.50	.33
AR	7	.929	.857	.57	.14
Total	67	.933	.915	.60	.49

S-CVI/Ave = scale-level content validity index, average method; S-CVI/UA = universal agreement method. Values below the .90 criterion for S-CVI/Ave: BR and AR clarity.

3.2 Agreement Among Experts

Mean pairwise agreement on the dichotomised ratings was 86.6% for relevance and 83.1% for clarity. Fleiss' kappa was nevertheless slightly negative (−.072 and −.092). This reflects the prevalence effect:

93.3% of relevance ratings and 91.5% of clarity ratings fell in the agree category, which inflates expected chance agreement and makes kappa uninformative. Of the 61 ratings of 2 (27 relevance, 34 clarity), all six experts contributed some; the chronic disease coordinator (E6) gave 10 of the 34 clarity ratings of 2, five of them on barrier items (BR4, BR5, BR7, BR10, BR11). The 402 written comments were brief and favourable (21 unique phrases, most often "Good", "Well constructed" and "Well written"); none explained a rating of 2, and all experts reported that no important content was missing. No item was removed at this stage, and items with a clarity rating of 2 were monitored in the pilot.

3.3 Pilot Sample, Feasibility and Face Validity

All 30 participants answered all 67 items (2,010 responses), with no missing or out-of-range values. No participant showed straight-lining; the within-person standard deviation ranged from 0.67 to 1.16. The sample included 14 men (46.7%) and 16 women (53.3%), aged 24–50 years (mean 38.2, SD 7.76); 18 (60.0%) were Saudi nationals, and the professional quotas (8/8/7/7) were met. Mean completion time was 15.0 minutes (SD 3.21, range 9–21). Participants rated overall clarity at 4.20 (SD 0.71) and relevance at 4.40 (SD 0.56). Thirteen participants (43.3%) marked at least one item as unclear, producing 18 flags across eight items. Four items reached the 10% review threshold: BR3 (5 participants, 16.7%), and PEOU10, BR12 and OR2 (3 participants each, 10.0%). ER4, TR8, ER2 and PU9 were each flagged by one participant.

3.4 Internal Consistency and Item Performance

Cronbach's alpha was at least .70 in all eight subscales, ranging from .740 for ER to .907 for OR (Table 4). Confidence intervals were wide, as expected with n = 30, and the lower limits for OC (.689) and ER (.564) fell below .70, so adequate reliability cannot yet be confirmed for these two subscales.

Table 4: Internal consistency by subscale (n = 30)

Subscale	Items	Cronbach's alpha	95% CI	Mean inter-item r (range)	Interpretation
PU	10	.898	.834–.945	.472 (.312 to .708)	Good
PEOU	10	.876	.798–.933	.421 (−.036 to .767)	Good
OR	9	.907	.847–.950	.516 (.320 to .766)	Excellent
OC	5	.818	.689–.904	.485 (.377 to .613)	Good
TR	8	.851	.755–.920	.417 (.053 to .698)	Good
ER	6	.740	.564–.861	.312 (.121 to .593)	Acceptable
BR	12	.818	.705–.901	.281 (−.028 to .616)	Good
AR	7	.865	.776–.928	.488 (.314 to .714)	Good

Confidence intervals from the F distribution (Feldt et al., 1987). Three items had corrected item–total correlations below .30, and alpha would rise if each were deleted: PEOU10 (r = .163; alpha .876 to .897), ER2 (r = .238; .740 to .761) and BR12 (r = .284; .818 to .823). PEOU10 was also near zero with other PEOU items, for example PEOU5 (r = −.036) and PEOU6 (r = .009). TR1 (r = .419) and BR6 (r = .323) produced negligible alpha gains on deletion (.002 and .001) and were retained. All other items had item–total correlations between .323 and .878.

3.5 Distributions and Relationships Between Subscales

Subscale scores were approximately normal (skewness −0.70 to 0.32; kurtosis −0.98 to 0.16); the Shapiro–Wilk test was significant only for OC (W = .922, p = .030). Item means were highest for perceived usefulness (3.87–4.33) and lowest for data quality (TR6, 2.93) and technological

compatibility (TR8, 2.87). As TAM predicts, PU and PEOU correlated strongly (r = .69), and both correlated with AR (r = .51 and .55) (Table 5). OR and OC were highly correlated (r = .74), suggesting conceptual overlap. The BR subscale was unrelated to every other subscale (|r| ≤ .21), including AR (r = −.04). TR correlated weakly with the TAM subscales (r = .02 and .04) and with AR (r = .27).

Table 5: Pearson correlations between subscale scores (n = 30)

Subscale	PU	PEOU	OR	OC	TR	ER	BR	AR
PU	1							
PEOU	.69**	1						
OR	.37*	.39*	1					
OC	.44*	.36	.74**	1				
TR	.02	.04	.43*	.34	1			
ER	.59**	.36	.34	.33	.26	1		
BR	.05	.02	.00	.08	−.13	.21	1	
AR	.51**	.55**	.53**	.53**	.27	.28	−.04	1

* p < .05 (|r| ≥ .37); ** p < .01 (|r| ≥ .47), two-tailed. Significance shown for two-decimal values; .36 lies at the .05 boundary (critical r = .361, df = 28).

3.6 Preliminary Structure of the Adoption Readiness Subscale

Principal axis factoring of the seven AR items gave a Kaiser–Meyer–Olkin value of .806 and a significant Bartlett test ($\chi^2 = 88.448$, df = 21, p < .001). The first eigenvalue was 3.959 and the second .985, so one factor was retained, explaining 49.9% of variance. Loadings ranged from .536 (AR1) to .833 (AR2). The two PHC-level items had the lowest loadings (AR1 .536, AR4 .566) and communalities (.288 and .321), a weak sign of a separate organisational facet that did not amount to a two-factor solution.

3.7 Integration of Expert and Pilot Evidence

Expert and pilot evidence agreed for some items and diverged for others (Table 6). BR3, PEOU10 and TR8 had an I-CVI of .83 on both criteria and were also flagged by participants. BR12 and ER4 were rated 1.00 by all experts but were flagged in the pilot. Six of the eight pilot-flagged items (75%) had a clarity I-CVI below 1.00, compared with 28 of the 59 unflagged items (47%) (Fisher's exact p = .26), so the two sources did not predict each other.

Table 6: Expert and pilot evidence for items flagged in the pilot

Item	Pilot flags (n)	I-CVI relevance / clarity	Experts rating 2 (relevance / clarity)	Corrected item–total r	Reading
BR3	5	.83 / .83	E5 / E3	.440	Both sources concur
PEOU10	3	.83 / .83	E4 / E1	.163	Both sources concur
BR12	3	1.00 / 1.00	– / –	.284	Pilot only
OR2	3	1.00 / .83	– / E6	.720	Weak expert signal
ER4	1	1.00 / 1.00	– / –	.522	Pilot only
TR8	1	.83 / .83	E1 / E2	.593	Both sources concur
ER2	1	1.00 / .83	– / E6	.238	Weak expert signal
PU9	1	1.00 / .83	– / E1	.645	Weak expert signal

Items were revised when the item–total correlation was below .30 with alpha rising on deletion (PEOU10, ER2, BR12) or when at least 10% of participants found the wording unclear (BR3, PEOU10, BR12). Four items were revised (Table 7). OR2 was flagged by three participants but correlated highly with its subscale (r = .72), so its validated source wording was kept. ER4, TR8 and PU9 were retained. The four revised items were re-rated by the same six experts and each achieved an I-CVI of 1.00 for relevance and clarity (S-CVI/Ave 1.00). The resulting version 2 of the instrument retains all 67 items for the main survey.

Table 7: Items revised after the pilot

Item	Original wording	Evidence	Revised wording
PEOU10	Familiarity with general AI technologies (e.g., generative AI tools) makes me more comfortable using AI in my work.	Item–total r = .16; alpha would rise; flagged by 3; measured familiarity rather than ease of use	I would find it easy to use general AI tools (e.g., generative AI tools) in my work.
BR3	Using AI does not guarantee the completeness and accuracy of patient data in my PHC.	Flagged by 5 (16.7%); negation confusing	I am concerned that AI cannot ensure that patient data in my PHC are complete and accurate.
BR12	AI has the potential to disrupt current work processes, requiring the adoption of new ones.	Item–total r = .28; alpha would rise; flagged by 3; direction unclear	I am concerned that AI would disrupt current work processes in my PHC.
ER2	My PHC receives adequate regulatory guidance from the Ministry of Health regarding AI and BDA use.	Item–total r = .24; alpha would rise; flagged by 1	The Ministry of Health has issued clear guidance on the use of AI and BDA in Primary Health Care.

4. Discussion

4.1 Principal Findings

This study provides the first psychometric evidence for an instrument that measures readiness to adopt AI and BDA among practising PHC professionals in Saudi Arabia across both TAM and TOE dimensions. Content validity was strong: every item reached an I-CVI of .83 or higher, all subscales met the .90 criterion for relevance, and the instrument as a whole met it for clarity. The pilot showed reasonable feasibility (15 minutes), good face validity with a small number of items needing rewording, and acceptable internal consistency in all eight subscales. Reliability was, however, imprecisely estimated, and two subscales (OC and ER) could not yet be confirmed as reliable. Table 8 summarises the evidence by source of validity evidence.

Table 8: Summary of validity and reliability evidence

Evidence source	Finding	Status
Test content	All 67 items I-CVI ≥ .83; S-CVI/Ave .93 (relevance) and .92 (clarity); experts judged the item pool complete	Supported; clarity of BR and AR below .90
Response processes	Mean clarity 4.20/5; 43.3% flagged at least one item; four items above the 10% threshold; 15 minutes	Supported after revision of four items
Internal structure	Alpha .740–.907; three items with r < .30; AR one-factor	Preliminary; CFA required
Relations to other variables	PU–PEOU .69; PU and PEOU with AR .51 and .55; OR–OC .74; BR unrelated to other subscales	Partly consistent with theory; OR–OC overlap and BR isolation to be tested
Reliability	Internal consistency acceptable in all subscales; wide intervals; no test–retest	Preliminary

4.2 Interpreting the Content Validity Evidence

The high content validity indices should be read alongside the raw rating distribution. Ratings clustered at 3 and 4, and no item received a rating of 1. Experts tend to endorse items adapted from published scales, and the brief, uniform written comments added little qualitative information to the numerical indices. The panel combined administrators, clinicians and digital health experts, matching the intended respondents, but it was small and drawn entirely from Al Madinah PHC centres, which increases contextual relevance while limiting an external perspective. The contrast between high raw agreement (about 85%) and slightly negative Fleiss' kappa illustrates the known paradox of chance-corrected statistics when almost all ratings fall in one category. In this situation the modified kappa, which is tied to the I-CVI, is more informative (Polit et al., 2007). The weak correspondence between expert ratings and pilot flags (Fisher's $p = .26$) shows that expert review and respondent testing capture different problems. Experts judged items against construct definitions, whereas frontline staff struggled with negations (BR3), mixed constructs (PEOU10) and statements of unclear direction (BR12). Both stages were therefore needed.

4.3 Reliability, Item Performance and Subscale Relationships

The alpha values (.740–.907) are comparable with those reported for the source instruments, including ORIC (Shea et al., 2014) and MAIRS-MS (Karaca et al., 2021). The three weak items had identifiable causes. PEOU10 described the effect of familiarity on comfort rather than ease of use, ER2 combined adequacy and regulatory source in one statement, and BR12 could be read as either a benefit or a concern. The revised wording addresses each problem, and the reliability of the revised subscales will be estimated in the main survey. The pattern of correlations is partly consistent with theory. The strong PU–PEOU correlation and the moderate associations of both with adoption readiness agree with TAM (Davis, 1989) and with extended TAM findings in health care (Holden & Karsh, 2010; Rathore et al., 2026). The OR–OC correlation of .74 is high enough to question whether culture and climate form a construct distinct from organisational readiness. This agrees with evidence that culture shapes structure and technology readiness in PHC organisations (Kruszyńska-Fischbach et al., 2022), and discriminant validity should be tested in the main survey using the Fornell–Larcker criterion and the heterotrait–monotrait ratio (Fornell & Larcker, 1981; Henseler et al., 2015).

The isolation of the barriers subscale is notable. Saudi professionals combine positive attitudes with substantial worries about data quality, deskilling and job displacement (Alsaedi et al., 2024; Elnaggar et al., 2023). The present data suggest that such concerns may coexist with perceived usefulness and intention rather than oppose them. BR may therefore function as a parallel dimension of perceived risk rather than as the negative pole of readiness, a possibility consistent with findings that perceived risk predicts intention independently of ease of use (Zheng et al., 2025). With $n = 30$ this interpretation remains tentative. The low means for data quality (TR6) and compatibility of existing systems (TR8) agree with national evidence of incomplete EHR coverage and data-quality concerns in Saudi PHC (Alzghaibi & Hutchings, 2025; Alsaedi et al., 2024). The weak association of TR with the TAM subscales suggests that professionals judge AI's usefulness independently of their centre's infrastructure, which supports measuring individual and organisational readiness together.

4.4 Implications for Instrument Use and Practice

The one-factor solution for AR supports scoring it as a single outcome, but the lower loadings of the two PHC-level items (AR1, AR4) suggest that organisational and personal readiness may separate in a larger sample. Analysts should check this before combining them. For health clusters, the instrument is intended to provide a baseline across individual, organisational, technological and environmental dimensions before AI or BDA tools are procured. Once validated, it can help identify centres where

training, data infrastructure or governance should come first, in line with national AI and data policies (Saudi Data and Artificial Intelligence Authority [SDAIA], 2020).

4.5 Limitations and Further Validation

The pilot sample of 30 was adequate for detecting comprehension problems and for preliminary reliability, but it produced wide confidence intervals and did not permit exploratory or confirmatory factor analysis of the full instrument. Participants came from five centres in one region, which limits generalisability. Test–retest reliability, convergent validity with established instruments and measurement invariance across professional groups were not assessed. The expert panel was small and local, and written comments provided little qualitative evidence. The chronic disease coordinator role was captured under "Other" in Section A and should be listed separately. The item sources Alam et al. (2024) and Alorfi and Alsaadi (2025) should be checked for exact item provenance, since the latter concerns manufacturing rather than health care. The main survey should therefore (a) estimate the reliability of the revised subscales, including composite reliability and test–retest stability in a subsample; (b) conduct exploratory factor analysis on one random half and confirmatory factor analysis on the other, judged against conventional fit criteria (Hu & Bentler, 1999); (c) test discriminant validity between OR and OC and the position of BR within the measurement model; and (d) conduct cognitive debriefing of the BR and AR items, whose clarity indices were lowest.

5. Conclusion

The AI and BDA Adoption Readiness Instrument combines TAM-based individual appraisals with TOE-based organisational, technological and environmental readiness, perceived barriers and an overall readiness outcome in one 67-item questionnaire for PHC professionals. Expert review supported its content validity, and the pilot showed that it is feasible to complete in about 15 minutes, understandable after revision of four items, and internally consistent at an acceptable level in all eight subscales. The evidence is preliminary: confidence intervals were wide, the distinction between organisational readiness and organisational culture is uncertain, the barriers subscale behaved as a separate dimension, and the factor structure has not yet been tested. Version 2 of the instrument is ready for the main survey, where confirmatory factor analysis, discriminant validity testing and test–retest reliability should establish whether it can serve as a dependable tool for assessing AI and BDA readiness in Saudi health clusters.

References

Items marked [verify] have a bibliographic field that could not be confirmed and must be checked before submission.

Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. npj Digital Medicine, 1, 39. https://doi.org/10.1038/s41746-018-0040-6

Alam et al. (2024). [verify: full reference for the technological readiness item source not identified]

Alfadhli, E., Surrati, A. M. Q., Masoud, R. S., Gadi, Y. A., Alahmadi, W. A., & Turkistani, M. K. (2025). The status of metabolic control in patients with diabetes attending primary care clinics in Madinah, Saudi Arabia. Medicina, 61(10), 1856. https://doi.org/10.3390/medicina61101856

Alorfi, A. S., & Alsaadi, N. (2025). Barriers to the adoption of big data analytics in Saudi Arabia's manufacturing sector: An interpretive structural modeling approach. Systems, 13(4), 250. https://doi.org/10.3390/systems13040250

Alsaedi, A. R., Alneami, N., Almajnoni, F., Alamri, O., Aljohni, K., Alrwaily, M. K., Eid, M., Budayr, A., Alrehaili, M. A., Alghamdi, M. M., Almutairi, E. D., & Eid, M. H. (2024). Perceived worries in the adoption of artificial intelligence among healthcare professionals in Saudi Arabia: A cross-sectional survey study. Nursing Reports, 14(4), 3706–3721. https://doi.org/10.3390/nursrep14040271

Alsudairy, N., Alahdal, A., Alrashidi, M., Altashkandi, D., Alzaidi, S., Alghamdi, A., & Alzahrani, S. (2025). Artificial intelligence in primary care decision-making: Survey of healthcare professionals in Saudi Arabia. Cureus, 17(4), e81960. https://doi.org/10.7759/cureus.81960

Alzghaibi, H., & Hutchings, H. A. (2025). Barriers to the implementation of large-scale electronic health record systems in primary healthcare centers: A mixed-methods study in Saudi Arabia. Frontiers in Medicine, 12, 1516714. https://doi.org/10.3389/fmed.2025.1516714

Boettiger, D. C., Lin, T. K., Almansour, M., Hamza, M. M., Alsukait, R., Herbst, C. H., ... Kattan, F. (2023). Projected impact of population aging on non-communicable disease burden and costs in the Kingdom of Saudi Arabia, 2020–2030. BMC Health Services Research, 23, 1381. https://doi.org/10.1186/s12913-023-10309-w [verify full author list]

Borges do Nascimento, I. J., Marcolino, M. S., Abdulazeem, H. M., Weerasekara, I., Azzopardi-Muscat, N., Gonçalves, M. A., & Novillo-Ortiz, D. (2021). Impact of big data analytics on people's health: Overview of systematic reviews and recommendations for future studies. Journal of Medical Internet Research, 23(4), e27275. https://doi.org/10.2196/27275

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. American Journal of Mental Deficiency, 86(2), 127–137.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research, and Evaluation, 10(7), 1–9. https://doi.org/10.7275/jyj1-4868

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334. https://doi.org/10.1007/BF02310555

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

El Bcheraoui, C., Basulaiman, M., Tuffaha, M., Daoud, F., Robinson, M., Jaber, S., Mikhitarian, S., Memish, Z. A., Al Saeedi, M., AlMazroa, M. A., & Mokdad, A. H. (2014). Status of the diabetes epidemic in the Kingdom of Saudi Arabia, 2013. International Journal of Public Health, 59(6), 1011–1021. https://doi.org/10.1007/s00038-014-0612-4

Elnaggar, M., Alharbi, Z. A., Alanazi, A. M., Alsaiari, S. O., Alhemaidani, A. M., Alanazi, S. F., & Alanazi, M. M. (2023). Assessment of the perception and worries of Saudi healthcare providers about the application of artificial intelligence in Saudi health facilities. Cureus, 15(8), e42858. https://doi.org/10.7759/cureus.42858

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. Applied Psychological Measurement, 11(1), 93–103. https://doi.org/10.1177/014662168701100107

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378–382. https://doi.org/10.1037/h0031619

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50. https://doi.org/10.2307/3151312

Ghaleb, E. A. A., Dominic, P. D. D., Fati, S. M., Muneer, A., & Ali, R. F. (2021). The assessment of big data adoption readiness with a technology–organization–environment framework: A perspective towards healthcare employees. Sustainability, 13(15), 8379. https://doi.org/10.3390/su13158379 [check published correction]

Health Sector Transformation Program. (2023). Health Sector Transformation Program: Achievements report 2023. Saudi Vision 2030. https://vision2030.gov.sa/media/c3ld5ltz/hstp-annual-report-2023-en.pdf

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Holden, R. J., & Karsh, B.-T. (2010). The technology acceptance model: Its past and its future in health care. Journal of Biomedical Informatics, 43(1), 159–172. https://doi.org/10.1016/j.jbi.2009.07.002

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. Educational and Psychological Measurement, 70(3), 394–400. https://doi.org/10.1177/0013164409355692

Karaca, O., Çalışkan, S. A., & Demir, K. (2021). Medical artificial intelligence readiness scale for medical students (MAIRS-MS): Development, validity and reliability study. BMC Medical Education, 21, 112. https://doi.org/10.1186/s12909-021-02546-6

Kruszyńska-Fischbach, A., Sysko-Romańczuk, S., Napiórkowski, T. M., Napiórkowska, A., & Kozakiewicz, D. (2022). Organizational e-health readiness: How to prepare the primary healthcare providers' services for digital transformation. International Journal of Environmental Research and Public Health, 19(7), 3973. https://doi.org/10.3390/ijerph19073973

Kueper, J. K., Terry, A. L., Zwarenstein, M., & Lizotte, D. J. (2020). Artificial intelligence and primary care research: A scoping review. Annals of Family Medicine, 18(3), 250–258. https://doi.org/10.1370/afm.2518

Lynn, M. R. (1986). Determination and quantification of content validity. Nursing Research, 35(6), 382–385.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health, 29(5), 489–497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health, 30(4), 459–467. https://doi.org/10.1002/nur.20199

Rahimi, B., Nadri, H., Lotfnezhad Afshar, H., & Timpka, T. (2018). A systematic review of the technology acceptance model in health informatics. Applied Clinical Informatics, 9(3), 604–634. https://doi.org/10.1055/s-0038-1668091

Rathore, J., Singh, A., & Rao, U. (2026). Validation of an extended technology acceptance model framework incorporating organisational culture and trust in AI usage within hospitals: A

cross-sectional study. Journal of Clinical and Diagnostic Research, 20(4), IC06–IC11. https://doi.org/10.7860/JCDR/2026/84554.22732

Saudi Data and Artificial Intelligence Authority. (2020). National strategy for data & AI. https://sdaia.gov.sa/en/SDAIA/SdaiaStrategies/Pages/NationalStrategyForDataAndAI.aspx [verify on official page]

Shea, C. M., Jacobs, S. R., Esserman, D. A., Bruce, K., & Weiner, B. J. (2014). Organizational readiness for implementing change: A psychometric assessment of a new measure. Implementation Science, 9, 7. https://doi.org/10.1186/1748-5908-9-7

Tornatzky, L. G., & Fleischer, M. (1990). The processes of technological innovation. Lexington Books.

Turner, M., Kitchenham, B., Brereton, P., Charters, S., & Budgen, D. (2010). Does the technology acceptance model predict actual use? A systematic literature review. Information and Software Technology, 52(5), 463–479. https://doi.org/10.1016/j.infsof.2009.11.005

Viechtbauer, W., Smits, L., Kotz, D., Budé, L., Spigt, M., Serroyen, J., & Crutzen, R. (2015). A simple formula for the calculation of sample size in pilot studies. Journal of Clinical Epidemiology, 68(11), 1375–1379. https://doi.org/10.1016/j.jclinepi.2015.04.014

Weiner, B. J. (2009). A theory of organizational readiness for change. Implementation Science, 4, 67. https://doi.org/10.1186/1748-5908-4-67

Wolf, R. M., Channa, R., Liu, T. Y. A., Zehra, A., Bromberger, L., Patel, D., Ananthakrishnan, A., Brown, E. A., Prichett, L., Lehmann, H. P., & Abramoff, M. D. (2024). Autonomous artificial intelligence increases screening and follow-up for diabetic retinopathy in youth: The ACCESS randomized control trial. Nature Communications, 15, 421. https://doi.org/10.1038/s41467-023-44676-z

World Health Organization. (2018). Noncommunicable diseases country profiles 2018: Saudi Arabia. https://cdn.who.int/media/docs/default-source/country-profiles/ncds/sau_en.pdf

Zheng, R., Jiang, X., Shen, L., He, T., Ji, M., Li, X., & Yu, G. (2025). Investigating clinicians' intentions and influencing factors for using an intelligence-enabled diagnostic clinical decision support system in health care systems: Cross-sectional survey. Journal of Medical Internet Research, 27, e62732. https://doi.org/10.2196/62732