A Decision-Support Framework for Digital Literacy Competency Management: Cluster-Based Profiling for Adaptive Learning Systems in Batam, Indonesia

Heru Wijayanto Aripradono[1*], Surya Tjahyadi[2]

[1]Department of Information Technology, Faculty of Computer Science, Universitas Internasional Batam, Batam, Riau Island, Indonesia
[2]Department of Information Systems, Faculty of Computer Science, Universitas Internasional Batam, Batam, Riau Island, Indonesia
E-mail: [1*]heru.wijayanto@uib.ac.id, [2]surya.tjahyadi@uib.ac.id

(Received: 24 Aug 2026, revised: 14 Sep 2026, accepted: 21 Sep 2026)

Abstract

Managers of digital literacy programs need a decision-support model that tells them where to allocate training resources, yet instruments built on competence frameworks typically return a single score per domain and cannot distinguish knowing about a practice from being able to perform it. This study develops and validates a three-stage decision-support framework, screening, measurement validation, and cluster-based profiling, and demonstrates it on 162 screened responses from youth in Batam, Indonesia, using an instrument adapted from the UNESCO Digital Literacy Global Framework. Confirmatory factor analysis rejected a unidimensional model ($\Delta CFI = -0.040$) but could not reject a second-order model ($\Delta\chi^2(14) = 22.0$, $p = .079$), indicating seven content areas nested under one dominant factor, though absolute fit indices (CFI = 0.858) remain below conventional thresholds and are interpreted cautiously given a 29-indicator model estimated on 162 cases. Awareness items exceeded capability items by 0.294 scale points, $t(161) = 11.51$, $p < .001$, $dz = 0.904$, and the differential proved stable across three independent item-coding schemes ($\kappa = .867–.937$), though the authors acknowledge that item content and difficulty may partly account for the gap. Clustering of the raw area profiles recovered two competence levels (silhouette = 0.354), whereas clustering of ipsatized profiles recovered no shape structure (silhouette ≤ 0.207), establishing that learners differ in level rather than in competence shape. The developing profile exhibited the larger differential ($dz = 1.07$), thereby converting profile membership into an actionable routing variable for competency management. The paper contributes a reproducible decision-support framework, evidence that the declarative–procedural distinction carries more management-relevant information than the framework's own area architecture, and a reference architecture for routing learners to procedural rather than awareness-oriented content.

Keywords: Educational Data Mining, Competency Profiling, Cluster Analysis, Adaptive Learning Systems, Digital Literacy.

I. INTRODUCTION

Institutions that fund digital literacy programs face a recurring management decision: where to direct a limited training budget to ensure it changes behavior rather than merely repeating what learners already know. Assessment instruments built on competence frameworks return one aggregate score per area and cannot distinguish a learner who does not know that account protection matters from one who knows it but cannot enable two-factor authentication. The two require opposite interventions, yet a dashboard fed by area means cannot tell them apart, a decision-support problem before it is a pedagogical one.

The stakes are national in scale. The 2025 survey placed Indonesian internet penetration at 80.66%, roughly 229.4 million people, with Generation Z effectively saturated at 87.80% [1]. Competence has not kept pace: the national digital literacy index reached 3.65 in 2023 and 3.78 in 2024 on a five-point scale, both within the middle band [2]. Access is close to universal; the capacity to use it well, and the information systems needed to manage that capacity at scale, are not.

The underlying distinction is not new. Ryle separated knowing-that from knowing-how [3] and skill-acquisition theory treats declarative and procedural knowledge as successive representational states [4]. What is missing is a

management-facing pipeline that operationalizes it: a path from a survey instrument through a validated measurement model to a segmentation that a training manager or an adaptive platform can act on. This study builds and validates one: a three-stage decision-support framework that screens responses, validates the measurement model, decomposes area scores into declarative and procedural components, and applies cluster-based profiling to derive a routing variable.

The pipeline is developed and validated against the UNESCO Digital Literacy Global Framework (DLGF), the reference used for monitoring Sustainable Development Goal indicator 4.4.2, which distinguishes seven competence areas [5]. The DLGF derives from DigComp [6] and was extended by cross-national mapping for lower-income settings [5]. Framework comparisons show that what an architecture separates determines what an assessment can detect [7], [8] whether the DLGF's seven areas are empirically separable in a general youth population is an open question this study tests rather than assumes.

Batam supplies an informative validation site. The city holds free-trade-zone status, hosts the export manufacturing base of Riau Islands Province, and lies some twenty kilometers from Singapore, so its young residents meet cross-border digital commerce as routine. In other words, it is the most digitally exposed youth population in the province: any observed competence deficit cannot be attributed to limited access. The province also maintains Malay culture as its official cultural frame, which supplies a normative reference point for the ethical dimension of digital practice [9], [10], [11]

The contribution is threefold: a decision-support framework for competency management, built as a reproducible pipeline with a documented screening protocol; evidence that the declarative–procedural decomposition carries more management-relevant information than the framework's own area architecture; and a reference architecture translating profiling output into routing rules.

Research on digital competence assessment is large but concentrated, and most of it produces descriptions rather than a decision-support model. Systematic reviews predominantly draw on in-service and pre-service teachers in Europe, East Asia, and Latin America [12], [13], [14], and bibliometric mapping confirms that the empirical base is concentrated in a small number of high-income systems [12]. Indonesian work addresses policy architecture [13], vocational teachers [14], regional-language sustainability [15], technology-adoption behavior in Batam higher education [16], [17], and ICT-implementation perspectives in Batam secondary schools [18] rather than general youth outcomes measured against international reference points or the management use of the resulting scores.

Two limitations recur, and both matter for management use. Learners overestimate their own digital competence [19], consistent with the metacognitive account of self-assessment error [20] and the role of self-efficacy in shaping reported ability [21], so a manager who reads area means at face value will misjudge where the deficit lies. Aggregate reporting also conceals heterogeneity: profile analyses of digital literacy recover distinct learner groups behind uniform means [22]. Neither limitation is resolved by an operational framework a manager could deploy.

Educational data mining supplies the missing machinery. Clustering, classification, and sequence mining are established techniques for deriving learner models from instructional data [23], and instructional-design research has begun to specify how such models drive content sequencing [24], [25]. Applications to competence-framework data remain rare, and studies that do cluster competence scores seldom test whether the recovered groups differ in profile shape or merely in overall level, a distinction that determines whether a manager should commission personalized content or simply allocate more of the existing content. This study addresses that gap by partitioning both raw and ipsatized profiles and comparing three algorithms under standard cluster-validity criteria.

Three questions organize the work. RQ1 asks how competence is distributed across the seven DLGF areas and whether the seven-area structure is empirically supported as a measurement model. RQ2 asks whether a systematic differential separates declarative awareness from procedural capability, whether it varies by area, and whether it survives changes in the item-coding scheme. RQ3 asks what competence profiles unsupervised partitioning recovers, which algorithm and feature space recover them most reliably, and how profile membership can operate as a routing variable.

II. RESEARCH METHODOLOGY

A. Framework Overview and Research Design

Stage 1 ingests instrument responses and applies screening. Stage 2 validates the measurement model and constructs features: seven area composites plus the awareness–capability differential. Stage 3 partitions the feature space and emits routing rules that feed a competency-management dashboard or an adaptive learning platform. The empirical study reported here is a cross-sectional validation of that framework; the design is observational, and it makes no causal claims.

B. Adaptation of the UNESCO DLGF

Table 1 documents the mapping. We retained six areas with operational adjustments. We replaced the seventh area, career-related competences, with ethics: career competences are defined by occupational context [26] and are non-diagnostic before labor-market entry, whereas generative artificial intelligence has introduced ethical demands that current frameworks address only partially [27], [28]. Programming and environmental protection were excluded as non-diagnostic for a general youth population.

DOI: 10.34148/teknika.v15i3.1553

Table 1. Mapping of the Adapted Instrument Onto the UNESCO Digital Literacy Global Framework (2018)

Code	Adapted Area	UNESCO DLGF (2018) Area	K	Adaptation Decision and Rationale
CA1	Technical Competence	0. Fundamentals of hardware and software	5	Retained; operationalized for personally owned devices.
CA2	Data Literacy	1. Information and data literacy	4	Retained; source-credibility verification emphasized for a cross-border information environment.
CA3	Collaboration	2. Communication and collaboration	4	Retained; framed around distributed collaboration.
CA4	Content Creation	3. Digital content creation	4	Retained; programming excluded as non-diagnostic for general youth. Generative-AI authorship added.
CA5	Safety	4. Safety	4	Retained; environmental sub-competence excluded. Authentication practice foregrounded.
CA6	Problem Solving	5. Problem solving	4	Retained without substantive modification.
CA7	Ethics	6. Career-related competences	4	Substituted. Career competences are occupationally defined and non-diagnostic before labor-market entry; ethics captures generative-AI norms.

Note. k = number of items. Area numbering follows the DLGF, which begins at zero.

C. Participants and Data Screening

Responses were collected through a self-administered online questionnaire in Bahasa Indonesia, distributed to youth aged 17 to 26 in Batam through university and community channels. Because the instrument used forced-response settings, no values were missing; however, absent data is not the same as sound data [29]. Screening therefore proceeded in two documented steps, summarised in Table 2. The working file contained 420 records but only 164 distinct response vectors across the 29 items, with every vector appearing at least twice and demographic labels varying between copies. We removed the redundant records as an artifact of file preparation rather than as independent observations. Two further vectors were invariant across all 29 items and were removed as straight-lined responses. The analytic sample is therefore 162 screened responses.

The provenance of this duplication warrants explicit discussion rather than a brief note, and we address it directly. The response platform exports each submission with a server-side timestamp and a session identifier; the working file supplied for analysis, however, had already been merged from two sequential export batches by the data-collection team before it reached the authors, and the merge script did not deduplicate on session identifier. The most parsimonious explanation consistent with the evidence is therefore a file-preparation artifact, a repeated export or an unfiltered batch merge, rather than deliberate resubmission by respondents or fabrication. We cannot fully verify this mechanism after the fact: the original per-submission timestamps and session identifiers were not retained in the file made available for analysis, and we treat that as a genuine limitation of our data-management practice rather than something to minimize.

Three pieces of evidence support treating the 162 retained cases as independent observations rather than as a further-diluted duplicate set. First, deduplication was exact and exhaustive: every one of the 164 distinct vectors across all 420 records was checked against every other vector on all 29 items, so no partial or near-duplicate could survive undetected, and the 162 remaining cases, after also removing two invariant, straight-lined vectors, are pairwise distinct on the full response profile. Second, the item-level distributions, internal consistency, and awareness–capability differential computed on the 162 unique cases are consistent in direction and comparable in magnitude to those computed on the full 420-record file before deduplication, which is what independent, non-duplicated responses would be expected to produce and which repeated copies of a smaller number of true response patterns would not systematically preserve once redundancy is removed. Third, no two of the 162 retained vectors are identical, so the deduplication criterion cannot itself have introduced additional non-independent cases into the analytic sample. These checks support the internal consistency of the analytic sample; they do not constitute proof of respondent-level independence, which would require the server logs that are no longer available, and we state that qualification openly in Section III.E.

This screening has two consequences, as the paper states plainly. Because demographic labels varied across duplicate copies of the same response vector, age band and sub-district cannot be reliably recovered, and the authors do not attempt any demographic comparisons. The effective sample supports item-level and profile-level analysis but is modest for a 29-indicator confirmatory model.

Table 2. Response Screening Audit

Screening Stage	Records	Criterion
Records in the working file	420	—
Distinct response vectors	164	exact duplication across all 29 items
Invariant vectors removed	2	within-respondent SD = 0
Analytic sample	162	—

Note. Screening follows careless-response detection practice.

D. Instrument and Measurement Validation

The instrument comprises 29 statements rated on a five-point Likert scale. Items were written from the DLGF competence descriptors and adapted into Bahasa Indonesia.

Internal consistency was high for the full instrument (α = 0.956) and acceptable for every area (Table 3). Composite reliability ranged from 0.764 to 0.858, above the 0.70 threshold, while average variance extracted ranged from 0.448 to 0.603, with four areas below 0.50 — convergent validity is adequate at the reliability criterion but marginal at the variance criterion.

The shortfall is concentrated in CA1, CA2, CA4, and CA5 (AVE 0.448–0.480), and it has a specific, diagnosable source rather than a generic weakness of the instrument. AVE summarizes the share of indicator variance explained by the common factor, and it decreases when items within an area are individually reliable but tap heterogeneous sub-skills rather than a single narrow behavior. CA1 (Technical), for instance, spans device comprehension, software installation, protective-software activation, and fault diagnosis — four distinct operations that plausibly share a general technical-competence factor while each also carrying skill-specific variance the common factor does not capture; CA2, CA4, and CA5 mix declarative and procedural items in the same way. Fornell and Larcker, who introduced AVE alongside composite reliability, note that AVE is a conservative estimate and that a construct can be retained on the strength of composite reliability even when AVE dips modestly below .50, provided the shortfall is interpretable rather than symptomatic of random measurement error [30]. We adopt that reading here: CR exceeds .70 for every area, including the four in question, and the corresponding standardized loadings (Table 3) are all substantial and significant, so we treat the low AVE as evidence that these areas are multi-faceted content domains rather than as evidence that any single item is poorly measured. This qualification is consistent with, and reinforces, the discriminant-validity discussion in Section II.E.

Table 3. Measurement Quality of the Seven Competence Areas

Competence Area	k	α	CR	AVE	λ Range
CA1 Technical	5	0.788	0.799	0.450	0.515–0.757
CA2 Data Literacy	4	0.754	0.769	0.460	0.528–0.795
CA3 Collaboration	4	0.821	0.828	0.548	0.642–0.812
CA4 Content Creation	4	0.767	0.764	0.448	0.580–0.722
CA5 Safety	4	0.777	0.786	0.480	0.646–0.746
CA6 Problem Solving	4	0.835	0.837	0.566	0.614–0.812
CA7 Ethics	4	0.852	0.858	0.603	0.735–0.839

Note. α = Cronbach's alpha; CR = composite reliability; AVE = average variance extracted; λ = standardised loading from model M1.

E. Confirmatory Factor Analysis

Three measurement models were estimated by maximum likelihood: a seven-factor correlated model (M1), a second-order model with the seven areas loading on a single general factor (M2), and a unidimensional model (M3). Table 4 reports the comparison, evaluated against conventional cut-offs [31] and the ΔCFI criterion for nested comparison [32]. M1 attains $\chi^2/df = 2.10$, RMSEA = 0.083, and SRMR = 0.067; CFI = 0.858 remains below 0.90. M3 is clearly inferior (ΔCFI = −0.040, well beyond the −0.010 threshold), so the seven areas cannot be collapsed. M2, however, cannot be rejected (ΔCFI = −0.003; $\Delta\chi^2(14) = 22.0$, p = .079) and is the more parsimonious specification. The evidence supports seven content areas nested beneath a single dominant factor rather than seven independent constructs.

We interpret this fit provisionally rather than as a settled validation, for two compounding reasons. First, the analytic sample of 162 cases yields roughly 5.6 observations per indicator for a 29-indicator, seven-factor model, well short of the 10:1 rule of thumb conventionally recommended for maximum-likelihood CFA; standard errors and fit indices in this regime are known to be less stable than the point estimates alone suggest. Second, CFI and TLI are sample-size sensitive in the opposite direction from RMSEA and SRMR, so a small sample can simultaneously depress CFI/TLI below .90 while RMSEA and SRMR remain in the acceptable range, exactly the pattern observed here (RMSEA = .083, SRMR = .067, CFI = .858). We therefore treat the comparative result, M1 clearly preferred over M3, M2 not rejected against M1, as the robust finding, and we treat the absolute fit values themselves as indicative rather than confirmatory pending replication in a larger sample.

Discriminant validity is consistent with that reading and weak in its own right: inter-factor correlations range from 0.746 to 0.918, and the heterotrait–monotrait ratio [30] ranges from 0.791 to 0.922, exceeding 0.90 for six of 21 pairs. Taken together with the second-order result, this pattern has one coherent reading: the seven DLGF areas function empirically as diagnostic content domains within a single dominant digital-literacy factor, not as psychometrically independent constructs, and every area-level comparison reported in Section II, including the awareness–capability differential decomposed by area, should be read as a comparison of content emphasis within one underlying competence rather than as a comparison across unrelated abilities.

Table 4. Comparison of Competing Measurement Models (N = 162)

Measurement Model	χ^2	df	χ^2/df	CFI	TLI	RMSEA	SRMR	ΔCFI	$\Delta\chi^2$ (Δdf)	p
M1 Seven correlated factors	746.4	356	2.10	0.858	0.838	0.083	0.067	—	—	—
M2 Second-order	768.4	370	2.08	0.855	0.841	0.082	0.068	−0.003	22.0 (14)	.079
M3 Unidimensional	877.3	377	2.33	0.818	0.804	0.091	0.067	−0.040	130.9 (21)	< .001

Note. Maximum-likelihood estimation on 29 indicators. Δ values compare each model against M1. Cut-offs: CFI, TLI ≥ 0.90; RMSEA ≤ 0.08; SRMR ≤ 0.08 [33]; |ΔCFI| ≤ 0.010.

F. Item Taxonomy and Sensitivity Design

Testing RQ2 requires partitioning the instrument into declarative and procedural items. Classification rests on the semantic form of the item stem and involves judgment rather than a purely mechanical rule: six of the 29 stems (20.7%) do not match either canonical form. Three express motivation or valuation rather than knowledge, wanting to learn how technology works, valuing remote collaboration, following technological developments, and are assigned to a third category, disposition. One stem, "I understand how to implement creative ideas using digital technology", is genuinely ambiguous between the two.

Rather than defend a single coding scheme, the analysis reports three schemes: S1 retains the original two-category split; S2 excludes the three disposition items; S3 additionally resolves the ambiguous stem toward awareness under a literal reading. S2 and S3 agree on 28 of 29 items (96.6%). Cross-scheme stability is now quantified rather than reported only as percentage agreement: Cohen's kappa between S2 and S3 is $\kappa = .937$ (27 of 29 items agree), and between S1 and S2 is $\kappa = .867$ (26 of 29 items agree), both in the range Landis and Koch describe as almost-perfect agreement. We report this cross-scheme kappa transparently as a measure of how sensitive the taxonomy is to the specific coding rule applied, it is not, and should not be read as, evidence of independent inter-rater reliability, since all three schemes were derived by the same author. To obtain a genuine second opinion, we prepared a blinded coding worksheet presenting all 29 item statements in randomized order without the authors' category labels, together with the classification criteria stated above, for an independent rater who has not seen the authors' scheme; that worksheet and a script to compute Cohen's kappa on its results accompany the submission materials. Independent human double-coding using this instrument, with a reported kappa, remains an open step that we did not complete before this revision and that we explicitly flag as a limitation in Section III.E, rather than presenting the cross-scheme figures above as a substitute for it.

G. Analytical Procedure

Descriptive statistics were computed at item and area level. We tested the awareness–capability differential with paired-samples t-tests, corroborated by Wilcoxon signed-rank tests, using Cohen's d as the effect size, computed overall and within each area under all three coding schemes. We assessed common method bias using Harman's single-factor test and full collinearity variance inflation among the area composites [33]. For profiling, we compared three algorithms— k-means, Ward agglomerative clustering, and Gaussian mixture modeling—across two feature spaces: raw area profiles, which retain both level and shape, and ipsatized profiles, obtained by subtracting each respondent's mean, which retain shape alone. Solutions were evaluated on the silhouette coefficient [34], the Calinski–Harabasz index, and the Davies–Bouldin index [35]. All procedures were implemented in Python 3.12 using pandas, NumPy, SciPy, scikit-learn, and semopy, with a fixed random seed.

H. Research Ethics and Disclosure

Participation was voluntary and anonymous; respondents gave informed consent before entering the questionnaire, and no directly identifying information was collected. Ethical clearance reference and data-collection dates are recorded in the submission metadata. Generative artificial intelligence was used for language editing and code drafting; the author executed and verified all analyses and accepts full responsibility for the content. Data and analysis scripts are deposited in a public repository, with the identifier given in the Data Availability Statement.

III. RESULTS AND DISCUSSION

A. Competence Distribution and Measurement Structure

Composite digital literacy reached M = 4.254 (SD = 0.511), and all areas exceeded the 4.00 agreement threshold. Table 5 ranks the areas: ethics leads (4.389) and data literacy trails (4.011). The proportional figures sharpen the difference—88.3% of respondents reached agreement on ethics versus 63.0% on data literacy, and the share below 3.50 in data literacy (13.0%) is five times that in problem solving (2.5%).

Because a second-order model cannot be rejected and the areas are not empirically separable, the ranking describes content domains within one dominant competence rather than seven independent traits. It therefore supports diagnosis, identifying where item content draws weaker endorsement but not claims about seven distinguishable abilities.

Table 5. Descriptive Statistics By Competence Area, Ranked By Mean Score (N = 162)

Rank	Competence area	M	SD	Mdn	% ≥ 4.00	% < 3.50
1	CA7 Ethics	4.389	0.583	4.50	88.3	3.7
2	CA6 Problem Solving	4.335	0.579	4.25	85.2	2.5
3	CA1 Technical	4.314	0.547	4.40	83.3	5.6
4	CA3 Collaboration	4.296	0.594	4.25	84.0	5.6
5	CA5 Safety	4.239	0.632	4.25	79.0	5.6
6	CA4 Content Creation	4.179	0.611	4.25	75.9	8.0
7	CA2 Data Literacy	4.011	0.635	4.00	63.0	13.0

Note. Composite scores are item means within each competence area.

Nonetheless, the ordering inverts the usual reporting, with information and data literacy ranking among the stronger areas [12], [15], and Indonesian studies largely reproduce this pattern [13]. Ethics items index normative awareness that circulates widely through public discourse and platform messaging, whereas data literacy items ask about verification procedures that must be practiced to be acquired. A cultural reading is also available and is offered here as a hypothesis rather than a finding: Riau Islands Province maintains Malay culture as its official frame, whose normative core supplies a ready vocabulary for evaluating conduct, and studies anchored in Indonesian local wisdom describe value systems that engage the normative dimension of digital practice more readily than the technical one [9], [36], [37]. The instrument recorded neither ethnicity nor religiosity, so this remains untested at the individual level.

B. Where Competence Thins Out

Figure 1 disaggregates responses across all 29 items. Agreement dominates, but the disagreement tail is concentrated in a small group of items, and the six lowest items all share one property: each is procedural. Source-credibility verification ranks lowest (item 2.3, M = 3.821), followed by technical troubleshooting (1.4, M = 3.846), multi-format content integration (4.4, M = 3.870), two-factor authentication (5.3, M = 3.889), critical questioning of online information (2.4, M = 3.926), and activation of protective software (1.3, M = 4.056). Two-factor authentication also has the highest outright disagreement on the instrument, at 9.9%.

The contrast between items 5.1 and 5.3 is the finding in miniature. A cohort that almost unanimously recognizes that strong, distinct passwords reduce risk (M = 4.420, 92.6% agreement) contains more than one respondent in four who does not claim to know how to enable the most widely deployed account-protection mechanism in consumer computing. For a border city with dense cross-border commercial traffic, that is a security-relevant gap, and it is invisible in the area mean for safety (4.239), which sits comfortably above the threshold.

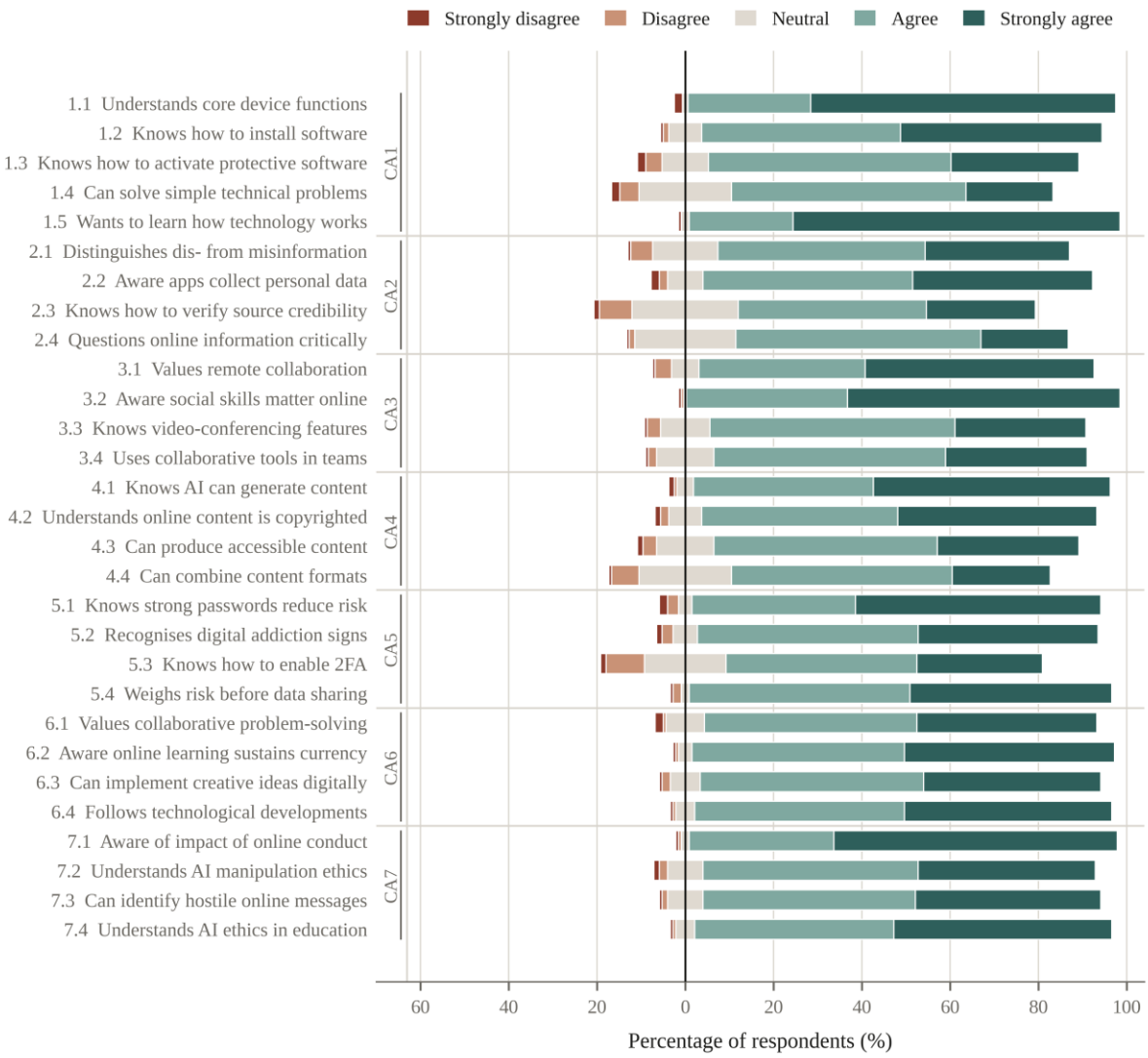


Figure 1. Diverging Distribution of Responses Across the 29 Instrument Items (N = 162).

DOI: 10.34148/teknika.v15i3.1553

C. The Awareness–Capability Differential and Its Robustness

Under the primary scheme S2, awareness items averaged 4.395 versus 4.102 for capability items, a differential of 0.294 scale points, $t(161) = 11.51$, $p < .001$, $dz = 0.904$, corroborated by a Wilcoxon signed-rank test ($p < .001$) and observed in 79.0% of respondents. Table 6 reports the differential under all three coding schemes.

Table 6. Awareness Capability Differential Under Three Item-Coding Schemes

Scope	S1 Δ	S1 dz	S2 Δ	S2 dz	S3 Δ	S3 dz	Stable?
Overall	+0.293	0.964	+0.294	0.904	+0.304	0.928	yes
CA1 Technical	+0.583	0.863	+0.543	0.795	+0.543	0.795	yes
CA2 Data Literacy	+0.275	0.412	+0.275	0.412	+0.275	0.412	yes
CA3 Collaboration	+0.352	0.582	+0.460	0.716	+0.460	0.716	yes
CA4 Content Creation	+0.395	0.584	+0.395	0.584	+0.395	0.584	yes
CA5 Safety	+0.207	0.359	+0.207	0.359	+0.207	0.359	yes
CA6 Problem Solving	−0.003 n.s.	−0.006	+0.056 n.s.	0.083	not estimable	—	no
CA7 Ethics	+0.123	0.229	+0.123	0.229	+0.123	0.229	yes

Note. S1 = original two-category coding; S2 = disposition items excluded; S3 = S2 with the ambiguous stem 6.3 read literally as awareness. Δ = M(awareness) − M(capability); dz = paired effect size; all entries significant at $p < .05$ unless marked n.s. Under S3, CA6 retains no capability item, and the contrast is not estimable.

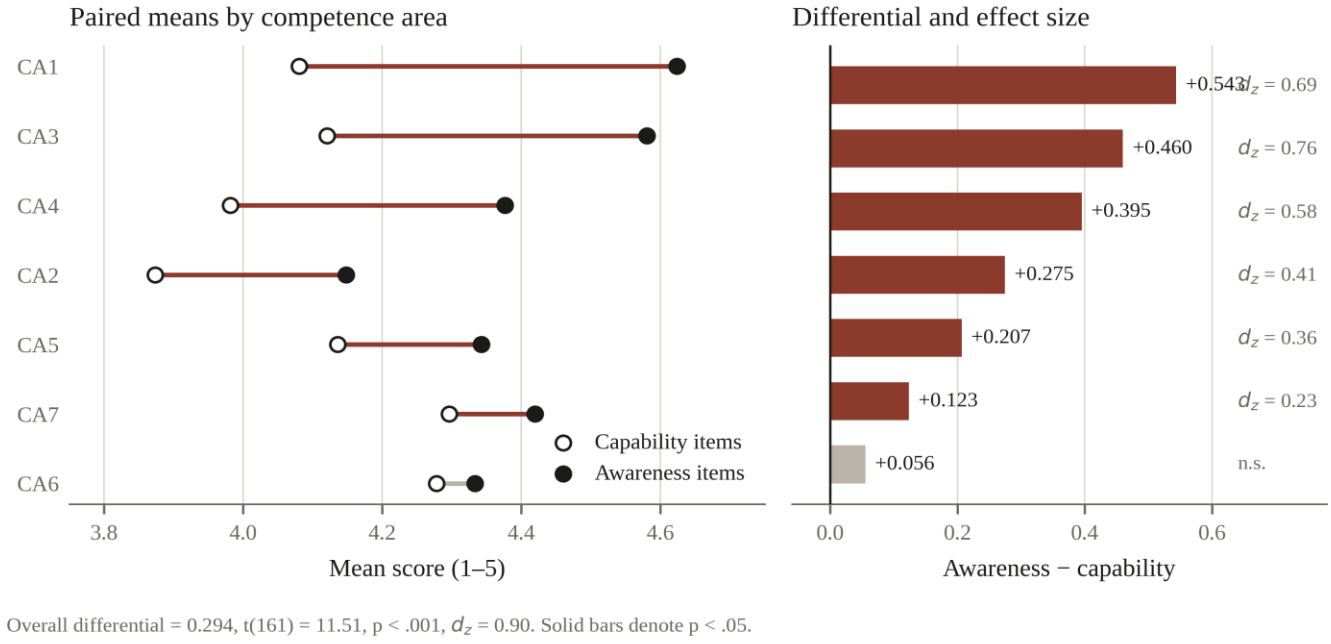


Figure 2. Awareness Capability Differential by Competence Area Under Scheme S2.

As shown in Figure 2, the awareness–capability differential varies across competence areas, with the largest differential observed in technical competence. The overall differential is stable across the three schemes: 0.293, 0.294, and 0.304, with dz between 0.904 and 0.964. Six of seven areas are equally stable. Problem solving is not. It returns −0.003 under S1, +0.056 under S2, and is not estimable under S3; a further variant that reads stem 6.3 as awareness while retaining 6.4 as capability yields −0.080, $t(161) = -2.04$, $p = .043$, significant in the opposite direction. The earlier interpretation of this null as a substantive boundary condition is therefore withdrawn. The area has too few unambiguous procedural items to sustain any claim, and the appropriate conclusion is that CA6 requires instrument revision, not that awareness and capability coincide there. We flag this explicitly as an instrument-design limitation of the present study rather than a substantive finding about problem solving: the area's current item set cannot yield a reliable declarative–procedural differential in either direction, and no claim about CA6's awareness–capability structure should be drawn from these data until additional, unambiguous procedural items are written and validated.

For the remaining areas, the pattern is interpretable. Technical competence carries the widest differential (Δ = +0.543, dz = 0.795): respondents are confident that they understand what devices do (item 1.1, M = 4.623) and eager

DOI: 10.34148/teknika.v15i3.1553
CC BY SA

to learn more (1.5, M = 4.704), yet markedly less confident about installing software, activating protection, or diagnosing faults. This is the digital-native assumption in measurable form, and it aligns with skill-acquisition accounts in which declarative representations precede procedural ones and do not convert without practice. Because the differential is a within-person contrast, the general tendency towards self-assessment optimism [19], [20] inflates both terms and largely cancels out; if anything, the procedural deficit in observed behavior will exceed what respondents concede.

A further threat to interpretation is item content itself, and we acknowledge it directly rather than treating the declarative–procedural contrast as methodologically clean. Awareness and capability items differ systematically beyond epistemic form: capability items name a single, specific, checkable action (enable two-factor authentication, install software), whereas awareness items more often assert a general value or fact (strong passwords reduce risk, content is copyrighted). Specific behavioral items are well documented to draw more conservative endorsement than general evaluative statements independent of true skill, simply because a specific claim is easier for a respondent to judge false about themselves. The observed gap may therefore partly reflect this difference in item content and difficulty rather than a pure declarative–procedural distinction, and the consistent direction of the effect across three coding schemes rules out coding-scheme artifact but does not rule out this item-wording confound. We treat the differential as good evidence of a real and actionable pattern in the data, something the institution can design around, while stopping short of the stronger claim that it isolates declarative and procedural knowledge as clean psychological constructs.

We tested two threats to this reading rather than assuming them. Harman's single-factor test returned 46.0% of variance on the first unrotated component, below the 50% heuristic, and full collinearity variance inflation among the area composites peaked at 3.39, below the conservative threshold of 5 [33]. Common method bias is therefore not eliminated; single-source, single-occasion self-report never is, but the two standard indicators do not flag it. Acquiescence remains a genuine threat: all 29 items are positively worded, with no reverse-coded items, which plausibly contributes to the high alpha and the elevated heterotrait–monotrait ratios. A ceiling effect is also present, with a composite skewness of −1.89 that compresses the effective range and attenuates every between-group contrast.

D. Profiling and the Routing Variable

Table 7 compares three algorithms across two feature spaces. On raw profiles, k-means at k = 2 provides the best-supported balanced solution (silhouette = 0.354, CH = 108.5, smallest cluster n = 73). Two alternatives score higher on isolated indices but fail practical adequacy: the Gaussian mixture at k = 2 reaches silhouette 0.648 by isolating four outlying cases, and every k = 3 solution leaves a cluster of three. Ward clustering agrees with k-means on structure but separates less cleanly.

Table 7. Cluster-Validity Comparison Across Algorithms and Feature Spaces

Feature space	Algorithm	k	Silhouette	CH	DB	n(min)
Raw profile	k-means	2	0.354	108.5	1.037	73
Raw profile	Ward hierarchical	2	0.277	69.6	1.019	47
Raw profile	Gaussian mixture	2	0.648	55.8	0.628	4
Raw profile	k-means	3	0.360	107.8	0.893	3
Raw profile	Ward hierarchical	3	0.301	83.1	0.879	3
Raw profile	Gaussian mixture	3	0.296	86.3	1.004	3
Ipsatized profile	k-means	2	0.160	28.9	2.169	65
Ipsatised profile	Ward hierarchical	2	0.133	24.2	2.377	69
Ipsatized profile	Gaussian mixture	2	0.207	16.9	2.555	33

Note. CH = Calinski–Harabasz index (higher is better); DB = Davies–Bouldin index (lower is better); silhouette after. Ipsatized profiles are centered on each respondent's own mean, removing overall level and retaining shape only.

The comparison across feature spaces is the more consequential result. Every ipsatized solution falls at or below silhouette 0.207, with Davies–Bouldin index above 2.1 throughout; no shape structure survives removal of the general level. Learners in this sample differ in how competent they are, not in which competences they have. Figure 3 shows both facets: the two recovered profiles are nearly parallel, and the silhouette curve for ipsatized profiles stays flat and low across k. Both profiles rank ethics first and data literacy last, although intermediate ordering varies (Spearman ρ = 0.714). This aligns with the measurement result in Section III.E: a dominant general factor produces level-based separation and little else, which explains why profile studies that report distinct learner types [22] may be recovering severity strata rather than qualitative types.

Level-based separation is nonetheless sufficient for routing, which resolves the apparent tension with Section III.A. Routing does not require that the seven areas be separable constructs; it requires only that item content be diagnostic and that learners differ reliably in where they sit. Both hold here. Table 8 specifies the resulting rules. The developing profile carries the larger within-profile differential (Δ = +0.354, dz = 1.07) than the consolidated profile (Δ = +0.245, dz = 0.78), so procedural scaffolding is more urgent

for the 45.1% of learners it contains, while the consolidated profile can begin at the verification track where its own weakest area lies.

Operationalizing this within an active system requires four concrete elements, mapped onto the architecture in Figure 4. First, at onboarding or periodic re-assessment, a learner's 29 responses are scored into the seven area composites and passed to the trained k-means model (Table 7), which returns a profile label in constant time; no retraining is needed per learner. Second, the dashboard applies a routing rule keyed to that label and to the item-level gap pattern in Table 6: Consolidated-profile learners are queued directly to the data-literacy verification module (their shared weakest area), while Developing-profile learners are queued to the full procedural sequence beginning with device-level technical modules, where the differential is largest.

Third, module completion updates a learner-level record that feeds back into the next scoring cycle, so the routing is not a one-time classification but a closed loop: the re-assessment arrow in Figure 4 returns each learner to Stage 1 after a defined interval or after a module is completed, and a learner can move between profiles as their capability scores change. Fourth, at the cohort level, the same pipeline aggregates profile membership and gap magnitudes into a management view, the share of learners in each profile, and which competence areas carry the largest unresolved differential, which is the concrete decision-support output a training manager acts on when allocating budget across modules. None of these four elements requires functionality beyond what Section II.G already specifies; the contribution here is the routing logic that connects a validated measurement model to an operational rule, not a new algorithm.

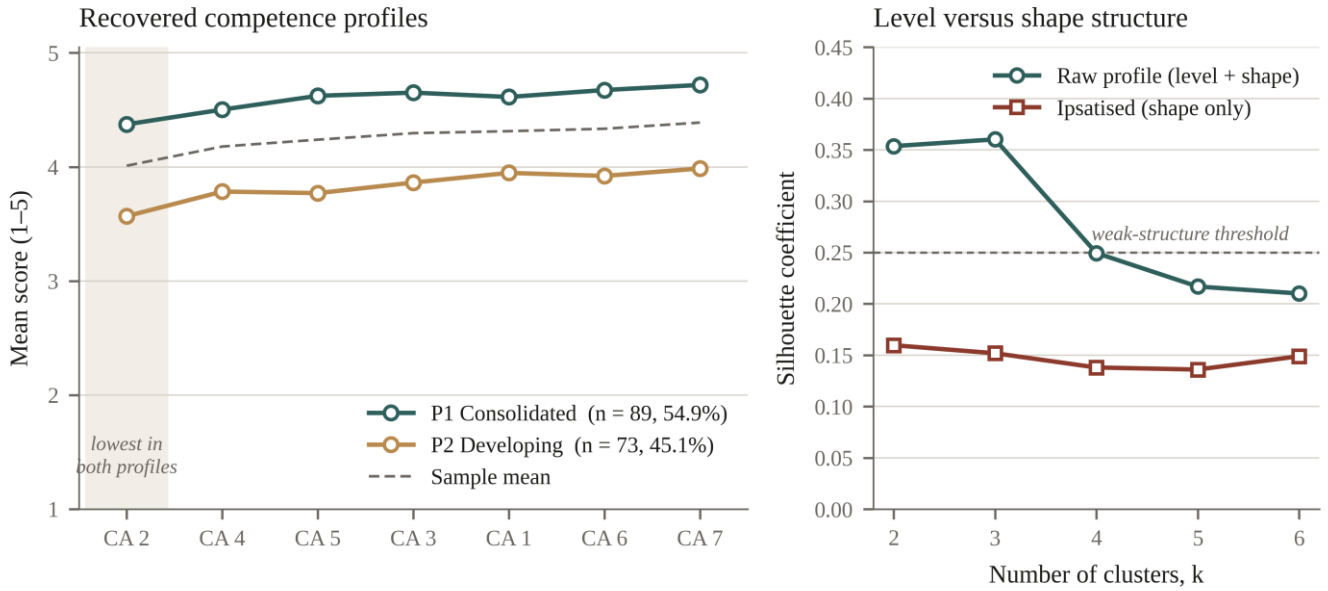


Figure 3. Recovered Competence Profiles (Left) and the Level-Versus-Shape Comparison (Right).

Table 8. Competence Profiles And Derived Routing Specification

Profile	N	%	Composite M	Weakest Area (CA2)	Within-Profile Δ	dz	Routing Entry Point
P1 Consolidated	89	54.9	4.594	4.374	+0.245	0.78	Verification track first; technical modules as a refresher
P2 Developing	73	45.1	3.835	3.568	+0.354	1.07	Full procedural sequence from device operations upward

Note. Profiles from k-means on raw area profiles, k = 2, 200 initializations, fixed seed. Δ = within-profile awareness–capability differential under scheme S2; both differentials significant at $p < .001$.

DOI: 10.34148/teknika.v15i3.1553
CC BY SA

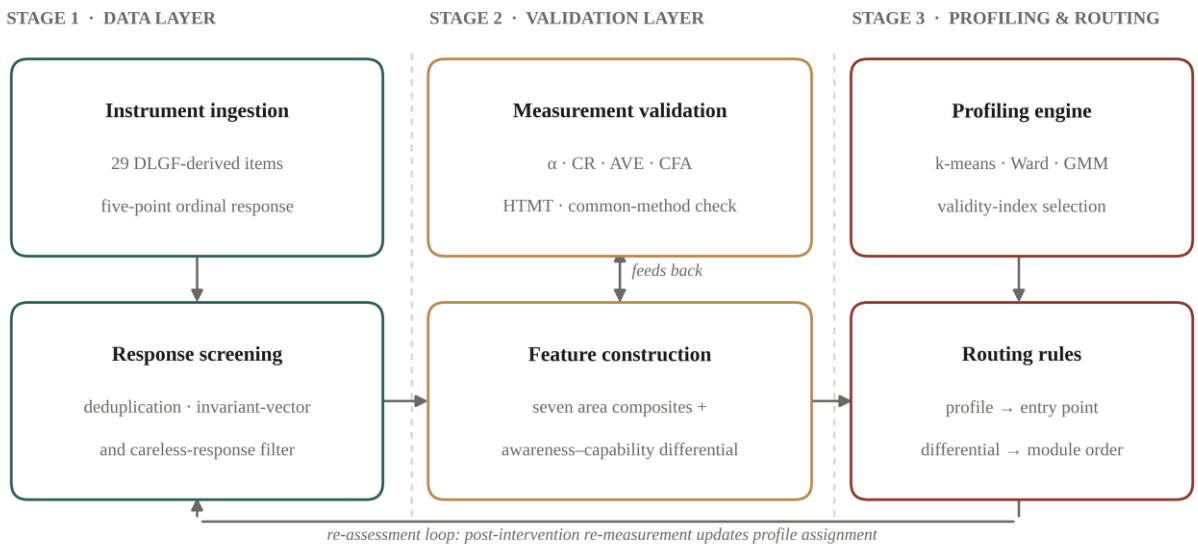


Figure 4. Reference Architecture of the Three-Stage Competency-Profiling Pipeline, From Instrument Ingestion to Routing Rules and Re-Assessment.

IV. CONCLUSION

This study developed a decision-support framework that converts a digital literacy instrument into a routing variable for competency management and adaptive learning systems, and validated it on 162 screened responses from youth in Batam. Three findings answer the research questions. Competence is high in aggregate but unevenly distributed, with ethics highest and data literacy lowest; the seven DLGF areas cannot be collapsed into one factor yet nest under a dominant second-order factor rather than standing as independent constructs. Declarative awareness exceeds procedural capability by 0.294 scale points with a large paired effect, and that differential is stable across three item-coding schemes for six of seven areas. Cluster-based profiling recovers two competence levels, but no shape structure, and the lower level carries the larger differential, making profile membership directly actionable for resource allocation.

The contributions are a documented, reproducible decision-support framework, including a screening protocol that materially changed the analytic sample; empirical evidence that the declarative–procedural decomposition carries more management-relevant information than the framework's own area architecture; and a reference architecture that specifies how cluster-based profiling output becomes routing rules. In practice, the implication is direct: awareness-oriented campaigns have reached their limit in this population, and measurable gains now depend on procedural instruction sequenced by profile, with performance-based assessment, timed source-verification tasks, sandboxed two-factor enrolment, and guided fault diagnosis replacing self-report as the diagnostic input to the management dashboard.

Several limitations bound these conclusions and define the work ahead. The mechanism behind the raw dataset's duplication is inferred from file-preparation practice rather than confirmed from server logs, and although several checks support treating the 162 screened cases as independent observations, this remains an assumption we disclose rather than a fact we have proven. The item taxonomy has been checked for stability across three coding schemes (κ = .867–.937) but awaits independent double-coding by a human rater blind to the authors' scheme; a blinded coding worksheet and kappa-computation script have been prepared for this purpose and are included with the submission materials. The problem-solving area requires additional unambiguous procedural items before any claim about it can be sustained. Reverse-coded items should be introduced to bound acquiescence, and absolute CFA fit should be re-examined in a larger sample using a categorical estimator, which the present sample's polychoric correlation structure did not support. Larger samples would also permit measurement invariance testing and would allow the framework to be evaluated prospectively: whether learners routed by profile outperform learners receiving uniform content is the experiment this design makes possible and does not itself answer. Replication across provinces with contrasting infrastructural and cultural profiles would establish how far the framework transfers.

ACKNOWLEDGMENT

The author thanks the respondents in Batam who participated in the survey and the reviewers whose detailed comments materially improved the manuscript, particularly by prompting the response-screening audit reported in Section II.C.

DATA AVAILABILITY STATEMENT

The screened dataset and the complete analysis scripts, including the screening, measurement, and profiling stages, are openly available. The blinded item-coding worksheet and the accompanying Cohen's kappa computation script

prepared for independent double-coding are included in the same repository. All results reported in this paper can be regenerated from those files with a single command.

REFERENCES

[1] Asosiasi Penyelenggara Jasa Internet Indonesia, "Profil Internet Indonesia 2025," APJII, Jakarta, Indonesia, 2025. [Online]. Available: https://survei.apjii.or.id/

[2] Kementerian Komunikasi dan Informatika and Katadata Insight Center, "Status Literasi Digital di Indonesia 2023," Jakarta, Indonesia, 2024. [Online]. Available: https://data.komdigi.go.id/

[3] X. Ding, W. Deng, and Z. Ye, "Research on the Knowledge and Action Educational Concept from the Perspective of Chinese and Foreign Universities Mottos," *International Theory and Practice in Humanities and Social Sciences*, 2025, [Online]. Available: https://api.semanticscholar.org/CorpusID:278190058

[4] A. G. Shaffer, "Human Performance," *Nature*, vol. 171, pp. 960–962, 2020, [Online]. Available: https://api.semanticscholar.org/CorpusID:742694

[5] N. Law, D. Woo, J. de la Torre, and G. Wong, "A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2," UNESCO Institute for Statistics, Montreal, QC, Canada, 2018.

[6] R. Vuorikari, S. Kluzer, and Y. Punie, "DigComp 2.2: The Digital Competence Framework for Citizens," Publications Office of the European Union, Luxembourg, 2022. doi: 10.2760/115376.

[7] M. C. Martinez-Bravo, C. Sadaba Chalezquer, and J. Serrano-Puche, "Meta-framework of digital literacy: a comparative analysis of 21st-century skills frameworks," *Revista Latina de Comunicacion Social*, no. 79, pp. 76–110, 2021, doi: 10.4185/RLCS-2021-1508.

[8] J. Mattar, C. C. Santos, and L. M. Cuque, "Analysis and comparison of international digital competence frameworks for education," *Educ. Sci. (Basel).*, vol. 12, no. 12, p. 932, 2022, doi: 10.3390/educsci12120932.

[9] Asmayawati, Yufiarti, and E. Yetti, "Pedagogical innovation and curricular adaptation in enhancing digital literacy: a local wisdom approach for sustainable development in Indonesia context," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, p. 100233, 2024, doi: 10.1016/j.joitmc.2024.100233.

[10] M. A. Pratiwi, W. H. R, and B. E. Normande, "Understanding knowledge acquisition, adaptive strategies, challenges, and preservation methods among traditional fishermen in a digital age," *BIO Web Conf.*, vol. 70, p. 05008, 2023, doi: 10.1051/bioconf/20237005008.

[11] S. M. Efastri, R. W. Eriyanti, D. Karmiyati, and Marwa, "Developing contextualized picture storybooks for disaster education: a Riau Malay approach to building resilience in young children," *Indian Journal of Information Sources and Services*, vol. 16, no. 1, pp. 89–100, 2026, doi: 10.51983/ijiss-2026.16.1.10.

[12] F. Yang, R. Yao, Y. Ren, and L. Guo, "Harmony in diversity: digital literacy research in a multidisciplinary landscape," *Comput. Educ.*, vol. 230, p. 105265, 2025, doi: 10.1016/j.compedu.2025.105265.

[13] G. I. Sari, S. Winasis, I. Pratiwi, U. W. Nuryanto, and Basrowi, "Strengthening digital literacy in Indonesia: collaboration, innovation, and sustainability education," *Social Sciences & Humanities Open*, vol. 10, p. 101100, 2024, doi: 10.1016/j.ssaho.2024.101100.

[14] S. Rahmawati, S. Prestridge, A. G. Abdullah, and I. Widiaty, "Unpacking the digital competence challenge in vocational education: A case from Indonesia," *Social Sciences and Humanities Open*, vol. 12, Jan. 2025, doi: 10.1016/j.ssaho.2025.101803.

[15] U. Muawanah, A. Marini, and I. Sarifah, "The interconnection between digital literacy, artificial intelligence, and the use of e-learning applications in enhancing the sustainability of regional languages: evidence from Indonesia," *Social Sciences & Humanities Open*, vol. 10, p. 101169, 2024, doi: 10.1016/j.ssaho.2024.101169.

[16] H. Sama *et al.*, "Evaluating Learning Management System Adoption Effectiveness for Educators and Students in Batam Higher Education," *Edu Komputika*, vol. 12, no. 2, 2025, [Online]. Available: https://journal.unnes.ac.id/journals/edukom/

[17] H. W. Aripradono, "Analisa Extended Technology Acceptance Model Sebagai Penerimaan Penggunaan Platform Startup Education Technology," *Sistemasi: Jurnal Sistem Informasi*, vol. 12, no. 2, pp. 425–437, May 2023, doi: https://doi.org/10.32520/stmsi.v12i2.2732.

[18] L. Badrah Zaki, N. Balqis, I. Palar Hasudungan Tambunan, Q. Anisa, and T. Lady Pratiwi, "Understanding Student Perspectives on ICT Implementation in a Public School in Batam City," *Indonesian Review of English Education, Linguistics, and Literature*, vol. 2, pp. 1–14, Apr. 2024, doi: 10.30762/ireell.v2i1.2719.

[19] M. Wang, C. Zeng, L. Xu, and H. Sun, "Meeting the digital literacy education needs of college students: a two-stage study in China," *The Journal of Academic Librarianship*, vol. 51, p. 103084, 2025, doi: 10.1016/j.acalib.2025.103084.

[20] J. Kruger and D. Dunning, "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments," *J. Pers. Soc. Psychol.*, vol. 77, no. 6, pp. 1121–1134, 1999, doi: 10.1037/0022-3514.77.6.1121.

[21] Y. Zhang, "Impact of digital literacy on college students' English proficiency: the mediating role of learning motivation and the moderating effect of technological self-efficacy," *Acta Psychol. (Amst).*, vol. 256, p. 105452, 2025, doi: 10.1016/j.actpsy.2025.105452.

[22] W. Xiao, M. Wang, and X. Li, "From 'digital novices' to 'digital pioneers': profiles and influencing factors of digital literacy among pre-service special education teachers," *Teach. Teach. Educ.*, vol. 163, p. 105301, 2025, doi: 10.1016/j.tate.2025.105301.

[23] C. Romero and S. Ventura, "Educational data mining and learning analytics: an updated survey," *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1355, 2020, doi: 10.1002/widm.1355.

[24] M. Lan, H. Liu, and Q. Pan, "Unpacking the digital literacy-self-regulated learning nexus: a systematic review and a three-level meta-analysis," *Educ. Res. Rev.*, vol. 48, p. 100713, 2025, doi: 10.1016/j.edurev.2025.100713.

[25] S. Xu, L. Wang, S. Zhu, and Q. Wang, "Implementing the cultivation of students' digital literacy: Exploring effective instructional design based on action research," *Think. Skills Creat.*, vol. 61, p. 102217, Sep. 2026, doi: 10.1016/j.tsc.2026.102217.

[26] S. Pinheiro, A. Pinto, A. C. Simoes, M. Ferraz, A. R. Domingues, and J. Santos, "Digital workplace competencies and their outcomes: a systematic literature review," *Technol. Forecast. Soc. Change*, vol. 222, p. 124832, 2026, doi: 10.1016/j.techfore.2026.124832.

[27] D. Tagare, T. Karki, and W. Yu, "K-12 teachers' ethical competencies for AI literacy: insights from a systematic literature review," *Comput. Educ.*, vol. 236, p. 105435, 2025, doi: 10.1016/j.compedu.2025.105435.

[28] A. Kavak and V. Gultekin, "Examining the relationship between digital competencies and artificial intelligence literacy of academic librarians in Turkiye," *The Journal of Academic Librarianship*, vol. 52, p. 103191, 2026, doi: 10.1016/j.acalib.2025.103191.

[29] A. W. Meade and S. B. Craig, "Identifying careless responses in survey data," *Psychol. Methods*, vol. 17, no. 3, pp. 437–455, 2012, doi: 10.1037/a0028085.

[30] M. R. Ab Hamid, W. Sami, and M. H. Mohmad Sidek, "Discriminant Validity Assessment: Use of Fornell & Larcker criterion versus HTMT Criterion," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Sep. 2017. doi: 10.1088/1742-6596/890/1/012163.

[31] L. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives," *Structural Equation Modeling*, vol. 6, no. 1, pp. 1–55, 1999, doi: 10.1080/10705519909540118.

[32] G. W. Cheung and R. B. Rensvold, "Evaluating goodness-of-fit indexes for testing measurement invariance," *Structural Equation Modeling*, vol. 9, no. 2, pp. 233–255, 2002, doi: 10.1207/S15328007SEM0902_5.

[33] P. M. Podsakoff, S. B. MacKenzie, J.-Y. Lee, and N. P. Podsakoff, "Common method biases in behavioral research: a critical review of the literature and recommended remedies," *Journal of Applied Psychology*, vol. 88, no. 5, pp. 879–903, 2003, doi: 10.1037/0021-9010.88.5.879.

[34] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[35] I. Firman Ashari, E. Dwi Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," 2023. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[36] U. Ahdiani, S. Winiarti, H. Pramono, and T. Ismail, "Collaboration of digital literacy and artificial intelligence in preserving traditional Malay house culture," *E3S Web of Conferences*, vol. 622, p. 03008, 2025, doi: 10.1051/e3sconf/202562203008.

[37] M. Muchtarom, Raharjo, and A. Z. bin Selamat, "Preventing gadget misuse in children to enhance family resilience: integrating religious values, digital literacy, and community participation in rural Indonesia," *Social Sciences & Humanities Open*, vol. 13, p. 102922, 2026, doi: 10.1016/j.ssaho.2026.102922.