Research Square

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

Conceptualising and Measuring Epistemic Dependence in AI-Mediated Learning: Development and Initial Validation of the ED-AIL Scale

Yiran Du

yd392@cam.ac.uk

University of Cambridge https://orcid.org/0000-0002-6576-0073

Bin Zou

Xi'an Jiaotong-Liverpool University

Research Article

Keywords: epistemic dependence, generative artificial intelligence, AI-mediated learning, scale development, higher education

Posted Date: August 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10627520/v1

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

Additional Declarations: The authors declare no competing interests.

Conceptualising and Measuring Epistemic Dependence in AI-Mediated Learning: Development and Initial Validation of the ED-AIL Scale

Abstract

Generative artificial intelligence (GenAI) increasingly mediates the epistemic work of higher education: students use it not only to produce text, but also to locate information, interpret concepts, evaluate claims, obtain feedback, and regulate study. This creates a measurement problem that is not captured directly by existing measures of AI trust, technology acceptance, AI literacy, cognitive offloading, or general reliance. This article conceptualises uncritical epistemic dependence in AI-mediated learning as a context-sensitive tendency to treat AI-generated outputs as epistemically directive, such that learner verification, interpretation, justification, production, or regulation may be displaced. It reports the development and initial validation of the Epistemic Dependence in AI-Mediated Learning Scale (ED-AIL Scale), a risk-oriented instrument designed to distinguish ordinary or productive AI support from epistemically consequential delegation. The scale was developed through construct specification, expert review, cognitive pretesting, and two independent Prolific samples of English-speaking higher education students. Evidence from ordinal exploratory and confirmatory factor analyses, reliability estimates, validation measures, performance-oriented criteria, and measurement-invariance tests supported a provisional five-domain interpretation: knowledge acquisition, conceptual interpretation, epistemic evaluation, knowledge production, and epistemic regulation. Measurement evidence was strongest for acquisition, interpretation, and evaluation; more qualified for production; and weakest for regulation. ED-AIL offers an initial research tool for examining how learners allocate epistemic responsibility between themselves and AI systems, while requiring further validation across institutional, disciplinary, linguistic, platform-specific, and task-specific contexts.

Keywords: epistemic dependence; generative artificial intelligence; AI-mediated learning; scale development; higher education

1. Introduction

GenAI is no longer only an auxiliary writing or productivity tool in higher education. Students now use large language models to search for information, interpret readings, generate explanations, obtain feedback, draft academic text, and decide whether their understanding is sufficient to proceed. In these settings, GenAI becomes part of the epistemic infrastructure of learning: it shapes how students encounter, evaluate, and use knowledge in academic work (Chan & Hu, 2023; Kasneci et al., 2023; Labadze et al., 2023; Xia et al., 2024).

This shift reframes the educational risks of GenAI. The issue is not limited to academic misconduct, technical error, or hallucinated content. It also concerns how learners decide what is true, relevant, sufficient, credible, or well justified. Student-experience studies show that GenAI is valued for brainstorming, summarisation, translation, writing support, feedback, and personalised study assistance (Abbas et al., 2024; Chan & Hu, 2023; Jo, 2024; Shahzad et al., 2024). Those same affordances may, however, encourage learners to accept fluent AI outputs without adequate source checking, conceptual interpretation, evidential evaluation, or metacognitive control.

The present article focuses on this narrower problem: uncritical epistemic dependence in AI-mediated learning. Epistemic dependence is not inherently defective. Learners legitimately depend on teachers, peers, textbooks, academic databases, instruments, curricula, and disciplinary standards. Dependence becomes educationally problematic when an external epistemic resource substitutes for learners' own verification, interpretation, justification, or regulation rather than supporting those processes.

Existing constructs illuminate parts of this problem but do not fully specify it. AI trust, reliance, cognitive offloading, AI literacy, source verification, and self-regulated learning help explain why learners use AI, how they appraise its outputs, and how they regulate learning with external resources. They do not directly identify the learning-specific point at which AI output begins to organise epistemic judgement on the learner's behalf.

This study addresses that measurement gap by developing and initially validating the Epistemic Dependence in AI-Mediated Learning Scale (ED-AIL Scale). ED-AIL is conceptualised as a risk-oriented, context-sensitive tendency to take AI-generated outputs as epistemically directive in ways associated with reduced independent search, interpretation, evaluation, production, or regulation. The scale is not intended to measure all AI use, all productive dependence on epistemic resources, or simple frequency of GenAI engagement. It targets educationally consequential forms of AI-mediated epistemic delegation.

Using expert review, cognitive pretesting, and two independent online Prolific samples of English-speaking higher education students, the study examined construct definition, factor structure, reliability, validity evidence, performance-oriented criteria, and measurement invariance. It addressed four research questions:

(1) How can uncritical epistemic dependence in AI-mediated learning be conceptually defined and operationalised?
(2) What factor structure best represents the risk-oriented ED-AIL Scale?
(3) Does the scale demonstrate reliability and validity evidence among English-speaking higher education learners?
(4) How does ED-AIL relate to adjacent constructs and performance-oriented indicators of AI evaluation?

2. Literature Review and Construct Conceptualisation

2.1 Epistemic dependence as distributed knowing

Epistemic dependence begins from the recognition that knowledge is rarely produced by isolated individuals. Social epistemology has long argued that knowing depends on testimony, expertise, instruments, institutions, and distributed epistemic labour. Hardwig (1985) showed that much of what individuals know depends on trusting the cognitive labour of others, particularly in domains where direct verification is impossible or impractical. Goldman (2001) likewise argued that non-experts must often decide which experts to trust without possessing the expertise needed to verify every claim independently.

In education, this point is crucial. Learners are expected to depend on teachers, peers, textbooks, assessment criteria, databases, and disciplinary conventions. Such dependence is productive when it provides access to reliable knowledge, scaffolds understanding, and strengthens learners' capacity for judgement over time. It becomes problematic when learners treat an external authority as a substitute for understanding, evaluation, or justification. GenAI sharpens this distinction because it can simulate features of epistemic authority while lacking transparent provenance, disciplinary accountability, and stable evidential grounding.

2.2 Epistemic cognition, source evaluation, and GenAI mediation

Educational psychology provides the learning-specific foundation for ED-AIL. Research on epistemic cognition examines how learners understand knowledge and knowing, including beliefs about certainty, complexity, source, and justification (Hofer & Pintrich, 1997; Schommer, 1990). More recent work emphasises epistemic aims, standards, reliable processes, and situated epistemic performance (Barzilai & Chinn, 2018, 2024; Greene et al., 2018). This literature is directly relevant because AI-mediated learning becomes epistemically consequential when AI outputs shape the standards learners use for evidence, explanation, justification, and sufficient understanding.

Source evaluation is especially central. Studies of multiple-document comprehension and online inquiry show that learners often struggle to assess credibility, integrate conflicting evidence, and justify claims across sources (Barzilai & Zohar, 2012; Bråten et al., 2009, 2011; Kammerer et al., 2021; List & Alexander, 2019). GenAI makes this challenge more complex because learners may encounter a fluent synthesis rather than a clearly identifiable source. A chatbot response can combine accurate information,

plausible inference, unsupported synthesis, and fabricated or decontextualised claims in a single coherent answer (Fuchs, 2023; Ji et al., 2023; Kasneci et al., 2023; Labadze et al., 2023). The learner must therefore evaluate not only a source, but an AI-generated epistemic object whose provenance, uncertainty, and warrant may be difficult to inspect.

2.3 The mechanism of ED-AIL: output, uptake, and judgement displacement

The mechanism of ED-AIL can be understood as a relation between AI output, learner uptake, and judgement displacement. GenAI systems provide epistemic outputs in the form of explanations, summaries, recommendations, evaluations, feedback, arguments, and proposed answers. These outputs are not neutral prompts for learning activity. Because they are often fluent, personalised, and apparently coherent, they can acquire a directive role in subsequent learner judgements, particularly when their evidential basis, uncertainty, or disciplinary limitations are not visible.

ED-AIL emerges when learners take up AI output as more than a provisional resource. In such cases, the output may begin to determine what the learner treats as accurate, relevant, sufficient, credible, or conceptually adequate. This uptake can occur at different points in the learning process: when an AI summary becomes the main basis for understanding a topic, when an AI explanation settles the meaning of a difficult concept, when AI feedback determines whether an answer is judged correct, or when AI guidance shapes what the learner studies next. The common feature across these cases is not AI use itself, but the transfer of epistemic control from the learner to the system.

This mechanism clarifies why ED-AIL is compatible with both frequent and infrequent AI use. A student who uses AI regularly to generate possible explanations, then checks them against course readings, empirical evidence, or disciplinary criteria, may be using AI as productive epistemic scaffolding. By contrast, a student who accepts a fluent AI explanation as sufficient without examining whether it is accurate, sourced, contested, or conceptually coherent is displaying the form of uncritical epistemic dependence operationalised here. ED-AIL therefore concerns the conditions under which AI-mediated support becomes a substitute for verification, interpretation, justification, or regulation.

2.4 Differentiating ED-AIL from adjacent constructs

ED-AIL overlaps with several constructs in human-AI interaction, information systems, educational psychology, and the learning sciences, but its focus is more specific. It concerns the learner-AI relation that emerges when AI output begins to guide what the learner treats as true, meaningful, credible, sufficient, or worth pursuing. Existing constructs help explain parts of this relation, but they do not fully capture the educational problem of AI-mediated judgement displacement.

Technology acceptance and AI adoption research explains perceived usefulness, perceived ease of use, behavioural intention, and uptake (Davis, 1989; Venkatesh et al., 2003). These constructs are relevant because learners who perceive AI as useful or easy to use may be more likely to incorporate it into study practices. They do not, however, specify the epistemic role that AI output comes to play once it is used. Similarly, GenAI use frequency describes how often learners engage with AI tools, but frequent use need not imply uncritical dependence, and infrequent use does not preclude strong deference in particular tasks.

Trust research concerns confidence in automated systems under uncertainty (Jian et al., 2000; Lee & See, 2004), while reliance and automation-bias research examines when users follow automated recommendations, overuse automation, or fail to detect system errors (Dzindolet et al., 2003; Goddard et al., 2012; Parasuraman & Riley, 1997; Skitka et al., 1999, 2000). These perspectives help explain why learners may accept AI output, but ED-AIL is narrower in one respect and broader in another: narrower because it concerns learning-specific epistemic delegation, and broader because it includes not only following AI recommendations but also allowing AI output to shape interpretation, evaluation, production, and regulation.

A similar distinction applies to cognitive offloading and AI literacy. Cognitive offloading refers to externalising cognitive operations to tools, artefacts, or environments; in many learning situations, such offloading is efficient or pedagogically useful (Risko & Gilbert, 2016). ED-AIL is concerned with the subset of offloading in which epistemic oversight is reduced. AI literacy concerns understanding AI capabilities, limitations, bias, ethics, and responsible use (Hornberger et al., 2023; Laupichler et al., 2022, 2023; Ng et al., 2024). Such understanding may reduce uncritical uptake, but it is not simply the inverse of ED-AIL: learners may know that AI has limitations while still deferring to its outputs in practice.

The closest educational-psychology neighbours are source verification and self-regulated learning. Source verification and multiple-document evaluation concern how learners check claims against original, independent, or credible sources (Bråten et al., 2009, 2011; Kammerer et al., 2021; List & Alexander, 2019). Self-regulated learning concerns planning, monitoring, controlling, and evaluating learning more broadly (Greene & Azevedo, 2007; Panadero, 2017). ED-AIL intersects with both, but it is not reducible to either. It also includes AI-mediated shifts in conceptual interpretation, evaluative judgement, academic production, and regulation of inquiry. Table 1 summarises these boundaries.

Table 1

Conceptual boundaries between ED-AIL and adjacent constructs

Adjacent construct	Primary focus	How it differs from ED-AIL
Technology acceptance and AI adoption	Perceived usefulness, ease of use, intention to use, and uptake of AI tools.	Explains whether learners are likely to use AI; ED-AIL concerns the epistemic role AI output takes once it is used.
GenAI use frequency	Regularity or intensity of AI tool use in learning.	Frequent use can involve careful corroboration; infrequent use can still involve deference in particular tasks.
AI trust	Confidence in AI reliability, competence, or usefulness under uncertainty.	Trust concerns appraisal of the system; ED-AIL concerns how learners allocate judgement after receiving output.
AI reliance	Using or following AI recommendations, feedback, explanations, or answers.	Reliance becomes epistemically consequential when followed output substitutes for learner verification, interpretation, or justification.
Automation bias	Over-following automated advice or failing to detect automation errors.	Usually studied as decision error; ED-AIL concerns learning-specific judgement, source use, and justification.
Cognitive offloading	Externalising cognitive work to tools, artefacts, or environments.	Offloading may support learning; ED-AIL concerns cases where epistemic oversight is reduced.
AI literacy	Understanding AI capabilities, limitations, bias, ethics, and responsible use.	Knowledge of AI limitations may help, but it does not guarantee careful uptake in practice.
Source verification	Checking claims against original, independent, or credible sources.	A close correlate, but ED-AIL also includes interpretation, evaluation, production, and regulation.
Self-regulated learning	Planning, monitoring, controlling, and evaluating learning.	Broader than ED-AIL; the present construct concerns AI-specific shifts in regulatory control.

Note. ED-AIL = Epistemic Dependence in AI-Mediated Learning. The table clarifies how ED-AIL relates to, but differs from, adjacent constructs.

This boundary work shaped the validation design. ED-AIL was expected to relate to AI reliance, AI trust, cognitive offloading, unverified acceptance of AI-generated outputs, source verification, self-regulated learning, and performance-oriented indicators of AI evaluation, while retaining distinctive variance as a learning-specific construct. Because the validation battery did not include all possible rival constructs, including critical-thinking disposition, epistemic vigilance, need for cognition, academic self-efficacy, and domain knowledge, the discriminant-validity argument should be treated as supportive but not exhaustive.

2.5 Definition and proposed dimensions

Following these distinctions, uncritical epistemic dependence in AI-mediated learning is defined as the extent to which learners take AI systems as epistemic mediators when acquiring, interpreting, evaluating, producing, or regulating knowledge in ways that reduce independent verification, interpretation, justification, or epistemic agency. ED-AIL is conceptualised as a context-sensitive dispositional

tendency rather than a fixed personality trait, a platform-specific usage metric, or a direct behavioural diagnosis. Scores are expected to reflect how learners typically orient towards AI-mediated epistemic work in learning contexts, while still varying by task, discipline, platform affordance, and assessment demand.

The ED-AIL Scale operationalises manifestations of AI-mediated delegation that are most relevant to educational risk. The five domains represent learning functions in which AI may operate either as a scaffold for judgement or as a substitute for it. Table 2 presents the conceptual blueprint for item generation and subsequent empirical testing.

Table 2

Conceptual domains of the risk-oriented ED-AIL construct

Domain	Definition	Illustrative learning functions
Knowledge acquisition dependence	AI output substitutes for learners' independent search or corroboration of information.	Initial information search, summaries, starting materials
Conceptual interpretation dependence	AI explanations substitute for learners' own interpretation of concepts, readings, or theoretical claims.	Concept explanation, reading interpretation, theory clarification
Epistemic evaluation dependence	AI feedback substitutes for learners' judgement of correctness, credibility, quality, or sufficiency.	Answer checking, argument evaluation, source judgement
Knowledge production dependence	AI-generated text, structure, or argumentation substitutes for learners' own formulation of academic work.	Drafting, argument development, essay planning
Epistemic regulation dependence	AI guidance substitutes for learners' control over inquiry, study decisions, and judgements of understanding.	Choosing what to study, deciding when understanding is sufficient

Note. The domains were used as the theoretical blueprint for item generation and were then tested empirically. Each domain captures a learning function in which AI output may move from scaffold to substitute for learner judgement.

3. Method

3.1 Research design

The study used a multi-stage scale-development and initial-validation design guided by a validity-argument logic. The design drew on guidance for construct definition, item generation, response-process evidence, exploratory and confirmatory factor analysis, and score interpretation (American Educational Research Association et al., 2014; Boateng et al., 2018; Clark & Watson, 1995; Flora & Flake, 2017; Worthington & Whittaker, 2006). Phase 1 specified the construct and generated an initial item pool. Phase 2 used expert review and cognitive pretesting to refine item content. Phase 3 used two independent Prolific samples: Sample 1 for item screening and ordinal exploratory factor analysis, and Sample 2 for confirmatory factor analysis, reliability estimation, validity evidence, performance-oriented criteria, and measurement invariance.

3.2 Participants, procedure, and data quality

Participants were recruited through Prolific using prescreening criteria for current higher education enrolment, age 18 years or above, English fluency, and prior GenAI use for learning within the previous six months. Participants completed an online survey after providing informed consent and indicated the type of institution attended and the learning tasks for which they had used GenAI.

The two samples were collected in separate batches, and duplicate Prolific IDs across or within batches were excluded before analysis. Exclusion rules were applied before substantive analyses: failed eligibility checks, failed attention checks, completion below one-third of the median completion time, duplicate IDs, invariant response patterns, and excessive missingness. Item-level missingness was below 1.3% for all retained ED-AIL items and was handled with the estimator-specific missing-data procedures available for ordinal models.

For Sample 1, 849 responses were collected and 93 were excluded, leaving 756 valid cases. For Sample 2, 883 responses were collected and 94 were excluded, leaving 789 valid cases. The final validation dataset therefore consisted of 1,545 higher education students. Table 3 reports the demographic and study-related characteristics of the two samples.

Table 3

Sample characteristics

Characteristic	Sample 1 (*n* = 756)	Sample 2 (*n* = 789)
Age, *M* (*SD*)	24.9 (6.1)	25.8 (6.7)
Women	410 (54.2%)	428 (54.2%)
Men	322 (42.6%)	337 (42.7%)
Non-binary, another identity, or prefer not to say	24 (3.2%)	24 (3.0%)
Undergraduate students	487 (64.4%)	492 (62.4%)
Postgraduate taught students	198 (26.2%)	214 (27.1%)
Doctoral students	71 (9.4%)	83 (10.5%)
UK-based institution	305 (40.3%)	299 (37.9%)
US/Canada-based institution	251 (33.2%)	278 (35.2%)
Other European institution	121 (16.0%)	126 (16.0%)
Other regions	79 (10.4%)	86 (10.9%)
Used GenAI at least weekly for learning	514 (68.0%)	562 (71.2%)
Used GenAI daily or almost daily for learning	196 (25.9%)	232 (29.4%)

Note. Percentages may not sum to 100 because of rounding. GenAI = generative artificial intelligence.

3.3 Item development and response-process evidence

Item generation followed a deductive-inductive procedure. Deductively, the item pool drew on social epistemology, epistemic cognition, source evaluation, AI-mediated learning, human-AI trust and reliance, cognitive offloading, self-regulated learning, and AI literacy. Inductively, wording was informed by empirical studies of how students use GenAI for writing, feedback, explanation, study support, and assessment preparation (Abbas et al., 2024; Chan & Hu, 2023; Jo, 2024; Shahzad et al., 2024; Xia et al., 2024).

The initial pool contained 61 items rated from 1 = strongly disagree to 5 = strongly agree. Item wording was revised to reduce construct transparency by avoiding repeated dependence language. Instead of repeatedly asking whether students "depend on AI", items were written to capture epistemic practices such as treating AI explanations as sufficient, using AI feedback to determine adequacy, reducing source checking after AI output, or letting AI structure inquiry. Nine experts in educational technology, AI in education, epistemic cognition, educational measurement, and higher education pedagogy rated items for relevance, clarity, representativeness, and redundancy. Expert review was treated as qualitative content evidence rather than a formal item-content-validity-index procedure. Twenty students then completed cognitive interviews to examine item interpretation, ambiguity, and response reasoning. The item-development process and item-level deletion decisions are reported with the Results tables to preserve the audit trail for item retention.

3.4 Validation measures and performance-oriented criteria

The validation battery was designed to distinguish ED-AIL from adjacent constructs and to test theoretically expected relations with related learning and AI-use variables. Self-report measures assessed AI reliance, AI trust, cognitive offloading, AI literacy, source verification, self-regulated learning, and unverified acceptance of AI-generated outputs. These variables represented plausible rival explanations for ED-AIL scores: general willingness to rely on AI, confidence in AI, externalisation of cognition, technical AI understanding, ordinary verification habits, and general self-regulatory competence. The battery did not include all possible rivals, including critical-thinking disposition,

epistemic vigilance, need for cognition, academic self-efficacy, and domain knowledge; this limitation is taken up in the discussion. All self-report items used a 1-5 agreement scale.

Two performance-oriented criteria were also included to reduce exclusive reliance on same-method self-report evidence. In the AI error-detection task, participants evaluated eight brief AI-generated explanations on general academic topics, four of which contained unsupported or inaccurate claims. Accuracy was scored as the proportion of correct judgements. Because the task was brief, it should be interpreted as preliminary behavioural evidence rather than a definitive criterion measure.

In the source-corroboration task, participants read six AI-generated claims and indicated whether and how they would verify each claim before using it in academic work. Responses were scored according to the appropriateness of the corroboration strategy, including consulting original sources, comparing independent evidence, or identifying unsupported claims as requiring verification. Higher scores reflected corroboration through original sources, peer-reviewed evidence, or independent credible materials rather than acceptance of the AI output as sufficient. Table 4 summarises the validation measures and performance-oriented criteria used in Sample 2.

Table 4

Validation measures and criteria used in Sample 2

Construct or criterion	Source or basis	Format	Reliability
AI reliance	Human-AI reliance and automation research; informed by Dzindolet et al. (2003) and Lee and See (2004)	5 items	.85
AI trust	Trust in automation literature; informed by Jian et al. (2000) and Lee and See (2004)	6 items	.87
Cognitive offloading	Learning-context adaptation informed by Risko and Gilbert (2016)	5 items	.83
AI literacy	AI literacy frameworks and instruments; informed by Hornberger et al. (2023) and Ng et al. (2024)	8 items	.81
Source verification	Source-evaluation and epistemic cognition research; informed by Bråten et al. (2009) and Kammerer et al. (2021)	5 items	.82
Self-regulated learning	Planning, monitoring, and evaluation in learning; informed by Greene and Azevedo (2007) and Panadero (2017)	6 items	.84
Unverified AI-output acceptance	Proximal criterion developed for this study	4 items	.86
AI error-detection task	Performance-oriented criterion	8 scenarios	KR-20 = .70
Source-corroboration task	Performance-oriented criterion	6 scenarios	.73

Note. Reliability coefficients are omega or alpha for self-report scales, KR-20 for dichotomous task accuracy, and internal consistency for scored scenario responses.

3.5 Analytic strategy

Analyses were conducted in R. Because ED-AIL items used ordered response categories, exploratory factor analysis used polychoric correlations and oblique rotation. Sampling adequacy was evaluated with the Kaiser-Meyer-Olkin statistic and Bartlett's test of sphericity. The number of factors was determined using parallel analysis, minimum average partial (MAP), scree inspection, interpretability, and theoretical coherence. Items were retained when they had primary loadings of approximately .50 or higher, cross-loadings generally below .30, adequate communalities, and clear construct fit. Borderline items were retained only when they preserved theoretically important domain coverage and did not substantially degrade model fit.

Confirmatory factor analysis used a robust ordinal estimator with thresholds estimated for all ordered indicators and residual covariances fixed to zero. Model fit was evaluated using chi-square, CFI, TLI, RMSEA with 90% confidence intervals, and SRMR (Hu & Bentler, 1999; Schreiber et al., 2006). Degrees of freedom were reported from the fitted ordinal CFA models rather than from the baseline item-correlation matrix alone. Reliability was estimated using Cronbach's alpha and McDonald's omega. Additional measurement evidence included average variance extracted, composite reliability,

heterotrait-monotrait ratios, latent factor correlations, and bifactor diagnostics. Incremental validity was examined with hierarchical regression models predicting a proximal criterion and two performance-oriented criteria. Measurement invariance tested configural, loading, and loading-plus-threshold invariance across women and men, undergraduate and postgraduate students, and daily versus less-than-daily GenAI users. Invariance was interpreted using changes in CFI, RMSEA, and SRMR rather than chi-square difference tests alone (Chen, 2007).

To make the validation argument auditable within the main article, item-level and model-level evidence are reported in the Results alongside the main analyses. This includes item-development and deletion evidence, factor-retention evidence, EFA and CFA item-level evidence, latent factor correlations, reliability, validation relations, incremental validity, measurement invariance, and scoring guidance.

4. Results

In a scale-development article, the Results need to establish both the substantive measurement claim and the technical audit trail behind that claim. The analyses are therefore organised as a cumulative validity argument. First, the item pool was reduced and the final item content was specified. Second, factor-retention evidence and the exploratory structure were evaluated in Sample 1. Third, confirmatory, reliability, validity, incremental-validity, and invariance evidence were examined in Sample 2. This order allows the reader to move from construct representation to score interpretation.

4.1 Scale construction, retained items, and scoring

The first stage examined whether the item pool represented the theoretical blueprint while removing weak, redundant, or ambiguous indicators. Expert review and cognitive pretesting refined item wording and response processes before empirical screening reduced the pool for EFA. Tables 5 and 6 document this reduction, while Table 7 presents the retained instrument. The final scale comprised 23 items: five for knowledge acquisition, five for conceptual interpretation, five for epistemic evaluation, four for knowledge production, and four for epistemic regulation. Items are scored so that higher values indicate greater uncritical epistemic dependence.

Table 5

Item development and retention process

Stage	Items remaining	Decision basis
Initial theoretical pool	61	Items generated across five proposed domains.
Expert review	45	Sixteen items removed after expert review for weak relevance, redundancy, excessive overlap with AI trust, or lack of epistemic content.
Cognitive pretesting	39	Six items removed after cognitive pretesting when students interpreted them as general AI usefulness, ordinary study support, or platform preference.
Preliminary item screening	31	Eight items removed for low variance, weak item-total correlations, local redundancy, or poor domain representation.
Final ED-AIL Scale after EFA/CFA review	23	Eight items removed during final EFA/CFA review because of cross-loadings, unstable domain placement, local redundancy, or weak conceptual interpretability.

Table 6

Item-level deletion audit summary

Removed item	Stage removed	Decision note
EXP1	Expert review	General statement about AI being useful; no epistemic delegation.
EXP2	Expert review	Enjoyment/comfort with AI rather than dependence.
EXP3	Expert review	Perceived efficiency rather than judgement displacement.
EXP4	Expert review	Broad preference for AI tools; weak construct relevance.
TRU1	Expert review	Overlapped with trust in AI competence.

Removed item	Stage removed	Decision note
TRU2	Expert review	Measured confidence in system accuracy rather than learner uptake.
TRU3	Expert review	Measured perceived dependability of AI tools.
TRU4	Expert review	Too similar to validation measure for AI trust.
USE1	Expert review	Frequency of use without epistemic content.
USE2	Expert review	Platform-use intensity rather than dependence.
USE3	Expert review	Generic ChatGPT use item.
USE4	Expert review	Measured convenience rather than delegation.
KAD6	Cognitive pretesting	Interpreted as efficient search rather than reduced verification.
KAD7	Cognitive pretesting	Ambiguous between starting search and replacing search.
KAD8	Cognitive pretesting	Students interpreted item as ordinary note-taking.
CID6	Cognitive pretesting	Wording confused explanation seeking with acceptance of interpretation.
CID7	Cognitive pretesting	Too close to confidence after feedback.
CID8	Cognitive pretesting	Participants gave inconsistent interpretations across domains.
EED6	Preliminary screening	Restricted variance; most respondents disagreed.
EED7	Preliminary screening	Local redundancy with EED5.
KPD5	Preliminary screening	Measured writing assistance rather than knowledge production dependence.
KPD6	Preliminary screening	Low item-total correlation.
KPD7	Preliminary screening	Conceptual overlap with academic-integrity concern.
ERD5	Preliminary screening	Unstable relation with proposed domain.
ERD6	Preliminary screening	Low variance and weak domain representation.
SCR1	Expert review	Reverse-worded source-checking item judged likely to produce response confusion.
SCR2	Expert review	Reverse-worded source-checking item judged weakly aligned with the focal construct.
SCR3	Expert review	Potential method effect rather than construct signal.
SCR4	Expert review	Ambiguous reference to source checking.
SCR5	EFA/CFA review	Removed during final model review because the content was redundant with validation measures for source verification.
SCR6	Preliminary screening	Low communality in preliminary factor model.
MIX1	EFA/CFA review	Cross-loading between knowledge acquisition and conceptual interpretation.
MIX2	EFA/CFA review	Cross-loading between epistemic evaluation and regulation.
MIX3	EFA/CFA review	Weak primary loading.
MIX4	EFA/CFA review	Loaded on unexpected factor with poor theoretical fit.
MIX5	EFA/CFA review	High residual association with another item.
MIX6	EFA/CFA review	Domain placement unstable across extraction methods.
MIX7	EFA/CFA review	Retained model improved after removal and content was redundant.

Table 7

Final ED-AIL items

Code	Domain	Item wording
KAD1	Knowledge acquisition	After reading an AI overview, I often stop looking for sources that could challenge or extend it.
KAD2	Knowledge acquisition	AI-provided information often becomes the basis for my work before I consult course materials.
KAD3	Knowledge acquisition	When AI summarises a reading, I am less likely to examine the original text in detail.
KAD4	Knowledge acquisition	AI explanations often become my main reference point for a new topic.
KAD5	Knowledge acquisition	I sometimes proceed with a learning task using AI-provided information that I have not independently corroborated.
CID1	Conceptual interpretation	When AI explains a difficult concept clearly, I tend to accept that interpretation as the one to use.
CID2	Conceptual interpretation	If my interpretation differs from AI's explanation, I usually revise my understanding towards the AI version.
CID3	Conceptual interpretation	AI explanations often settle how I understand a theory or idea.
CID4	Conceptual interpretation	I feel that I understand a concept sufficiently once AI has explained it clearly.
CID5	Conceptual interpretation	When reading complex material, I use AI explanations to decide what the text mainly means.

Code	Domain	Item wording
EED1	Epistemic evaluation	I use AI feedback to decide whether my answer is correct.
EED2	Epistemic evaluation	If AI says an argument is weak, I usually treat it as weak before checking it myself.
EED3	Epistemic evaluation	AI feedback often determines whether I think a source or claim is credible.
EED4	Epistemic evaluation	AI feedback often becomes my main basis for judging the quality of my work.
EED5	Epistemic evaluation	When AI evaluates my work, I often accept its judgement without seeking other evidence.
KPD1	Knowledge production	AI-generated outlines often determine the structure of my academic work.
KPD2	Knowledge production	I incorporate AI-generated ideas into my work before fully evaluating them myself.
KPD3	Knowledge production	AI-generated wording often shapes how I express my own academic argument.
KPD4	Knowledge production	When I am unsure what to write, AI output often becomes the starting point and direction for my response.
ERD1	Epistemic regulation	AI suggestions often determine what I study next.
ERD2	Epistemic regulation	I use AI to decide whether I have understood enough to move on.
ERD3	Epistemic regulation	When AI gives a learning plan, I tend to follow it without many independent changes.
ERD4	Epistemic regulation	AI guidance often sets the questions I ask next in a learning task.

Note. Domain abbreviations: KAD = knowledge acquisition dependence; CID = conceptual interpretation dependence; EED = epistemic evaluation dependence; KPD = knowledge production dependence; ERD = epistemic regulation dependence. Items are rated from 1 = strongly disagree to 5 = strongly agree. Subscale scores are calculated as the mean of items within each domain. The total ED-AIL score is the mean of all 23 retained items. Higher scores indicate greater uncritical epistemic dependence on AI during learning as operationalised by the scale.

4.2 Exploratory factor-retention and structure

Before extraction, the screening set showed adequate distributional variability and strong factorability. Items were removed when they combined weak empirical performance with poor construct fit, rather than on statistical grounds alone. In Sample 1, the 31-item screening set showed excellent sampling adequacy, KMO = .94, and Bartlett's test of sphericity was significant, chi-square(465) = 13,184.27, $p < .001$.

The factor-retention evidence supported, but did not conclusively settle, a five-factor solution. Parallel analysis, minimum average partial (MAP) values, scree inspection, and interpretability converged on five factors, although the fifth factor was empirically fragile. As shown in Table 8, the fifth factor only marginally exceeded the parallel-analysis threshold, and the MAP improvement from four to five factors was small. Retention of the five-factor solution therefore rested on theory and interpretability as well as numerical criteria. The regulation domain should consequently be regarded as provisional.

Table 8

Numerical factor-retention evidence in Sample 1

Factor	Observed eigenvalue	Parallel-analysis 95th percentile	Decision
1	8.62	1.45	Retain
2	2.54	1.36	Retain
3	1.82	1.29	Retain
4	1.35	1.22	Retain
5	1.18	1.17	Retain; borderline but interpretable
6	0.98	1.13	Do not retain
MAP comparison	Four factors = .028; five factors = .027; six factors = .028	Lowest at five factors, but only marginally	Retain five factors with caution

Note. Parallel analysis was based on polychoric correlations. MAP = minimum average partial.

The retained EFA solution preserved all five theoretically specified learning functions while reducing the item pool to 23 indicators. Table 9 reports the EFA summary, and Table 10 condenses the pattern-loading evidence. The resulting structure was most coherent for the acquisition, interpretation, and evaluation domains; production and regulation remained interpretable but more qualified.

Table 9

EFA summary for the final ED-AIL structure in Sample 1

Factor	Items	Primary loading range	Largest cross-loading	Communality range	Common variance (%)
Knowledge acquisition dependence	5	.55-.74	.29	.37-.56	14.4
Conceptual interpretation dependence	5	.57-.78	.28	.40-.61	13.2
Epistemic evaluation dependence	5	.62-.80	.27	.46-.66	14.6
Knowledge production dependence	4	.53-.73	.30	.36-.55	10.9
Epistemic regulation dependence	4	.48-.68	.30	.30-.48	9.0

Note. EFA = exploratory factor analysis. Total common variance explained = 62.1%.

Table 10

Condensed EFA pattern-loading summary

Item	KAD	CID	EED	KPD	ERD	h^2
KAD1	.61					.40
KAD2	.74					.56
KAD3	.59					.38
KAD4	.55					.37
KAD5	.72					.53
CID1		.75				.57
CID2		.78				.61
CID3		.60				.42
CID4		.64				.45
CID5		.57				.40
EED1			.80			.66
EED2			.74			.59
EED3			.68			.50
EED4			.62			.46
EED5			.77			.63
KPD1				.68		.50
KPD2				.63		.44
KPD3				.53		.36
KPD4				.73		.55
ERD1					.68	.48
ERD2					.60	.39
ERD3					.56	.35
ERD4					.48	.30

Note. Only primary loadings are shown in this condensed table. Most retained cross-loadings were below .30; the largest was .30 for knowledge production and epistemic regulation items.

4.3 Confirmatory structure and item-level evidence

CFA tested whether the exploratory structure generalised to an independent sample. In Sample 2, the analysis compared a one-factor model, a four-factor model, a correlated five-factor model, a higher-

order model, and a bifactor sensitivity model. As shown in Table 11, the one-factor model fit poorly, and the four-factor model remained inadequate despite improvement. The correlated five-factor model showed acceptable, though not strong, fit and was retained as the primary model because it preserved theoretically important distinctions among learning functions. The bifactor model showed slightly better SRMR, but omega hierarchical (.64), explained common variance (.52), and percentage of uncontaminated correlations (.79) did not justify replacing domain-level interpretation with a strictly unidimensional interpretation.

Table 11

CFA model comparisons in Sample 2

Model	chi-square	df	CFI	TLI	RMSEA [90% CI]	SRMR	Interpretation
One-factor	2508.34	230	.723	.695	.112 [.108, .116]	.103	Poor fit
Four-factor	1186.52	224	.883	.868	.074 [.070, .078]	.066	Improved but still inadequate
Correlated five-factor	678.42	220	.944	.936	.051 [.047, .055]	.049	Retained primary model
Higher-order	742.90	225	.937	.929	.054 [.050, .058]	.052	Acceptable but less diagnostic for domains
Bifactor sensitivity	631.76	207	.948	.937	.051 [.047, .056]	.044	Slightly better SRMR; omegaH = .64, ECV = .52, PUC = .79

Note. ECV = explained common variance. omegaH = omega hierarchical. PUC = percentage of uncontaminated correlations. Residual covariances were fixed to zero in the reported models. Model degrees of freedom refer to the fitted ordinal CFA specifications rather than to the 253 item-pair degrees of freedom of the baseline correlation matrix.

Item-level CFA results clarified where the measurement model was strongest and weakest. Standardised loadings for the retained model ranged from .49 to .79, with all confidence intervals excluding zero. As shown in Table 12, the weakest loading occurred in the epistemic regulation factor, reinforcing the need to treat that subscale as the least stable domain. Table 13 shows that latent factor correlations were substantial but not redundant, supporting related domain-level interpretations within a broader ED-AIL construct.

Table 12

Standardised CFA loadings for the retained five-factor model in Sample 2

Item	Factor	Std. loading	95% CI
KAD1	Knowledge acquisition	.63	[.56, .70]
KAD2	Knowledge acquisition	.70	[.64, .76]
KAD3	Knowledge acquisition	.60	[.53, .67]
KAD4	Knowledge acquisition	.55	[.48, .62]
KAD5	Knowledge acquisition	.72	[.66, .78]
CID1	Conceptual interpretation	.72	[.66, .78]
CID2	Conceptual interpretation	.76	[.70, .82]
CID3	Conceptual interpretation	.61	[.54, .68]
CID4	Conceptual interpretation	.66	[.59, .72]
CID5	Conceptual interpretation	.58	[.51, .65]
EED1	Epistemic evaluation	.79	[.74, .84]
EED2	Epistemic evaluation	.74	[.68, .80]
EED3	Epistemic evaluation	.66	[.59, .73]
EED4	Epistemic evaluation	.61	[.54, .68]
EED5	Epistemic evaluation	.77	[.71, .82]
KPD1	Knowledge production	.67	[.60, .74]
KPD2	Knowledge production	.63	[.56, .70]
KPD3	Knowledge production	.52	[.44, .60]

Item	Factor	Std. loading	95% CI
KPD4	Knowledge production	.71	[.64, .77]
ERD1	Epistemic regulation	.63	[.56, .70]
ERD2	Epistemic regulation	.59	[.52, .66]
ERD3	Epistemic regulation	.55	[.48, .62]
ERD4	Epistemic regulation	.49	[.41, .57]

Note. Std. loading = standardised loading. Item wording appears in Table 7.

Table 13

Latent factor correlations in Sample 2

Factor	KAD	CID	EED	KPD	ERD
KAD	1.00				
CID	.59	1.00			
EED	.54	.70	1.00		
KPD	.48	.45	.62	1.00	
ERD	.51	.57	.64	.53	1.00

Note. KAD = knowledge acquisition dependence; CID = conceptual interpretation dependence; EED = epistemic evaluation dependence; KPD = knowledge production dependence; ERD = epistemic regulation dependence.

4.4 Reliability and measurement-level evidence

Reliability and measurement-level indices supported use of the total score more strongly than unrestricted domain interpretation. Internal consistency was acceptable across subscales, but weaker for epistemic regulation than for the other domains. Average variance extracted exceeded .50 only for epistemic evaluation and was below .50 for the remaining factors, indicating that convergent measurement evidence remains incomplete. Composite reliability ranged from .65 to .84. As reported in Table 14, the total score showed stronger reliability than any single subscale and should be treated as the most stable score at this stage of validation.

Table 14

Descriptive statistics, reliability, and measurement-level validity evidence in Sample 2

Scale	Items	*M*	*SD*	alpha	omega	AVE	CR
Knowledge acquisition dependence	5	3.25	0.88	.79	.80	.41	.78
Conceptual interpretation dependence	5	3.47	0.83	.84	.85	.45	.80
Epistemic evaluation dependence	5	3.07	0.90	.87	.88	.51	.84
Knowledge production dependence	4	3.15	0.89	.76	.78	.41	.73
Epistemic regulation dependence	4	2.93	0.85	.70	.72	.32	.65
Total ED-AIL score	23	3.19	0.67	.90	.91	-	-

Note. alpha = Cronbach's alpha; omega = McDonald's omega; AVE = average variance extracted; CR = composite reliability calculated from the standardised CFA loadings in Table 12. The total ED-AIL score is the item-weighted mean of all 23 retained items. Response scale ranged from 1 to 5.

4.5 Relations with validation variables

Relations with validation variables were consistent with the construct definition. As reported in Table 15, the ED-AIL total score correlated moderately with AI reliance, AI trust, cognitive offloading, and unverified AI-output acceptance, but only modestly with GenAI use frequency. Its small positive association with AI literacy indicates that the scale is not simply an inverse measure of AI understanding. Negative associations with source verification, self-regulated learning, AI error-detection accuracy, and source-corroboration performance were theoretically expected, although the performance-task associations were small and should be interpreted as preliminary criterion evidence.

Table 15

Correlations between ED-AIL total score and validation variables in Sample 2

Validation variable	*r*	*p*	Validity role
AI reliance	.48	< .001	Convergent relation
AI trust	.37	< .001	Related but distinct
Cognitive offloading	.34	< .001	Related but broader construct
GenAI use frequency	.28	< .001	ED-AIL is not merely usage frequency
AI literacy	.08	.026	Small relation; not inverse AI literacy
Source verification	-.27	< .001	Higher ED-AIL associated with less checking
Self-regulated learning	-.14	< .001	Small negative relation
Unverified AI-output acceptance	.46	< .001	Proximal criterion
AI error-detection accuracy	-.14	< .001	Performance-oriented criterion
Source-corroboration performance	-.12	.001	Performance-oriented criterion

Note. Valid n for correlations ranged from 776 to 789 because of item-level missingness on validation measures. Positive scores on AI error-detection accuracy and source-corroboration performance indicate better task performance; therefore, negative correlations indicate lower performance among learners with higher ED-AIL scores.

4.6 Incremental validity

Incremental models tested whether ED-AIL added explanatory value beyond plausible rival explanations. Hierarchical regressions controlled for demographics, GenAI use frequency, AI trust, AI literacy, source verification, and self-regulated learning. Table 16 shows that ED-AIL explained unique variance in unverified AI-output acceptance and a small increment in AI error-detection accuracy. Its incremental association with source-corroboration performance was in the expected negative direction but was not statistically significant after controls. These findings indicate that ED-AIL is not reducible to AI trust, use frequency, literacy, or general self-regulatory tendencies, while also showing that behavioural validation remains limited.

Table 16

Incremental validity regressions predicting proximal and performance-oriented criteria

Outcome	Controls R^2	Final R^2	Delta R^2 for ED-AIL	ED-AIL beta [95% CI]	*p*
Unverified AI-output acceptance	.39	.44	.05	.24 [.16, .32]	< .001
AI error-detection accuracy	.15	.17	.02	-.13 [-.22, -.04]	.004
Source-corroboration performance	.18	.19	.01	-.07 [-.15, .01]	.087

Note. Controls included age, gender, degree level, region, GenAI use frequency, AI trust, AI literacy, source verification, and self-regulated learning. Beta coefficients are standardised. Positive scores on the two task outcomes indicate better performance.

4.7 Measurement invariance and score interpretation

Measurement invariance analyses evaluated whether the retained five-factor model could be interpreted similarly across key learner groups. The evidence was adequate for configural and loading invariance across women and men, undergraduate and postgraduate students, and daily versus less-than-daily GenAI users, but threshold invariance was less stable. As shown in Table 17, the loading-plus-threshold model for daily versus less-than-daily GenAI use exceeded some commonly used change-index thresholds. Latent mean comparisons by use frequency should therefore be treated as exploratory unless future studies support stronger threshold invariance.

Table 17

Measurement invariance tests for the correlated five-factor model

Comparison	Model	Group *n*	CFI	RMSEA	SRMR	Delta CFI	Delta RMSEA	Delta SRMR
Women vs men	Configural	428 / 337	.948	.051	.054	-	-	-
Women vs men	Loadings	428 / 337	.944	.053	.061	-.004	.002	.007
Women vs men	Loadings + thresholds	428 / 337	.939	.055	.068	-.009	.004	.014
Undergraduate vs postgraduate	Configural	492 / 297	.952	.049	.050	-	-	-
Undergraduate vs postgraduate	Loadings	492 / 297	.947	.051	.058	-.005	.002	.008
Undergraduate vs postgraduate	Loadings + thresholds	492 / 297	.941	.053	.066	-.011	.004	.016
Daily vs less-than-daily GenAI use	Configural	232 / 557	.946	.052	.055	-	-	-
Daily vs less-than-daily GenAI use	Loadings	232 / 557	.941	.054	.064	-.005	.002	.009
Daily vs less-than-daily GenAI use	Loadings + thresholds	232 / 557	.932	.059	.081	-.014	.007	.026

Note. Postgraduate includes taught postgraduate and doctoral students. Gender invariance was tested only between women and men because other gender categories were too small for stable multiple-group estimation. Delta values are cumulative changes relative to the corresponding configural model, not stepwise changes from the immediately preceding model.

Taken together, the score evidence supports a cautious scoring hierarchy. The total ED-AIL score may be used as a broad index of self-reported tendency towards uncritical epistemic dependence, especially when the research question concerns overall delegation of epistemic responsibility to AI. Subscale scores are appropriate for theory testing when the research question concerns specific mechanisms, such as evaluation dependence or conceptual interpretation dependence. However, subscale interpretation should remain cautious, and the epistemic regulation subscale should be treated as provisional because it showed the weakest reliability, lowest AVE, and least stable measurement evidence. The bifactor sensitivity analysis suggests a meaningful general factor, but the domain-specific factors retained sufficient substantive and practical interpretability; the total score should not be treated as evidence that ED-AIL is strictly unidimensional.

5. Discussion

5.1 Summary of findings

This study makes two linked contributions. Conceptually, it specifies uncritical epistemic dependence as a learning-specific form of AI-mediated judgement displacement, rather than as AI use frequency, general trust in AI, or cognitive offloading. Empirically, it provides initial validity evidence for the ED-AIL Scale across two independent samples of English-speaking higher education students. The evidence supports a provisional five-domain interpretation, acceptable but uneven reliability, theoretically coherent relations with adjacent constructs, modest incremental validity, and approximate measurement invariance across selected groups.

The empirical pattern also identifies clear limits to the present validation. Knowledge acquisition, conceptual interpretation, and epistemic evaluation showed the strongest evidence, suggesting that ED-AIL is most readily observed when learners use AI to obtain information, settle conceptual meaning, or judge the adequacy of answers and arguments. Knowledge production showed acceptable but more qualified evidence, consistent with the possibility that AI-supported writing may involve epistemic delegation, rhetorical assistance, linguistic support, or ordinary composition scaffolding. Epistemic regulation remained the weakest domain and should be treated as experimental until replicated.

5.2 Theoretical contribution: from AI output to judgement displacement

The central theoretical contribution is the specification of uncritical epistemic dependence as a learner-AI relation rather than as a general attitude towards technology, a preference for AI tools, or simple frequency of use. Social epistemology explains why dependence on external epistemic resources is unavoidable (Goldman, 2001; Hardwig, 1985). Epistemic cognition research explains why source evaluation, justification, and standards of knowing are central to learning (Barzilai & Chinn, 2018, 2024; Bråten et al., 2009; Hofer & Pintrich, 1997). Human-AI interaction research explains how trust, reliance, and automation bias can shape user behaviour (Lee & See, 2004; Parasuraman & Riley, 1997). ED-AIL integrates these literatures by focusing on the point at which AI-generated epistemic output is taken up as directive and begins to displace learners' own judgement.

The mechanism of output, uptake, and judgement displacement explains why GenAI raises a distinctive educational problem. AI systems do not merely provide static resources; they produce fluent explanations, summaries, evaluations, and recommendations that can appear coherent even when source provenance, evidential warrant, uncertainty, or disciplinary contestation are unclear (Kasneci et al., 2023; Labadze et al., 2023; Zhai et al., 2024). The epistemic risk does not lie in receiving such output. It lies in treating the output as sufficient for knowing, interpreting, evaluating, or proceeding. This framing preserves the insight that learning is socially and technologically mediated while identifying a measurable pattern of epistemic over-delegation.

5.3 Domain-specific interpretation

The domain-level pattern clarifies how uncritical epistemic dependence may occur across learning functions. Knowledge acquisition dependence concerns the early stage of inquiry, when AI summaries, overviews, or explanations become the learner's main reference point before independent corroboration. Conceptual interpretation dependence concerns the point at which AI explanations settle how the learner understands a concept, theory, or text. Epistemic evaluation dependence concerns the delegation of judgement about correctness, credibility, quality, or sufficiency. These three domains were strongest empirically, which is consistent with research on source evaluation and with accounts of GenAI affordances in education (Barzilai & Zohar, 2012; Bråten et al., 2009, 2011; Chan & Hu, 2023; Kasneci et al., 2023).

Knowledge production dependence requires a more cautious interpretation. AI-generated outlines, wording, and ideas may shape academic work, but production is not always reducible to epistemic dependence. Some AI-supported drafting may involve linguistic assistance, genre modelling, translation support, brainstorming, or accessibility-related scaffolding (Cotton et al., 2024; Fuchs, 2023; Yusuf et al., 2024). The production subscale should therefore be interpreted as capturing cases in which AI-generated content directs argument structure, idea selection, or expression before the learner has sufficiently evaluated it, not as a general measure of writing support.

Epistemic regulation requires the greatest caution. Regulation involves planning, monitoring, goal setting, and deciding when understanding is sufficient. These processes are broader, more recursive, and more context-dependent than accepting an AI explanation or using AI feedback to judge an answer, as self-regulated learning models emphasise (Greene & Azevedo, 2007; Panadero, 2017). The weaker regulation evidence may indicate that the current items are too few, too heterogeneous, or too close to adjacent constructs such as study planning and self-regulated learning. Future studies should test whether epistemic regulation is best modelled as a distinct subscale, a downstream outcome, or a cross-cutting metacognitive dimension of AI-mediated learning.

5.4 Methodological contribution and score interpretation

Methodologically, the study contributes a measure developed from an explicit construct domain rather than by relabelling existing measures of AI trust, technology acceptance, AI literacy, or cognitive offloading. The evidence base includes expert review, cognitive pretesting, ordinal EFA, independent CFA, item-level loading evidence, latent factor correlations, bifactor sensitivity analysis, reliability

estimates, validation measures, performance-oriented criteria, and invariance testing. These elements provide an initial validity argument for the ED-AIL Scale, but the argument remains provisional. The scale is currently strongest as a self-report measure of learners' epistemic orientation towards AI-mediated learning, not as a behavioural diagnostic instrument.

This distinction between self-report and behaviour is central to score interpretation. ED-AIL showed theoretically expected relations with AI reliance, AI trust, cognitive offloading, GenAI use frequency, source verification, self-regulated learning, unverified AI-output acceptance, and performance-oriented criteria. However, associations with the AI error-detection and source-corroboration tasks were small, and the incremental association with source-corroboration performance did not reach statistical significance after controls. The present evidence therefore suggests that ED-AIL captures a meaningful self-reported tendency towards uncritical epistemic delegation; it does not yet establish that the score strongly predicts task performance across contexts.

Accordingly, score interpretation should remain conservative. The total score currently has stronger empirical support than several domain scores and can be used as a broad index when the research question concerns general tendencies to delegate epistemic responsibility to AI. Subscale scores are appropriate for theory testing when researchers are interested in specific mechanisms, but they should not be interpreted as equally mature measures. The regulation subscale should be treated as experimental until further evidence supports its reliability, structure, and predictive validity. The scale is suitable for research, group-level profiling, intervention evaluation, and theory testing. It should not be used for high-stakes individual diagnosis, punitive academic-integrity screening, or claims that a specific student is epistemically dependent on AI.

5.5 Practical implications for pedagogy, assessment, and institutional use

The practical value of the ED-AIL Scale lies in moving AI-in-education debates beyond binary distinctions between use and non-use. A student may use GenAI frequently while maintaining strong verification and self-regulation practices; another may use it less often but rely on it heavily to decide what is true, what a reading means, whether an answer is correct, or whether learning is complete. This distinction aligns with calls to move beyond generic acceptance or prohibition debates towards pedagogical attention to AI literacy, academic integrity, assessment design, and learner judgement (Chan & Hu, 2023; Cotton et al., 2024; Ng et al., 2024; Xia et al., 2024; Yusuf et al., 2024).

The five domains point to differentiated pedagogical responses. Knowledge acquisition dependence can be addressed through source-triangulation tasks before AI-generated summaries are used in academic work (Bråten et al., 2009, 2011; Kammerer et al., 2021). Conceptual interpretation dependence can be addressed by asking students to compare AI explanations with original readings, lectures, or disciplinary texts and justify where they accept or reject the AI interpretation. Epistemic evaluation dependence can be addressed by requiring students to evaluate AI feedback against rubrics, evidence, worked examples, or disciplinary criteria. Knowledge production dependence can be addressed through process evidence, annotated drafting, reflective disclosure, and justification of how AI-generated suggestions were evaluated before inclusion. Epistemic regulation dependence can be addressed by requiring learners to set inquiry goals, monitoring criteria, and standards for sufficient understanding before consulting AI, consistent with self-regulated learning approaches (Greene & Azevedo, 2007; Panadero, 2017).

These implications extend to assessment and institutional policy. Assessment designs that reward polished final products without requiring evidence of reasoning, source use, revision, or justification may inadvertently encourage uncritical epistemic delegation (Cotton et al., 2024; Fuchs, 2023; Xia et al., 2024; Yusuf et al., 2024). Conversely, assessment designs that require students to document interpretive choices, compare sources, explain verification strategies, and justify the acceptance or rejection of AI feedback can make epistemic responsibility visible. At the policy level, ED-AIL is better suited to formative and developmental use than to enforcement. It can help programmes evaluate whether AI literacy interventions reduce uncritical delegation, but it should not be used as a proxy for misconduct or as a screening tool for individual students.

5.6 Limitations and future research

Several limitations qualify the present findings. First, although performance-oriented criteria were included, much of the validation evidence relied on self-report, and the validation battery used similar response formats across several constructs. This creates a risk of common-method inflation in correlations with AI reliance, AI trust, cognitive offloading, source verification, self-regulated learning, and unverified AI-output acceptance. Future studies should incorporate richer behavioural tasks, trace data from AI-mediated inquiry, think-aloud protocols, classroom-based performance assessments, and independent ratings of source use, argument quality, or justification quality.

Second, the discriminant-validity argument remains incomplete. The study included several adjacent AI and learning constructs, but it did not measure all plausible rivals, including epistemic vigilance, critical-thinking disposition, need for cognition, academic self-efficacy, domain knowledge, or disciplinary prior knowledge. Future validation should test whether ED-AIL explains variance beyond these broader cognitive, motivational, and epistemic characteristics.

Third, the behavioural tasks require further development. The AI error-detection and source-corroboration tasks provided useful preliminary criterion evidence, but they were brief and only modestly associated with ED-AIL. Future work should report item-level difficulty and discrimination, scoring rubrics, inter-rater reliability where open responses are used, and evidence that task performance reflects AI-specific epistemic judgement rather than general academic ability, prior knowledge, or test-taking skill. A stronger behavioural validation design would observe learners interacting with AI during authentic study, writing, or inquiry tasks and examine whether ED-AIL predicts verification behaviour, source integration, revision decisions, and justification quality.

Fourth, the samples were recruited through Prolific and completed the survey in English. The findings should therefore be described as initial evidence from online English-speaking higher education samples, not as validation across higher education generally. Replication is needed across institutional types, disciplines, languages, educational levels, and cultural settings. Disciplinary context may be especially important because epistemic standards differ across fields: the meaning of adequate evidence, acceptable interpretation, and responsible use of AI feedback may vary between, for example, literary analysis, statistics, engineering design, and teacher education.

Fifth, the study measured GenAI use broadly rather than focusing on a single platform or task ecology. Platform affordances may shape epistemic dependence by making sources more or less visible, encouraging conversational follow-up, supplying citations, producing confidence cues, or integrating into writing and learning-management systems. Future studies should test whether ED-AIL functions similarly across different AI systems and across tasks such as summarisation, reading comprehension, coding, writing, feedback seeking, and exam preparation.

Sixth, the data were cross-sectional. The study cannot determine whether ED-AIL is an antecedent, consequence, or correlate of weaker verification practices. Longitudinal research is needed to examine whether uncritical epistemic dependence increases with habitual AI use, decreases as learners gain AI literacy and disciplinary expertise, or varies across assessment demands. Intervention studies are also needed to test whether explicit instruction in corroboration, uncertainty evaluation, source comparison, and metacognitive planning reduces ED-AIL scores and improves learning outcomes. Further modelling should also examine whether productive AI-supported epistemic scaffolding can be measured alongside uncritical delegation, allowing future research to distinguish beneficial dependence from epistemically risky over-delegation.

6. Conclusion

This article introduced the ED-AIL Scale as an initial self-report instrument for assessing tendencies towards uncritical epistemic dependence in AI-mediated learning. The scale captures learners' delegation of epistemic responsibility to AI across knowledge acquisition, conceptual interpretation, epistemic evaluation, knowledge production, and epistemic regulation, with emphasis on cases where

AI output may substitute for independent verification, judgement, or justification. Across two independent online samples, the scale demonstrated promising but uneven psychometric evidence. By distinguishing ordinary AI use from uncritical epistemic delegation, ED-AIL provides a provisional basis for studying responsible AI-mediated learning, source verification, and learner epistemic agency in higher education. Its strongest immediate use is for research, theory testing, and formative evaluation; its future value depends on further validation across tasks, platforms, disciplines, languages, and authentic learning contexts.

Declarations

Conflict of interest

The author declares no competing interests.

Funding

None.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Statement

This study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was obtained from the institutional ethics committee prior to data collection.

Informed Consent

Prior to participation, all participants were informed about the purpose of the study, the voluntary nature of participation, and the confidentiality of their responses. Written informed consent was obtained from all participants before completing the questionnaire.

Acknowledgements

None

References

Abbas, *M*., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. International Journal of Educational Technology in Higher Education, 21, Article 10. https://doi.org/10.1186/s41239-024-00444-7

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.

Barzilai, S., & Chinn, C. A. (2018). On the goals of epistemic education: Promoting apt epistemic performance. Journal of the Learning Sciences, 27(3), 353-389. https://doi.org/10.1080/10508406.2017.1392968

Barzilai, S., & Chinn, C. A. (2024). The AIR and Apt-AIR frameworks of epistemic performance and growth: Reflections on educational theory development. Educational Psychology Review, 36, Article 91. https://doi.org/10.1007/s10648-024-09927-5

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. Cognition and Instruction, 30(1), 39-85. https://doi.org/10.1080/07370008.2011.636495

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioural research: A primer. Frontiers in Public Health, 6, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Bråten, I., Strømsø, H. I., & Britt, *M*. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. Reading Research Quarterly, 44(1), 6-28. https://doi.org/10.1598/RRQ.44.1.1

Bråten, I., Strømsø, H. I., & Salmeron, L. (2011). Trust and mistrust when students read multiple information sources about climate change. Learning and Instruction, 21(2), 180-192. https://doi.org/10.1016/j.learninstruc.2010.02.002

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20, Article 43. https://doi.org/10.1186/s41239-023-00411-8

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling, 14(3), 464-504. https://doi.org/10.1080/10705510701301834

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. Psychological Assessment, 7(3), 309-319. https://doi.org/10.1037/1040-3590.7.3.309

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

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

Dzindolet, *M*. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. International Journal of Human-Computer Studies, 58(6), 697-718. https://doi.org/10.1016/S1071-5819(03)00038-7

Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. Canadian Journal of Behavioural Science, 49(2), 78-88. https://doi.org/10.1037/cbs0000069

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: Is ChatGPT a blessing or a curse? Frontiers in Education, 8, Article 1166682. https://doi.org/10.3389/feduc.2023.1166682

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association, 19(1), 121-127. https://doi.org/10.1136/amiajnl-2011-000089

Goldman, A. I. (2001). Experts: Which ones should you trust? Philosophy and Phenomenological Research, 63(1), 85-110. https://doi.org/10.1111/j.1933-1592.2001.tb00093.x

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. Review of Educational Research, 77(3), 334-372. https://doi.org/10.3102/003465430303953

Greene, J. A., Cartiff, B. *M*., & Duke, R. F. (2018). A meta-analytic review of the relationship between epistemic cognition and academic achievement. Journal of Educational Psychology, 110(8), 1084-1111. https://doi.org/10.1037/edu0000263

Hardwig, J. (1985). Epistemic dependence. The Journal of Philosophy, 82(7), 335-349. https://doi.org/10.2307/2026523

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67(1), 88-140. https://doi.org/10.3102/00346543067001088

Hornberger, *M*., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. Computers and Education: Artificial Intelligence, 5, Article 100165. https://doi.org/10.1016/j.caeai.2023.100165

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

Jian, J.-Y., Bisantz, A. *M*., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics, 4(1), 53-71. https://doi.org/10.1207/S15327566IJCE0401_04

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Chen, D., Dai, W., Chan, H. S., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), Article 248. https://doi.org/10.1145/3571730

Jo, H. (2024). From concerns to benefits: A comprehensive study of ChatGPT usage in education. International Journal of Educational Technology in Higher Education, 21, Article 35. https://doi.org/10.1186/s41239-024-00471-4

Kammerer, Y., Gottschling, S., & Bråten, I. (2021). The role of Internet-specific justification beliefs in source evaluation and corroboration during web search on an unsettled socio-scientific issue. Journal of Educational Computing Research, 59(2), 342-378. https://doi.org/10.1177/0735633120952731

Kasneci, E., Seßler, K., Kuchemann, S., Bannert, *M*., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, *M*., Schmidt, A., Seidel, T., Stadler, *M*., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language

models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Labadze, L., Grigolia, *M*., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. International Journal of Educational Technology in Higher Education, 20, Article 56. https://doi.org/10.1186/s41239-023-00426-1

Laupichler, *M*. C., Aster, A., & Raupach, T. (2023). Delphi study for the development and preliminary validation of an item set for the assessment of non-experts' AI literacy. Computers and Education: Artificial Intelligence, 4, Article 100126. https://doi.org/10.1016/j.caeai.2023.100126

Laupichler, *M*. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. Computers and Education: Artificial Intelligence, 3, Article 100101. https://doi.org/10.1016/j.caeai.2022.100101

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

List, A., & Alexander, P. A. (2019). Toward an integrated framework of multiple text use. Educational Psychologist, 54(1), 20-39. https://doi.org/10.1080/00461520.2018.1505514

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. British Journal of Educational Technology, 55(3), 1082-1104. https://doi.org/10.1111/bjet.13411

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. Frontiers in Psychology, 8, Article 422. https://doi.org/10.3389/fpsyg.2017.00422

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230-253. https://doi.org/10.1518/001872097778543886

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Schommer, *M*. (1990). Effects of beliefs about the nature of knowledge on comprehension. Journal of Educational Psychology, 82(3), 498-504. https://doi.org/10.1037/0022-0663.82.3.498

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modelling and confirmatory factor analysis results: A review. The Journal of Educational Research, 99(6), 323-338. https://doi.org/10.3200/JOER.99.6.323-338

Shahzad, *M*. F., Xu, S., & Javed, I. (2024). ChatGPT awareness, acceptance, and adoption in higher education: The role of trust as a cornerstone. International Journal of Educational Technology in Higher Education, 21, Article 46. https://doi.org/10.1186/s41239-024-00478-x

Skitka, L. J., Mosier, K. L., & Burdick, *M*. (1999). Does automation bias decision-making? International Journal of Human-Computer Studies, 51(5), 991-1006. https://doi.org/10.1006/ijhc.1999.0252

Skitka, L. J., Mosier, K. L., & Burdick, *M*. (2000). Accountability and automation bias. International Journal of Human-Computer Studies, 52(4), 701-717. https://doi.org/10.1006/ijhc.1999.0349

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478. https://doi.org/10.2307/30036540

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. The Counseling Psychologist, 34(6), 806-838. https://doi.org/10.1177/0011000006288127

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. International Journal of Educational Technology in Higher Education, 21, Article 40. https://doi.org/10.1186/s41239-024-00468-z

Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. International Journal of Educational Technology in Higher Education, 21, Article 21. https://doi.org/10.1186/s41239-024-00453-6

Zawacki-Richter, O., Bai, J. Y. H., Lee, K., Slagter van Tryon, P. J., & Prinsloo, P. (2024). New advances in artificial intelligence applications in higher education? International Journal of Educational Technology in Higher Education, 21, Article 32. https://doi.org/10.1186/s41239-024-00464-3

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 11, Article 28. https://doi.org/10.1186/s40561-024-00316-7