Jupiter: Jurnal Pendidikan Terapan
Homepage : https://journal.diginus.id/jupiter/index
Email: jurnalpendidikanterapan@gmail.com

p-ISSN : 2985-3214; e-ISSN : 2964-3171
Jupiter, Vol. 4, No. 5, November 2026
Page 612-629

Exploring the Factor Structure and Reliability of a Generative AI Use and Dependence Scale for English Writing among Indonesian University Students

Venny Karolina[1*], Urai Salam[2], Farina Tazijan[3], Carla Queiroz[4], Toby Xaixanith[5], Lindsay M. Ozarraga[6], Qonita Indah Arrusli[7]
[1,2,7]Universitas Tanjungpura, Indonesia
[3]Universiti Teknologi MARA, Selangor Malaysia
[4]Instituto Superior de Administração e Finanças, Academia BAI, Angola
[5]National University of Laos, Laos
[6]Mindanao State University, Phillipina

ARTICLE INFO

Article history:
Submitted: October 06, 2026
Final Revised: Nov. 07, 2026
Accepted: November 10, 2026
Published: November 22, 2026

Keywords:
Generative AI Dependence; Scale Development; Exploratory Factor Analysis; Cronbach's Alpha; University Students.

ABSTRACT

Purpose: This study explores the factor structure and internal consistency of a 24-item Generative AI Use and Dependence Scale for English writing among Indonesian university students and describes how the items perform. **Methods**: A quantitative exploratory scale-development design was applied to responses from 453 students, recruited by non-probability self-selection, on 24 five-point Likert items from six proposed dimensions. Analyses comprised descriptive statistics, corrected item-total correlations, exploratory factor analysis (principal axis factoring, Oblimin rotation), Cronbach's alpha, and McDonald's omega. **Findings**: The data were suitable for factor analysis (KMO = .93; Bartlett's $\chi^2(276) = 4{,}524.42$, $p < .001$), but the six proposed dimensions were not recovered. The Kaiser criterion supported four correlated factors (47.4% of variance): Habitual Use, Instrumental Reliance and Educational Value, Emotional–Behavioral Dependence, and Reduced Learner Autonomy. Parallel analysis supported only three. The four-factor solution was kept as a provisional exploratory model on interpretive grounds (fewer cross-loadings, autonomy kept distinct from dependence), not statistical superiority. Alpha was .91 overall and .73–.85 for the four factors (omega .74–.86); two behavioral dependence items loaded clearly below .40 (one functional reliance item marginally so). **Research Implications**: These preliminary findings do not validate the scale. Scores should be provisional and factor-level until the structure is confirmed independently; weak items need revision, content validity evidence is limited to a post hoc relevance review by three experts (mean Aiken's V = .77; 10 of 24 items at V ≥ .80), and the non-probability, regionally concentrated sample limits generalizability. **Originality**: Unlike existing AI dependence scales, which mainly address problematic use, this instrument combines use, reliance, dependence, autonomy, and perceived educational value in one measure anchored to English writing; whether it adds information beyond existing scales remains to be tested.

Doi: https://doi.org/10.61255/jupiter.v4i5.2236

INTRODUCTION

Generative artificial intelligence (AI) chatbots have moved quickly from novelty to everyday academic tool. Since the public release of ChatGPT, university students have used it to draft and edit text, summarize readings, explain concepts, and work through problems, and surveys of college students show that many now use such tools regularly (Baek et al., 2024; Stöhr et al., 2024). Studies of students' perceptions describe real benefits, including immediate feedback, personalized explanations, and writing support, alongside concerns about accuracy, academic integrity, and over-reliance (Chan & Hu, 2023; Tlili et al., 2023). Engagement differs across disciplines and student groups (Qu et al., 2024; Stöhr et al., 2024) , and adoption is shaped by trust and perceived usefulness (Shahzad et al., 2024).

Concern about over-reliance now has an empirical basis. A systematic review linked excessive dependence on AI dialogue systems to weaker critical thinking, reduced independent analysis, and less original thought (Zhai et al., 2024). Gerlich (2025) reported that more frequent use of AI tools was associated with lower critical thinking,

an association mediated by cognitive offloading, and Abbas et al. (2024) reported that ChatGPT use among university students was linked to procrastination, memory loss, and lower academic performance.

To explain when helpful use turns into dependence, researchers draw on behavioral addiction, technology acceptance, and cognitive offloading. Behavioral-addiction models treat problematic technology use as a multidimensional pattern that includes emotional reactions, salience, and impaired control rather than usage frequency alone (Andreassen et al., 2012; Griffiths, 2005). Technology acceptance research explains why perceived usefulness drives adoption (Davis et al., 1989), and cognitive offloading research suggests that delegating demanding thinking to external tools can improve immediate performance while weakening memory and independent processing (Grinschgl et al., 2021). Empirical work on AI dependency has linked it to students' academic self-efficacy, academic stress, and performance expectations (S. Zhang et al., 2024), and, for conversational AI, to social anxiety and loneliness (Hu et al., 2023).

Several instruments now exist (Table 1). Morales-García et al. (2024) developed a scale for dependence on AI among university students, which later authors describe as unidimensional with a limited item pool (Wu et al., 2025). Zhang et al. (2025) validated an AI chatbot dependence scale, and Maral et al. (2025) the Problematic ChatGPT Use Scale. Chen et al. (2025) developed a 20-item conversational AI dependence scale for Chinese college students, whose four dimensions (uncontrollability, withdrawal symptoms, mood modification, and negative impacts) correspond to components of behavioral addiction. Wu et al. (2025) proposed a four-dimensional AI dependence scale covering emotional, functional, and cognitive dependence and loss of control, and supported it with exploratory and confirmatory factor analyses in two Chinese student samples. Kasap & Babayiğit (2026) developed a 10-item AI addiction scale with two dimensions (compulsive use and cognitive dependence; functional impairment and emotional attachment) for English language learners in Türkiye.

Table 1. Selected Existing AI Dependence Instruments and the Measurement Needs They Leave Open

Instrument	Reported structure and sample	What it leaves open for the present purpose
Dependence on AI Scale (Morales-García et al., 2024)	Unidimensional; short item pool (as described by Wu et al. (2025)	Cannot show whether high scores reflect emotional comfort, functional convenience, or impaired control Wu et al. (2025) .
Conversational AI Dependence Scale (Chen et al., 2025)	20 items; four dimensions (uncontrollability, withdrawal symptoms, mood modification, negative impacts); Chinese college students	Content follows addiction components and sits at the problematic end; no indicators of habitual use, task-specific reliance, or perceived value; not tied to an academic task.
AI Dependence Scale, AIDep-22 (Wu et al., 2025)	22 items; four dimensions (emotional, functional, and cognitive dependence; loss of control); exploratory and confirmatory analyses in two samples of Chinese undergraduates (N = 400 each)	Measures dependence only; no separate appraisal of educational value or habitual use; general academic use rather than a specific writing task; Chinese samples only.
AI Addiction Scale, AIAS-10 (Kasap & Babayiğit, 2026)	10 items; two dimensions (compulsive use and cognitive dependence; functional impairment and emotional attachment); English language learners and working professionals in Türkiye	Brief and limited to addictive patterns; no use, value, or autonomy indicators; developed outside Indonesia.

Note. Structures are summarized from the cited reports. Maral et al. (2025) and Zhang et al. (2025) also target problematic use of, or dependence on, conversational AI.

Taken together, these instruments leave three measurement needs open for the present purpose. First, they concentrate on the problematic end of the continuum. Even the multidimensional scales (Chen et al., 2025; Wu et al., 2025) do not place dependence next to students' habitual use, the extent to which they rely on AI tools for particular tasks, or their appraisal of educational value, which is what allows productive reliance to be told apart from problematic dependence, and the unidimensional scale cannot separate these sources at all (Wu et al., 2025). Second, they are general-purpose. With the exception of the English-learner scale of Kasap & Babayiğit (2026) , none is anchored to a specific academic task, yet "relying on AI" means different things when solving a mathematical problem and when writing in a foreign language, where tools can scaffold learning but can also replace the learner's own practice. The English-learner scale is brief, restricted to addictive patterns, and was developed in Türkiye with a sample that mixed students and professionals. Third, to the authors' knowledge, psychometric evidence from Indonesian university students remains limited, and the cultural and educational settings in which the scales above were developed (mainly China and Türkiye) cannot be assumed to transfer. These points do not show that the existing scales are invalid for Indonesian students. They show that no available

instrument was designed to measure use, reliance, dependence, autonomy, and perceived educational value together within one academic task, which is the gap this study addresses. A Generative AI Use and Dependence Scale for English writing was therefore drafted around six theoretically proposed dimensions, with 24 items referring to the use of ChatGPT and similar AI tools when writing English assignments. In scale development, an exploratory analysis in an adequate sample precedes confirmatory testing (DeVellis, 2012; Watkins, 2018) , and it is first necessary to know whether the proposed dimensions emerge from students' responses at all.

The scale treats AI use in English writing as a continuum from productive use to problematic dependence, and the six proposed dimensions are intended to mark different positions on it. *Frequency of Use* refers to how habitually students turn to AI tools when writing in English. It is included because habitual use is a precondition of dependence but not dependence itself: behavioral-addiction models distinguish engagement from the emotional and control-related features of addiction (Andreassen et al., 2012; Griffiths, 2005). *Functional Reliance* is the instrumental delegation of specific writing tasks, such as grammar checking or organizing ideas, to AI tools. In line with technology acceptance research and the cognitive offloading literature (Davis et al., 1989; Grinschgl et al., 2021), such reliance can be a rational and even beneficial use of a tool as long as the learner keeps control of the task. Reliance becomes potentially problematic when it takes on affective and behavioral features. *Emotional Dependence* covers anxiety, fear of error, and greater trust in the tool than in oneself, and *Behavioral Dependence* covers difficulty starting or finishing a text without AI input and uncritical acceptance of its output; both correspond to the emotional reactions and impaired control that behavioral-addiction frameworks treat as markers of problematic use (Andreassen et al., 2012; Griffiths, 2005). *Reduction of Learner Autonomy* is the learner's perception that independent control over their own writing has declined. It is treated as a perceived consequence rather than a component of dependence, in line with concerns that delegating thinking to AI weakens independent processing (Grinschgl et al., 2021; Zhai et al., 2024). *Perceived Educational Value* is the appraisal of the tool as beneficial for learning; it makes it possible to distinguish a student who relies heavily on AI but judges the use to be beneficial from one whose reliance is accompanied by emotional and behavioral dependence. These concepts are related but are treated as distinct. Whether they are empirically separable, and whether the scale adds information beyond existing dependence scales, are empirical questions that this exploratory study can address only in part.

This study therefore explores the factor structure and internal consistency of the 24-item scale among Indonesian university students in the context of English writing. Specifically, it examines the underlying factor structure of the items (RQ1: What factor structure underlies the scale?), determines the reliability of the identified dimensions (RQ2: How reliable are the dimensions?), and describes the scale's descriptive characteristics and item-level performance (RQ3: What are the descriptive characteristics and item-level performance of the scale?). The findings are preliminary and do not constitute a full validation of the scale.

METHOD

Research design

The study used a quantitative, exploratory scale-development design (Carpenter, 2018; DeVellis, 2012). A single cross-sectional survey provided the data for examining the factor structure and internal consistency of the Generative AI Use and Dependence Scale. Confirmatory testing is left to later research.

Participants

Participants were 453 university students who completed an online questionnaire distributed through Google Forms. They came from eight groups of higher-education institutions, mostly in Pontianak, West Kalimantan, with additional respondents from UIN KHAS Jember, Universitas Negeri Manado, and other institutions (Table 2). The inclusion criteria were: (1) currently enrolled as a university student, (2) having used ChatGPT or similar AI tools for English writing tasks, and (3) willingness to participate voluntarily. Respondents were self-selected, so the sample is non-probability. Institution names were typed in freely by respondents and were grouped by the authors for reporting. Although respondents came from institutions in several provinces, more than two-thirds (n = 308, 68.0%) came from four institution groups in Pontianak, so the sample is geographically concentrated and cannot be treated as representative of Indonesian university students. Gender, age, and year of study were not recorded in the dataset analyzed here. The sample therefore cannot be described by demographic characteristics, and the analyses cannot show whether the scale performs equivalently across student groups. With 453 respondents and 24 items, the sample provides about 19 respondents per item, above the minimum ratios usually recommended for factor analysis (Hair et al., 2019), although sample size alone does not establish generalizability. Informed consent was obtained online from all respondents before participation, and the data were collected anonymously to protect respondents' identities and confidentiality.

Data were collected between June and November 2025 using an online questionnaire administered through Google Forms. The questionnaire link was distributed to university students through lecturers and student WhatsApp groups across participating higher education institutions in Indonesia. Participants were recruited using

purposive sampling based on three inclusion criteria: (1) being currently enrolled as university students, (2) having prior experience using generative AI tools, such as ChatGPT, DeepSeek, or Gemini, for English writing tasks, and (3) voluntarily agreeing to participate in the study. Before completing the questionnaire, participants were informed about the purpose of the research and the voluntary nature of their participation. The questionnaire was administered in Indonesian to ensure that participants could clearly understand the statements. The English versions of the 24 items presented in Appendix A were translated from the Indonesian questionnaire for reporting purposes. A total of 453 completed responses were obtained and included in the final analysis.

Table 2. Participants by Institution Group

Institution group	*n*	%
Universitas Muhammadiyah Pontianak	107	23.6
Universitas Tanjungpura	89	19.6
UIN KHAS Jember	80	17.7
Universitas Nahdlatul Ulama Kalimantan Barat	60	13.2
Poltekkes / Politeknik (Pontianak)	52	11.5
IAIN (Pontianak and Kerinci)	33	7.3
Universitas Negeri Manado	13	2.9
Other institutions	19	4.2
Total	**453**	**100.0**

Note. N = 453. Groups were formed from free-text responses. The four Pontianak groups (Universitas Muhammadiyah Pontianak, Universitas Tanjungpura, Universitas Nahdlatul Ulama Kalimantan Barat, and Poltekkes/Politeknik) together comprised 308 respondents (68.0%).

Instrument

The instrument has 24 items distributed across six theoretically proposed dimensions of four items each (Table 3). The items refer to the use of ChatGPT and similar AI tools (e.g., ChatGPT, DeepSeek, Gemini) when writing English assignments, and the full wording is given in Appendix A. Because every item names several tools together, responses refer to generative AI tools in general and cannot be attributed to ChatGPT specifically; the instrument is therefore called the Generative AI Use and Dependence Scale. Respondents rated each item on a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree). Higher scores indicate stronger agreement, and no item was reverse-scored in the analyses. All dimensions, including Frequency of Use, were measured with agreement statements. The Frequency of Use items describe habitual or routinized use (for example, using AI tools almost always, or immediately, when writing in English) and do not ask for counts of sessions or time spent, so the dimension reflects perceived habitual use and not behaviorally anchored frequency.

Table 3. Dimensions and Items of the Generative AI Use and Dependence Scale

Dimension	Code	Items
Frequency of Use	F	F1–F4
Functional Reliance	FR	FR1–FR4
Emotional Dependence	ED	ED1–ED4
Behavioral Dependence	BD	BD1–BD4
Reduction of Learner Autonomy	LA	LA1–LA4
Perceived Educational Value	EV	EV1–EV4

Item development

The 24 items were written by the authors on the basis of the behavioral-addiction, technology acceptance, and cognitive offloading literature and the instruments reviewed in the Introduction (DeVellis, 2012; Hinkin, 1998), with four items per dimension. Following the initial survey, the content relevance of the items was evaluated in an additional expert-assessment stage. Three evaluators with expertise in educational measurement, educational technology, and English language education each judged the relevance of every item to its proposed theoretical dimension on a five-point scale (1 = not relevant, 2 = slightly relevant, 3 = moderately relevant, 4 = relevant, 5 = highly relevant). They also gave written comments on clarity, conceptual relevance, and overlap between items; these comments were read item by item and summarized descriptively. Content relevance was quantified with (Aiken, 1985) V, computed for each item as $V = \Sigma(r - 1) / [n(c - 1)]$, where r is an expert's rating, n is the number of experts (3), and c is the number of rating categories (5). With these values the denominator is 12, and V ranges from 0 to 1. A value of $V \geq .80$ was used as a practical benchmark of satisfactory relevance; this is a convention and not a significance test. Statistical significance was examined separately, using the exact one-tailed probability of the observed rating sum under a uniform null distribution (all $5^3 = 125$ equally likely combinations of three ratings; Aiken, 1985) at $\alpha = .05$. Dimension-level values are the means of the item coefficients. These assessments were conducted after the original survey, and the items analyzed in this study were not altered in response to them. They are therefore treated as supplementary evidence concerning content relevance and not as validation

conducted before data collection. Only relevance was rated; clarity and representativeness were commented on qualitatively but were not scored.

Data analysis

Analyses were run in IBM SPSS Statistics 27. Parallel analysis and the minimum average partial test used (O'connor, 2000) SPSS syntax, and the polychoric analysis was also run in SPSS. McDonald's omega and a supplementary outlier screen on all 24 items were computed in Python (NumPy, SciPy) from the same raw data. Data screening examined missing values, univariate skewness and kurtosis, and multivariate outliers. Multivariate outliers were identified with Mahalanobis distance obtained from the SPSS linear regression procedure, in which five of the six dimension scores were entered as predictors and the remaining dimension (Frequency of Use) as the dependent variable. Because the distance is computed on the predictors only, the critical value was based on 5 degrees of freedom, $\chi^2(5) = 20.52$ at $p < .001$ (Tabachnick & Fidell, 2019). This procedure screens for atypical combinations of five composite scores and not of the full 24-item response pattern, and the dimension used as the dependent variable does not contribute to the distance. As a supplementary screen, Mahalanobis distance was therefore also computed on all 24 items (critical $\chi^2(24) = 51.18$ at $p < .001$), and the factor analysis was repeated without the flagged cases in both screens. Descriptive statistics (mean, standard deviation, minimum, maximum, skewness, and kurtosis) were computed for each item. Item analysis used corrected item-total correlations (CITC), with values of .30 or higher treated as acceptable, together with Cronbach's alpha if the item was deleted.

Exploratory factor analysis (EFA) was performed on the 24 items. Suitability was checked with the Kaiser–Meyer–Olkin (KMO) measure (Kaiser, 1974) and Bartlett's test of sphericity. Principal axis factoring (PAF) was used for extraction, and Oblimin rotation was applied because the dimensions were expected to correlate (Watkins, 2018). The number of factors was judged by several criteria: the Kaiser criterion (eigenvalue > 1), the scree plot, parallel analysis (Horn, 1965), and the minimum average partial test (Velicer, 1976). The Kaiser criterion is known to over-extract, whereas parallel analysis and the minimum average partial test are generally regarded as more accurate (Fabrigar et al., 1999; Hayton et al., 2004; Watkins, 2018). Where the criteria disagreed, adjacent solutions were therefore compared on loading patterns, cross-loadings, communalities, and interpretability (Fabrigar et al., 1999). Parallel analysis used the 95th percentile of eigenvalues from 500 random normal datasets of the same size (453 × 24) and was repeated on the reduced correlation matrix. Items were assigned to the factor on which they loaded most strongly. A primary loading of at least .40 was required, and items below that value or with cross-loadings of .30 or more were flagged for review (Hair et al., 2019; Watkins, 2018). Loadings are reported to two decimals, except FR4, whose loading falls just below the cutoff and is reported to three decimals. Communalities, explained variance, and factor correlations were also examined. Because five-point items are ordinal, the analysis was repeated on a polychoric correlation matrix as a sensitivity check (Lloret et al., 2017; Timmerman & Lorenzo-Seva, 2011). Polychoric correlations were estimated by two-step maximum likelihood, and the matrix was positive definite (smallest eigenvalue = .14), so no smoothing was needed. The main solution was also repeated after excluding multivariate outliers. Agreement between the Pearson-based and polychoric-based solutions was assessed with Tucker's congruence coefficient, with values of .95 or higher taken to indicate factor equivalence (Lorenzo-Seva & Ten Berge, 2006).

Exploratory factor analysis was chosen over confirmatory analysis because the item pool had no earlier empirical testing and the proposed six-dimension structure had no prior psychometric support (DeVellis, 2012; Fabrigar et al., 1999). A common-factor method (principal axis factoring) was used instead of principal component analysis because the aim was to identify latent constructs that explain the shared variance among the items and not to reduce the data (Fabrigar et al., 1999; Watkins, 2018), and an oblique rotation was used because the dimensions were expected to correlate. Because five-point items are ordinal, the Pearson-based solution was complemented with a polychoric sensitivity analysis. All results are interpreted as preliminary psychometric evidence and not as validation of the scale.

Reliability was assessed with Cronbach's alpha for the total scale, for the six proposed dimensions, and for the dimensions resulting from the EFA. Because alpha assumes equal loadings and does not index dimensionality (Cortina, 1993), McDonald's omega was also computed for each dimension and factor and for the total scale (Dunn et al., 2014; Hayes & Coutts, 2020). For each dimension and factor, omega was computed as $\omega = (\Sigma\lambda)^2 / [(\Sigma\lambda)^2 + \Sigma(1 - \lambda^2)]$, where λ are the standardized loadings of its items from a one-factor principal-axis solution. For the total scale, omega total was computed as $1 - \Sigma(1 - h^2) / 1'R1$, where h^2 are the communalities from the four-factor principal-axis solution and $1'R1$ is the sum of all elements of the 24 × 24 item correlation matrix (the variance of the unit-weighted sum of standardized items); the same formula with a one-factor solution gave a general-factor estimate. Empirical-factor scales used the item sets listed in Table 9. A coefficient of .70 or higher was treated as acceptable, and values above about .90 were read cautiously because they can indicate item redundancy (Boateng et al., 2018; Tavakol & Dennick, 2011).

RESULTS

Data screening and descriptive characteristics (RQ3)

No responses were missing, and all 453 cases were retained. Univariate skewness ranged from −1.20 (EV4) to 0.95 (BD1) and kurtosis from −1.30 (BD3) to 2.27 (EV4), values within the limits generally accepted for factor analysis (Kline, 2005). Mahalanobis distances (df = 5) ranged from 0.32 to 27.78 (M = 4.99), and three cases exceeded the critical value of $\chi^2(5) = 20.52$ at $p < .001$. These were valid responses within the 1–5 range, so they were retained. Repeating the factor analysis without them left every primary loading assignment unchanged and shifted loadings by no more than .03. As a supplementary check, Mahalanobis distance computed on all 24 items (critical $\chi^2(24) = 51.18$ at $p < .001$) flagged 18 cases (4.0%), including these three. They were valid responses and were retained. Repeating the factor analysis without them (n = 435) left 23 of 24 primary assignments unchanged; ED1 moved from Habitual Use (.42 to .37) to Emotional–Behavioral Dependence (.33 to .39), loadings changed by up to .11, FR4 rose to .43, and BD2 (.39) and BD3 (.38) remained below .40 (explained variance 49.3%).

Item means ranged from 2.22 (BD1) to 3.88 (EV4), and standard deviations from 0.80 to 1.13 (Table 4). Perceived Educational Value items (3.29–3.88) and Functional Reliance items (3.26–3.53) received the highest ratings, whereas Behavioral Dependence (2.22–2.78) and Emotional Dependence items (2.37–2.87) received the lowest. At the dimension level, mean scores were 3.66 for Perceived Educational Value, 3.38 for Functional Reliance, 2.75 for Frequency of Use, 2.62 for Reduction of Learner Autonomy, 2.54 for Emotional Dependence, and 2.48 for Behavioral Dependence. Respondents on average therefore agreed that generative AI tools are educationally valuable and that they rely on them for tasks, but tended to disagree that they are emotionally or behaviorally dependent on them. Every item used the full 1–5 range.

Item analysis within the six proposed dimensions showed corrected item-total correlations from .43 (BD4) to .68 (ED2), all above .30. The weakest items were BD1, BD3, and BD4 (.43–.44), LA4 (.47), and FR4 (.48). Deleting any single item would not have raised the alpha of its dimension (Table 4).

Table 4. Item Descriptive Statistics and Item Analysis by Proposed Dimension

Item	*M*	*SD*	Skewness	Kurtosis	CITC	*α* if item deleted
Frequency of Use (F), α = .806						
F1	3.02	1.00	−.14	−1.17	.64	.75
F2	2.89	1.00	.13	−1.28	.68	.73
F3	2.62	1.02	.35	−1.02	.59	.77
F4	2.46	0.94	.80	−.15	.58	.78
Functional Reliance (FR), α = .776						
FR1	3.53	0.99	−.71	−.27	.64	.69
FR2	3.27	1.04	−.35	−1.04	.61	.70
FR3	3.47	0.97	−.80	−.23	.60	.71
FR4	3.26	1.04	−.39	−.98	.48	.77
Emotional Dependence (ED), α = .821						
ED1	2.87	1.07	.18	−1.17	.59	.80
ED2	2.42	0.96	.76	−.14	.68	.76
ED3	2.37	1.05	.63	−.56	.66	.77
ED4	2.51	1.04	.58	−.68	.65	.77
Behavioral Dependence (BD), α = .680						
BD1	2.22	0.91	.95	.57	.44	.63
BD2	2.49	0.96	.56	−.62	.54	.56
BD3	2.78	1.02	.09	−1.30	.44	.63
BD4	2.42	0.96	.55	−.71	.43	.63
Reduction of Learner Autonomy (LA), α = .733						
LA1	2.41	0.95	.69	−.34	.53	.67
LA2	2.66	1.06	.34	−1.09	.60	.63
LA3	2.51	1.00	.54	−.65	.50	.69
LA4	2.91	1.13	.11	−1.17	.47	.71
Perceived Educational Value (EV), α = .798						
EV1	3.67	0.91	−1.00	.58	.66	.72
EV2	3.79	0.91	−1.07	.96	.59	.76
EV3	3.29	0.97	−.55	−.84	.61	.75
EV4	3.88	0.80	−1.20	2.27	.60	.76

Note. N = 453. CITC = corrected item-total correlation within the proposed dimension. Skewness SE = .115 and kurtosis SE = .229 for all items. Item scores range from 1 to 5.

Factor structure (RQ1)

The correlation matrix was suitable for factoring. The KMO measure was .93, individual measures of sampling adequacy ranged from .86 to .95, and Bartlett's test was significant, $\chi^2(276) = 4{,}524.42$, $p < .001$. No correlation exceeded .80 (largest $r = .66$), and the determinant (3.68×10^{-5}) was above .00001, so multicollinearity was not a concern.

Four initial eigenvalues exceeded 1 (8.07, 2.86, 1.42, and 1.07) and together accounted for 55.9% of the variance (Table 5), but the other retention criteria did not agree with the Kaiser criterion. The scree plot bent after the second and third factors (Figure 1). Parallel analysis supported three factors, because the fourth eigenvalue (1.07) fell below the random eigenvalue of 1.31. The minimum average partial test was lowest (.0124) at two and at three factors. Parallel analysis on the reduced correlation matrix, in contrast, favored five factors (observed eigenvalues of 0.52 and 0.34 for factors 4 and 5 against random values of 0.37 and 0.32). The criteria therefore agree that the six factors implied by the theoretical model are not supported (the fifth and sixth initial eigenvalues were 0.91 and 0.81), but they do not agree on whether three or four factors should be retained. When six factors were forced, the EV, LA, and FR1–FR3 items each formed their own factor, but the ED and BD items merged, the F items split between two factors, and the sixth factor was defined only by loadings of .40–.43 (F4, BD2, BD3). Both the three- and four-factor solutions were therefore examined. The four-factor solution is presented first as a working structure and should be regarded as tentative; the comparison follows Table 7.

Table 5. Eigenvalues, Parallel Analysis, and Explained Variance

Factor	Initial eigenvalue	% of variance	Cumulative %	Random eigenvalue (95th pct.)	Extraction SS loadings	% of variance (extracted)
1	8.07	33.6	33.6	1.52	7.56	31.5
2	2.86	11.9	45.5	1.42	2.33	9.7
3	1.42	5.9	51.5	1.36	0.92	3.8
4	1.07	4.5	55.9	1.31	0.57	2.4
5	0.91	3.8	59.7	1.26	–	–
6	0.81	3.4	63.1	1.22	–	–

Note. Extraction by principal axis factoring; the four extracted factors explain 47.4% of the variance. Random eigenvalues are 95th percentiles from 500 simulated datasets (453 × 24).

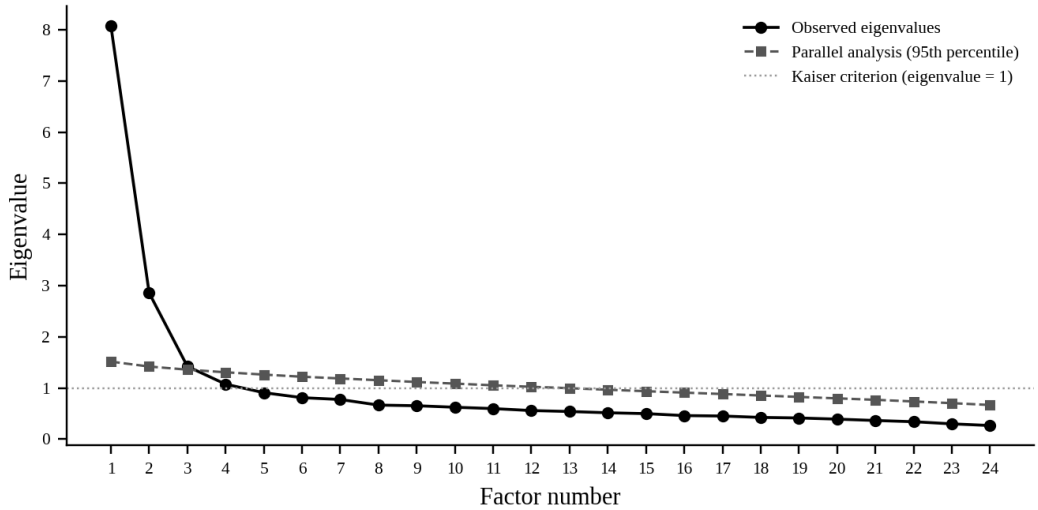

Figure 1. Scree Plot with Parallel Analysis Thresholds

Principal axis factoring with Oblimin rotation extracted four factors that explained 47.4% of the variance after extraction (Table 5). The factors were correlated ($r = .12$ to $.59$; Table 7), which supports the oblique rotation. Table 6 presents the pattern matrix.

Factor 1, Habitual Use, comprised F1–F4 (loadings .53–.81) and ED1 (.42, with a cross-loading of .33 on Factor 3). Factor 2, Instrumental Reliance and Educational Value, combined all four Functional Reliance items (.397–.63, with FR4 marginal) with all four Perceived Educational Value items (.59–.74). Factor 3, Emotional–Behavioral Dependence, was defined by ED2–ED4 (.61–.71), BD1 (.43), and BD4 (.47). Factor 4, Reduced Learner Autonomy, was defined by LA1–LA4 (.41–.71), although LA1 and LA3 also loaded .31 and .29 on Factor 3. Three items did not reach the .40 criterion on any factor: BD2 (.31), BD3 (.32, with .29 on Factor 1), and, marginally, FR4 (.397). Communalities after extraction ranged from .31 (BD4) to .62 (F2) with a mean of .47; the lowest were BD4 (.31), LA4 (.32), BD1 (.37), FR4 (.38), and BD2 (.38).

Table 6. Pattern Matrix and Communalities from Principal Axis Factoring with Oblimin Rotation

Item	F1	F2	F3	F4	h^2
Factor 1: Habitual Use					
F1	**.79**				.60
F2	**.81**				.62
F3	**.56**				.46
F4	**.53**				.49
ED1	**.42**		.33		.53
Factor 2: Instrumental Reliance and Educational Value					
FR1		**.63**			.51
FR2		**.53**			.49
FR3		**.56**			.44
FR4†		**.397**			.38
EV1		**.71**			.52
EV2		**.74**			.50
EV3		**.59**			.46
EV4		**.70**			.46
Factor 3: Emotional–Behavioral Dependence					
ED2			**.63**		.60
ED3			**.71**		.53
ED4			**.61**		.54
BD1			**.43**		.37
BD4			**.47**		.31
BD2†			**.31**		.38
BD3†			**.32**		.44
Factor 4: Reduced Learner Autonomy					
LA1			.31	**.43**	.47
LA2				**.71**	.55
LA3				**.41**	.43
LA4				**.53**	.32

Note. N = 453. Loadings below .30 are suppressed; the highest loading in each row is in bold. h^2 = communality after extraction. †Primary loading below .40 (BD2 and BD3 clearly; FR4 = .397, marginally). BD2 and BD3 are listed under the factor of their strongest loading for completeness but were excluded from the five-item Emotional–Behavioral Dependence reliability estimate (Table 9); FR4 was retained in the eight-item Instrumental Reliance and Educational Value scale.

Table 7. Correlations between the Four Factors

Factor	1	2	3	4
1. Habitual Use	1.00			
2. Instrumental Reliance and Educational Value	.43	1.00		
3. Emotional–Behavioral Dependence	.59	.26	1.00	
4. Reduced Learner Autonomy	.31	.12	.43	1.00

Comparison of the three- and four-factor solutions (RQ1)

In the three-factor solution (44.7% of variance), the Instrumental Reliance and Educational Value factor was unchanged. The first factor absorbed several emotional dependence items (ED1–ED4, .39–.60) and BD3 (.45) in addition to F1–F4, and the third factor combined the autonomy items (LA1–LA4, .52–.74) with the behavioral dependence items BD1, BD2, and BD4 (.39–.49), with ED2–ED4 also loading on it (.35–.40). Table 8 compares the two solutions. Complete pattern matrices for both solutions, showing every loading regardless of size, are given in Appendix B (Table B1).

Table 8. Comparison of the Three- and Four-Factor Solutions

Criterion	Three-factor solution	Four-factor solution
Retention criteria in favor	Parallel analysis; minimum average partial test	Kaiser criterion
Variance explained after extraction	44.7%	47.4%
Items with primary loading < .40	2 (FR4, BD2)	3 (FR4, BD2, BD3)
Items with a cross-loading ≥ .30	5 (FR2, ED2, ED3, ED4, BD2)	2 (ED1, LA1)
Smallest / mean communality	.26 / .45	.31 / .47
Highest factor correlation	.49	.59
Alpha of the factors	.82–.87	.73–.85
Interpretability	Autonomy merged with behavioral dependence; emotional items split	Autonomy separate; fourth factor marginal

Note. Alpha was computed for each factor using the items with their highest loading on it. FR4 had a loading of .38 (three-factor) and .397 (four-factor) on its primary factor.

Neither solution is clearly superior, and the choice between them rests on both empirical and theoretical considerations (full pattern matrices for both solutions are in Appendix B). Empirically, the four-factor solution explains more variance (47.4% vs. 44.7%), has fewer cross-loading items (2 vs. 5), and a higher smallest communality (.31 vs. .26), but the gain in variance is modest (2.7 percentage points) and does not by itself establish that the fourth factor exists. That factor also rests on a marginal eigenvalue (1.07, below the parallel-analysis threshold of 1.31). The three-factor solution is supported by parallel analysis and the minimum average partial test, which are generally the more dependable retention criteria, and its factors have higher alphas (.82–.87), but it merges the autonomy items with behavioral dependence and splits the emotional items across two factors. Theoretically, perceived autonomy was proposed as a consequence of dependence and not as a component of it, and only the four-factor solution keeps that distinction. The four-factor solution is therefore retained as a provisional exploratory model that requires confirmation in an independent sample. It is preferred on interpretive grounds and not because the statistical criteria favor it, and the three-factor solution remains an equally plausible alternative. Two sensitivity checks broadly reproduced the four-factor pattern, although some individual loadings and the primary assignment of ED1 were sensitive to outlier exclusion. When the three outlying cases were excluded, no primary assignment changed, but when the 18 cases flagged on all 24 items were excluded, ED1 changed primary factor and some loadings shifted by up to .11 (see Data screening and descriptive characteristics). When the analysis was repeated on a polychoric correlation matrix, the four factors were reproduced with congruence coefficients of .99 to 1.00 against the Pearson-based factors, and 22 of the 24 items kept the same primary factor; the exceptions were BD2 and BD3, the two weakest items. These checks show that the overall pattern does not depend on a few outlying cases or on treating the items as continuous, but individual item assignments, especially for ED1, BD2, and BD3, are less stable, and the checks cannot show that a fourth factor exists in the population.

Items flagged for review (RQ3)

Several items warrant review (Tables 4 and 6; item wording in Appendix A). BD2 (rarely writing in English without first consulting AI suggestions; primary loading .31, communality .38) and BD3 (waiting for AI input before finishing a text; .32, with .29 on Factor 1; communality .44) did not reach the .40 criterion on any factor, and they were the only two items that changed primary factor in the polychoric analysis. The content of BD2 is close to that of F4 (immediately opening an AI tool whenever writing in English), which may explain why it does not define a separate behavioral-dependence factor; this is an interpretation from item content and not a tested result. ED1 (feeling more confident when writing with AI tools) loaded .42 on Habitual Use with a cross-loading of .33 on Emotional–Behavioral Dependence, and its wording describes writing confidence rather than emotional dependence; it was also the only item whose primary factor changed when the 18 cases flagged on all 24 items were excluded. FR4 fell marginally below the criterion (.397; communality .38) and had the lowest corrected item-total correlation in its proposed dimension (.48). LA1 cross-loaded (.31 on Factor 3), and LA4 had a low communality (.32) and the lowest corrected item-total correlation in its dimension (.47). BD4 had the lowest communality (.31). Because the factor structure itself is unresolved, these items should be revised or replaced and the revised pool retested, rather than deleted on the basis of this sample alone. In this study, BD2 and BD3 are shown in the pattern matrices but were not used to define the Emotional–Behavioral Dependence factor or its reliability estimate, whereas FR4 was kept in Instrumental Reliance and Educational Value because its loading was only marginally below .40 (.397 in the four-factor and .38 in the three-factor solution).

Reliability (RQ2)

Alpha for the six proposed dimensions ranged from .68 (BD) to .82 (ED): .81 for F, .78 for FR, .73 for LA, and .80 for EV (Table 9). Alpha was thus acceptable (≥ .70) for five of the six proposed dimensions, and BD fell slightly below the .70 benchmark specified in the Methods. Alpha based on standardized items was virtually identical (.68–.82). For the dimensions from the EFA, alpha was .83 for Habitual Use (F1–F4 and ED1), .85 for Instrumental Reliance and Educational Value, .81 for Emotional–Behavioral Dependence (ED2–ED4, BD1, BD4), and .73 for Reduced Learner Autonomy. Adding the weakly loading items BD2 and BD3 to Factor 3 gave an alpha of .83 for seven items, and combining all eight ED and BD items gave .85, so these items still contribute to a coherent dependence factor even though their loadings were low. The 24-item total scale had an alpha of .91, with corrected item-total correlations from .34 to .66, which does not show that the 24 items measure a single construct (Cortina, 1993). McDonald's omega closely matched alpha: ω = .68–.82 for the six proposed dimensions (BD = .68, ED = .82), .74–.86 for the four empirical factors, and .93 for the total scale (.91 when a single general factor was assumed).

Table 9. Descriptive Statistics and Cronbach's Alpha of the Proposed and Empirical Dimensions

Scale	Items	*M*	*SD*	*α*	CITC range
Six proposed dimensions					
Frequency of Use (F)	4	2.75	0.79	.81	.58–.68
Functional Reliance (FR)	4	3.38	0.78	.78	.48–.64
Emotional Dependence (ED)	4	2.54	0.83	.82	.59–.68
Behavioral Dependence (BD)	4	2.48	0.69	.68	.43–.54
Reduction of Learner Autonomy (LA)	4	2.62	0.77	.73	.47–.60
Perceived Educational Value (EV)	4	3.66	0.71	.80	.59–.66
Dimensions from the exploratory factor analysis					
Habitual Use	5	2.77	0.78	.83	.59–.69
Instrumental Reliance and Educational Value	8	3.52	0.67	.85	.51–.66
Emotional–Behavioral Dependence	5	2.39	0.74	.81	.50–.65
Reduced Learner Autonomy	4	2.62	0.77	.73	.47–.60
Total scale					
All 24 items	24	2.90	0.57	.91	.34–.66

Note. N = 453. M and SD refer to the mean of each respondent's item scores. CITC = corrected item-total correlation within the scale. Empirical-factor scales use the items assigned to each factor by a primary loading of at least .40 (BD2 and BD3 excluded; FR4 retained).

Content validity assessment

The supplementary content validity assessment examined the relevance of the 24 items to their proposed dimensions (Table 10). Aiken's *V* ranged from .42 (BD3) to 1.00 (FR1, EV1, and EV4), with a mean of .77 (*SD* = .17). Ten items (41.7%) reached the benchmark of $V \geq .80$ (F1, F2, FR1, FR2, FR3, ED2, ED4, LA2, EV1, and EV4), and 14 items (58.3%) fell below it. With three experts, the exact one-tailed test reaches $p \leq .05$ only when $V \geq .92$, and eight items met this condition (F1, F2, FR1, FR2, FR3, ED4, EV1, and EV4). ED2 and LA2 ($V = .83$, $p = .08$) met the benchmark but not the significance criterion.

Table 10. Expert Ratings and Aiken's V for the 24 Items

Item	Expert 1	Expert 2	Expert 3	Σ(r − 1)	*V*	*p*	V ≥ .80
Frequency of Use (mean V = .83)							
F1	5	5	4	11	.92	.032	Yes
F2	5	4	5	11	.92	.032	Yes
F3	4	4	4	9	.75	.160	No
F4	4	4	4	9	.75	.160	No
Functional Reliance (mean V = .90)							
FR1	5	5	5	12	1.00	.008	Yes
FR2	5	4	5	11	.92	.032	Yes
FR3	5	5	4	11	.92	.032	Yes
FR4	4	4	4	9	.75	.160	No
Emotional Dependence (mean V = .75)							
ED1	3	3	3	6	.50	.576	No
ED2	5	4	4	10	.83	.080	Yes
ED3	4	4	4	9	.75	.160	No
ED4	5	4	5	11	.92	.032	Yes
Behavioral Dependence (mean V = .58)							
BD1	4	4	4	9	.75	.160	No
BD2	3	3	3	6	.50	.576	No
BD3	3	2	3	5	.42	.720	No
BD4	4	3	4	8	.67	.280	No
Reduction of Learner Autonomy (mean V = .71)							
LA1	4	4	4	9	.75	.160	No
LA2	4	4	5	10	.83	.080	Yes
LA3	4	4	4	9	.75	.160	No
LA4	3	3	3	6	.50	.576	No
Perceived Educational Value (mean V = .88)							
EV1	5	5	5	12	1.00	.008	Yes
EV2	4	4	4	9	.75	.160	No
EV3	4	4	4	9	.75	.160	No
EV4	5	5	5	12	1.00	.008	Yes

Note. Three experts rated each item from 1 (not relevant) to 5 (highly relevant). Σ(r − 1) is the sum of the three ratings minus 3; V = Σ(r − 1)/12 (Aiken, 1985). p is the exact one-tailed probability of the observed sum or a larger one under a uniform null distribution for three raters and five categories. Mean V = .77 for all 24 items. Item wording is given in Appendix A.

At the dimension level, mean *V* was highest for Functional Reliance (.90) and Perceived Educational Value (.88), followed by Frequency of Use (.83), Emotional Dependence (.75), and Reduction of Learner Autonomy (.71). Behavioral Dependence had the lowest mean (.58), and none of its four items reached .80 (range .42–.75). The five lowest coefficients were those of BD3 (.42), ED1, BD2, and LA4 (each .50), and BD4 (.67). All five were among the items flagged for review in the factor analysis (BD2 and BD3 with primary loadings below .40, ED1 with a cross-loading, BD4 with the lowest communality, and LA4 with a low communality). This correspondence is descriptive and was not tested, and FR4 and LA1, which were also flagged, had *V* = .75, the same value as seven items that were not flagged.

The written comments pointed to four recurring concerns. The first was construct boundaries. All three experts noted that confidence in writing with AI (ED1) is not necessarily emotional dependence and suggested reframing it toward distress without AI, and they noted that BD2 overlaps with the habitual-use item F4 and should be reworded toward an inability to start writing independently. The second was that several behavioral items could reflect preference or openness and not impaired control: waiting for AI input (BD3), changing one's ideas after AI responses (BD4), and copying suggestions without changes (BD1) were each said to need clearer reference to loss of control, and one expert judged that BD3 might need replacement and not only editing. The third concerned self-report and causal wording: LA4 was described as broad and causally loaded, EV3 as showing perceived improvement without establishing actual skill gains, and EV2 as mixing purpose with behavior. The fourth was vague or subjective wording, for example "almost always" (F1), "part of my routine" (F2), "simple writing task" (F3), "rely on" (FR4), and "writing correctly" (ED3), and the difficulty of judging how often one practices without AI relative to using it (LA3). Comments on the highest-rated items (FR1, EV1, and EV4) described them as clear and appropriate to their dimensions. The ratings and comments were not used to change the items analyzed here; they inform the recommendations for item revision below.

DISCUSSION

The study asked whether the six proposed dimensions of the Generative AI Use and Dependence Scale emerge from students' responses, how reliable the resulting dimensions are, and how the items behave. The theoretically proposed six-dimension model was not recovered. The Kaiser criterion supported four correlated factors and parallel analysis three, and both solutions are treated here as plausible. The four-factor solution matches the proposed structure only in part: frequency of use and learner autonomy each stood apart, whereas functional reliance merged with perceived educational value, and emotional dependence merged with behavioral dependence. All four empirical factors reached acceptable internal consistency (α = .73–.85), as did five of the six proposed dimensions; only the proposed Behavioral Dependence dimension (.68) fell slightly below the .70 benchmark. The data were well suited to factoring. A supplementary expert review of item relevance gave partial support (mean Aiken's *V* = .77, with 10 of the 24 items at or above .80), and the lowest-rated items were largely those that behaved weakly in the factor analysis. These results are preliminary evidence of structure, consistency, and content relevance, not a validation of the scale.

Habitual Use separated from the dependence items, which fits the argument that how often a tool is used is not the same as being dependent on it (Griffiths, 2005). Wu et al. (2025) likewise found only a moderate association (*r* = .37) between AI dependence and frequency of use. In the present data, Habitual Use and Emotional–Behavioral Dependence correlated .59, the strongest link among the factors, so the two are related but distinguishable. Because the Habitual Use items are agreement statements about routinized use, this factor reflects perceived habit, and its relation to actual frequency of use has not been tested. ED1 sat between Habitual Use and dependence, loading on both. Its wording describes feeling more confident when writing with AI tools, which is closer to a positive appraisal of the tool than to emotional dependence, so its position may partly reflect how often students use AI tools. More generally, this suggests that not every positive emotional response to AI use indicates dependence.

The merger of Functional Reliance and Perceived Educational Value suggests a shared positive, instrumental appraisal: students who find AI tools useful for study also report relying on them for study tasks. This is consistent with technology acceptance research, in which perceived usefulness underlies use (Davis et al., 1989; Shahzad et al., 2024). Item content may also contribute. Several Functional Reliance items and one value item describe what the respondent does with the tools ("I use…", "I consult…"), and other value items describe perceived skill improvement (EV3) rather than a separate evaluative judgment, so the factor may partly reflect a shared "purposeful use" component and not two substantively merged constructs. The factor correlated only weakly with Reduced Learner Autonomy (.12) and modestly with Emotional–Behavioral Dependence (.26), so instrumental reliance in this scale did not go together with signs of lost autonomy. The pattern may be consistent with a distinction between potentially productive reliance and problematic dependence; however, external learning

outcomes and behavioral measures are needed to establish this distinction, because the study measured no writing achievement, independent performance, or other learning outcome and therefore cannot show whether particular AI practices are beneficial or harmful. Because the factor combines two conceptually different constructs, it should not be read as a measure of dependence. Wu et al. (2025) , by contrast, treated functional dependence as part of AI dependence. Whether the functional reliance items here capture benign reliance or dependence in that sense cannot be settled by an exploratory analysis; criterion evidence, such as links with academic self-efficacy, learner autonomy, or established problematic-use measures (Wu et al., 2025; X. Zhang et al., 2025), is needed.

Emotional and behavioral dependence formed one factor, in line with behavioral-addiction frameworks that treat affective and behavioral indicators as parts of one pattern (Andreassen et al., 2012; Griffiths, 2005). Research on conversational AI also links problematic use to emotional states such as social anxiety and loneliness (Hu et al., 2023), so affective and behavioral indicators may co-occur in students. The merger can also be read as a discriminant-validity problem. When six factors were forced, the emotional and behavioral items still merged, and the eight items together reached an alpha of .85, which suggests that respondents did not distinguish the two sets empirically. Possible reasons include overlap in what the indicators capture (a general reluctance to write without AI input), low variance and positive skew in these items (means 2.22–2.87), and the weak wording of some behavioral items. The behavioral dependence items were the weakest part of the scale: BD had the lowest alpha (.68), and BD2 and BD3 did not reach a primary loading of .40. These two items, and probably ED1, should be reviewed in wording and content before the scale is used further (see Table 6), a conclusion that the expert ratings also support (V = .42 for BD3 and .50 for BD2 and ED1).

Reduced Learner Autonomy emerged as a distinct factor that correlated .43 with dependence. Concerns that delegating thinking to AI may be associated with weaker independent reasoning (Grinschgl et al., 2021; Zhai et al., 2024) motivate this dimension, but its scores reflect students' own perceptions of reduced motivation, confidence, and independent practice. They should not be read as evidence that critical thinking or writing ability has actually declined, particularly because the data are cross-sectional and self-reported. LA4, which asks about a decline in critical thinking, is itself a self-perception item, and it was among the weakest items in its dimension. A moderate rather than very high correlation supports treating autonomy as separate from dependence, but this factor is the one that disappears in the three-factor solution and is therefore the least secure. Its mean (2.62) was below the scale midpoint, as were the means for Habitual Use and dependence, while the means for instrumental reliance and educational value were above it. The average respondent thus reports high perceived value together with low to moderate dependence. However, self-reported autonomy may not fully reflect students' actual capacity to complete writing tasks independently, which should be examined using performance-based measures in future studies.

Reliability was acceptable for all four empirical factors (α = .73 to .85) and for five of the six proposed dimensions; only Behavioral Dependence (.68) fell below .70. The total-scale alpha of .91 is high, but alpha rises with the number of items and does not show that a scale is unidimensional (Cortina, 1993; Tavakol & Dennick, 2011). The 24 items also combine dependence indicators with items that reflect a positive appraisal, so summing them into one score would blur different constructs, and the total score should not be interpreted as a single dependence measure. Scores are better interpreted at the factor level until the structure is confirmed (Boateng et al., 2018). Omega differed from alpha by no more than .01 for any dimension or factor, so the equal-loadings assumption of alpha does not appear to have distorted the reliability estimates here.

Factor retention is the main uncertainty. The criteria disagree: the Kaiser criterion supports four factors, parallel analysis and the minimum average partial test support two to three, and parallel analysis on the reduced matrix favors more. Because parallel analysis is generally regarded as more dependable than the Kaiser criterion (Watkins, 2018), the three-factor solution is a serious alternative, and the fourth factor, Reduced Learner Autonomy, cannot be regarded as established. The four-factor solution is kept here as a working structure for three interpretive reasons: it explains more variance, it has fewer cross-loading items, and it avoids merging autonomy items with dependence indicators although the two are conceptually distinct. These reasons are not statistical proof. The four-factor pattern was reproduced after excluding outliers and with polychoric correlations, which speaks to numerical stability but not to whether the fourth factor exists in the population. A confirmatory comparison of the three-, four-, and six-factor models in an independent sample is the necessary next step.

The supplementary expert review adds partial, item-level evidence on content relevance, and it points in the same direction as several findings above. Experts rated the Functional Reliance, Perceived Educational Value, and Frequency of Use items most favorably (mean V = .90, .88, and .83), and these are the items that formed the two clearest factors. The Behavioral Dependence items were rated lowest (mean V = .58, with no item at .80), and the comments questioned whether they separate impaired control from workflow preference or from habitual use. This agrees with the factor analysis, in which BD2 and BD3 did not reach a primary loading of .40, Behavioral Dependence had the lowest alpha (.68), and the behavioral items merged with emotional dependence. The same convergence appears for ED1 (V = .50), where both the experts and the loading pattern indicate that the item taps writing confidence and not dependence, and for LA4 (V = .50), whose causally loaded wording all three experts

questioned and which had a low communality. The correspondence is not complete, however. The Frequency of Use items, two of which (F3 and F4) had $V = .75$, loaded .53–.81 on one factor, so a coefficient below .80 does not by itself mark an item as defective. Two readings of the convergence are possible, and the present data cannot separate them. Weak items may have degraded the empirical structure, or the boundaries between the constructs (emotional versus behavioral dependence, and habitual use versus behavioral dependence) may themselves be insufficiently sharp, as the comments on overlap suggest. The review therefore tempers the failure to recover six dimensions: part of the merger of emotional and behavioral dependence may reflect item wording and not theory alone, which can only be tested with revised items. The comments also give concrete directions for revision, such as reframing ED1 toward distress without AI, rewording BD2 and BD3 toward an inability to start or proceed, replacing the causal wording of LA4 with a concrete writing decision, and adding a reference period or operational definition to the Frequency of Use items and to FR4.

Sample characteristics and item formulation may also have shaped the structure. All respondents had used AI tools for English writing, most came from one region, and means on the dependence items were low with positive skew, so restricted variance in these items could weaken their loadings and reliability, especially for Behavioral Dependence. In addition, every item names the same set of tools, all items are keyed in the same direction, and several share an "I use… to…" structure, which can inflate correlations through shared method variance and acquiescence and may favor broader factors. Because gender, age, and year of study were not recorded, it is not possible to say whether the structure differs across student groups.

Theoretically, the findings suggest that AI dependence in English writing is more compact than the six proposed dimensions. Use and perceived value can be kept apart from dependence and autonomy indicators, but emotional and behavioral dependence were not empirically separable in this sample, and the boundary between adaptive reliance and problematic dependence is not located by these data. Instrumental reliance accompanied by a positive learning appraisal appeared as a separate and comparatively high-scoring factor, whereas the dependence and autonomy factors scored low to moderate; whether a given level of reliance is adaptive or maladaptive will depend on learning outcomes and on the learner's retained control, which this study did not measure. Practically, the scale may be used in research to describe students who use AI tools often and value them and students who also report emotional–behavioral dependence and reduced autonomy. High frequency of use alone should not be read as dependence. No cut-off scores have been established, factor-level scores are provisional, and until construct, criterion-related, and cross-sample validity evidence is available the scale should not be used to diagnose, screen, or make decisions about individual students.

Two limitations are fundamental. The first concerns content validity. The items were written by the authors from the literature. A supplementary expert review of item relevance was added after data collection, but it used only three experts, it scored relevance and not clarity or representativeness, it led to no revision of the items, and the study included no pilot testing or cognitive interviews with students. Fourteen of the 24 items did not reach $V \geq .80$, and with three experts the coefficient takes only a few coarse values ($V \geq .92$ is needed for significance at $\alpha = .05$), so agreement may also reflect a shared perspective among a small panel. The alignment between the items and the constructs they are meant to represent has therefore been examined only partially and after the fact. The factor structure provides only indirect evidence on this point, and several findings may partly reflect it: the merger of Functional Reliance with Educational Value, the overlap of emotional and behavioral dependence, the confidence wording of ED1, and the weak loadings of BD2 and BD3 may indicate items that respondents did not read as intended, and not only problems of theory. Items may also be ambiguous or tap more than one idea, and the present data cannot separate these possibilities. Before stronger claims about validity are made, the next stage of development should revise the weak items in line with the expert comments and then submit the revised pool to a larger expert panel for judgment of relevance, clarity, and representativeness (for example, content validity indices; Aiken, 1985; Lawshe, 1975; Polit & Beck, 2006), followed by cognitive interviews with students and a pilot test of the revised items. The second limitation concerns representativeness. Respondents were self-selected, and 68.0% came from four institution groups in Pontianak, so students with more experience of, or stronger views about, AI tools may be over-represented, and the results cannot be generalized to Indonesian university students as a whole. Gender, age, and year of study were not recorded, so the sample cannot be compared with student populations, differences across groups cannot be examined, and measurement invariance cannot be tested, for example across gender, where differences in AI chatbot use have been reported (Møgelvang et al., 2024). Larger and more varied samples, preferably drawn with quota or stratified sampling by region, institution type, and study program, are needed for confirmatory testing.

Further limitations apply. Frequency of Use was measured through agreement statements about habitual use and not through direct behavioral indicators such as sessions per week or time spent, so the relation between the Habitual Use factor and actual use is unknown. The original outlier screen used five composite scores; a supplementary screen on all 24 items flagged 18 cases, and excluding them changed the primary factor of ED1 and some loadings by up to .11, so individual loadings, particularly for ED1, FR4, BD2, and BD3, should be regarded as sample-specific. Only exploratory analysis was done on a single sample, and no convergent,

discriminant, or criterion validity evidence was gathered, so construct validity remains to be established (Cronbach & Meehl, 1955). Self-report data are open to response bias, and the cross-sectional design means that relations are associational. Pearson correlations were used for the main analysis, with a polychoric sensitivity check that gave a similar pattern. Future studies should repeat expert content validation of the revised items with a larger panel, conduct cognitive interviews, and pilot test the revised item pool, and should confirm and compare the three-, four, and six-factor models with confirmatory factor analysis in independent samples, assess convergent and discriminant validity, revise the behavioral dependence and other weak items, add direct behavioral frequency indicators, test invariance across groups, examine test–retest reliability, and relate factor scores to outcomes such as academic self-efficacy.

CONCLUSION

This study explored a 24-item Generative AI Use and Dependence Scale for English writing among 453 Indonesian university students. The data were suitable for factor analysis, but the theoretically proposed six-dimensional model was not adequately supported: habitual use and learner autonomy emerged as separate factors, whereas functional reliance merged with perceived educational value and emotional dependence merged with behavioral dependence. The Kaiser criterion supported four correlated factors (Habitual Use, Instrumental Reliance and Educational Value, Emotional–Behavioral Dependence, and Reduced Learner Autonomy), explaining 47.4% of the variance after extraction, whereas parallel analysis supported three. The four-factor solution is therefore a provisional exploratory model that requires confirmation in an independent sample, and the three-factor solution remains an equally plausible alternative. Reliability was acceptable for the four empirical factors (α = .73–.85), lower for Behavioral Dependence among the proposed dimensions (α = .68), and high for the total scale (α = .91), which does not demonstrate unidimensionality. BD2 and BD3 loaded clearly below .40 on every factor, and FR4 marginally so (.397). The main contribution of the study is a preliminary indication that instrumental reliance, habitual use, dependence, and perceived loss of autonomy can be partly separated in students' reports of AI use in English writing, which supports reading dependence against use and perceived value rather than in isolation. The behavior of ED1 also shows that a positive response such as greater writing confidence is not in itself an indicator of dependence. At the same time, the failure to recover six dimensions implies that the proposed distinctions between functional reliance and educational value, and between emotional and behavioral dependence, need to be sharpened in item content, and probably in theory, before the scale is developed further. A supplementary expert review of item relevance gave partial support (mean Aiken's $V = .77$; 10 of 24 items at $V \geq .80$), with the lowest coefficients concentrated in the Behavioral Dependence items (mean $V = .58$) and in ED1 and LA4, items that also performed weakly in the factor analysis. These results are preliminary psychometric evidence and do not validate the instrument.

Five recommendations follow. First, until the structure is confirmed in an independent sample, the scale should be treated as a preliminary research measure and not as a diagnostic or screening tool, and any scoring should be provisional and at the level of the empirical factors rather than as a single total score. No validated scoring procedure exists (for example, the mean of the items assigned to each factor is only a working convention), and BD2 and BD3 need revision before the Emotional–Behavioral Dependence score is used. Second, BD2, BD3, and ED1 should be revised or replaced, and FR4, LA1, and LA4 should be re-examined. Third, content validity should be strengthened before further data collection: BD2, BD3, ED1, and LA4 ($V \leq .50$) should be revised in line with the expert comments, and the revised pool should be rated again by a larger expert panel for relevance, clarity, and representativeness, followed by cognitive interviews and a pilot test. Fourth, confirmatory factor analysis on independent samples should compare the three-, four-, and six-factor models, together with evidence of convergent and discriminant validity from related measures such as academic self-efficacy, test–retest reliability, and demographic data to test invariance. Fifth, a direct behavioral indicator of frequency of use should be added so that habitual use can be checked against actual use.

REFERENCE

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*(1), 10.

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, *45*(1), 131–142.

Andreassen, C. S., Torsheim, T., Brunborg, G. S., & Pallesen, S. (2012). Development of a Facebook Addiction Scale. *Psychological Reports*, *110*(2), 501–517. https://doi.org/10.2466/02.09.18.PR0.110.2.501-517

Baek, C., Tate, T., & Warschauer, M. (2024). "ChatGPT seems too good to be true": College students' use and perceptions of generative AI. *Computers and Education: Artificial Intelligence*, *7*, 100294.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149.

Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, *12*(1), 25–44.

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*(1), 43.

Chen, Y., Wang, M., Yuan, S., & Zhao, Y. (2025). Development and validation of the conversational AI dependence scale for Chinese college students. *Frontiers in Psychology*, *16*, 1621540.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (Vol. 26). Sage publications.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272.

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, *15*(1), 6.

Griffiths, M. (2005). A 'components' model of addiction within a biopsychosocial framework. *Journal of Substance Use*, *10*(4), 191–197.

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477–1496.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2019). *Multivariate data analysis* . *Cengage Learning*. Hampshire, United Kingdom.

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185.

Hu, B., Mao, Y., & Kim, K. J. (2023). How social anxiety leads to problematic use of conversational AI: The roles of loneliness, rumination, and mind perception. *Computers in Human Behavior*, *145*, 107760.

Kasap, S., & Babayiğit, M. V. (2026). Development and Validation of the AI Addiction Scale (AIAS-10): Measuring Compulsive and Emotional Dependence on AI Tools Among EFL Learners. *Register Journal*, *19*(1), 1–27.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. Guilford press.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563.

Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2017). The exploratory factor analysis of items: Guided analysis based on empirical data and software. *Anales de Psicología*, *33*(2), 417–432.

Lorenzo-Seva, U., & Ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64.

Maral, S., Naycı, N., Bilmez, H., Erdemir, E. İ., & Satici, S. A. (2025). Problematic ChatGPT use scale: AI-human collaboration or unraveling the dark side of ChatGPT. *International Journal of Mental Health and Addiction*, 1–27.

Møgelvang, A., Bjelland, C., Grassini, S., & Ludvigsen, K. (2024). Gender differences in the use of generative artificial intelligence chatbots in higher education: Characteristics and consequences. *Education Sciences*, *14*(12), 1363.

Morales-García, W. C., Sairitupa-Sanchez, L. Z., Morales-García, S. B., & Morales-García, M. (2024). Development and validation of a scale for dependence on artificial intelligence in university students. *Frontiers in Education*, *9*, 1323898.

O'connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396–402.

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497.
Qu, Y., Tan, M. X. Y., & Wang, J. (2024). Disciplinary differences in undergraduate students' engagement with generative artificial intelligence. *Smart Learning Environments*, *11*(1), 51.
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*(1), 46.
Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. *Computers and Education: Artificial Intelligence*, *7*, 100259.
Tabachnick, B. G., & Fidell, L. S. (2019). Using multivariate statistics (7th Ed.). *Published Online*.
Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53.
Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209.
Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1), 15.
Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327.
Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219–246.
Wu, H., Ni, H., Luo, W., & Wu, T. (2025). Development and validation of the AI dependence scale for Chinese undergraduates and a preliminary exploration. *Frontiers in Psychology*, *16*, 1725393.
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*(1), 28.
Zhang, S., Zhao, X., Zhou, T., & Kim, J. H. (2024). Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. *International Journal of Educational Technology in Higher Education*, *21*(1), 34.
Zhang, X., Yin, M., Zhang, M., Li, Z., & Li, H. (2025). The development and validation of an artificial intelligence chatbot dependence scale. *Cyberpsychology, Behavior, and Social Networking*, *28*(2), 126–131.

ACKNOWLEDGEMENT

The authors thank all students who took part in the survey for their time and responses.

AUTHOR CONTRIBUTION STATEMENT

VK: conceptualization, methodology, data collection, and writing (original draft). US: formal analysis, data collection, writing. FT: supervision and writing (review and editing). CQ: formal analysis and writing (review and editing). TX: investigation and writing (review and editing). LMO: validation and writing (review and editing). QIA: data curation, formal analysis, and writing (original draft). All authors reviewed and approved the final version of the manuscript.

AI DISCLOSURE STATEMENT

ChatGPT was used solely to check grammar and improve the linguistic accuracy of the manuscript. After using the tool/service, the author thoroughly reviewed and edited the content as needed and takes full responsibility for the content of the publication.

APPENDIX A. ITEMS OF THE SCALE

Items are listed with their proposed dimension. Respondents rated each item from 1 (strongly disagree) to 5 (strongly agree).

Table A1. Item Wording

Code	Dimension	Item wording
F1	Frequency of Use	I almost always use ChatGPT, DeepSeek, Gemini, or other AI tools when writing in English.
F2	Frequency of Use	ChatGPT, DeepSeek, Gemini, or other AI tools have become part of my routine when writing in English.

Code	Dimension	Item wording
F3	Frequency of Use	I use ChatGPT, DeepSeek, Gemini, or other AI tools even for simple English writing tasks.
F4	Frequency of Use	I immediately open ChatGPT, DeepSeek, Gemini, or other AI tools whenever I have to write in English.
FR1	Functional Reliance	I use ChatGPT, DeepSeek, Gemini, or other AI tools to help with grammar and vocabulary when writing in English.
FR2	Functional Reliance	I consult ChatGPT, DeepSeek, Gemini, or other AI tools to structure or organize my writing in English.
FR3	Functional Reliance	I use ChatGPT, DeepSeek, Gemini, or other AI tools when I do not know how to start writing in English.
FR4	Functional Reliance	I rely on ChatGPT, DeepSeek, Gemini, or other AI tools to change or improve my sentences so that they sound more natural.
ED1	Emotional Dependence	I feel more confident writing in English when using ChatGPT, DeepSeek, Gemini, or other AI tools.
ED2	Emotional Dependence	I feel anxious when I have to write in English without assistance from ChatGPT, DeepSeek, Gemini, or other AI tools.
ED3	Emotional Dependence	I trust ChatGPT, DeepSeek, Gemini, or other AI tools more than myself when it comes to writing correctly in English.
ED4	Emotional Dependence	I am afraid of making mistakes if I do not use ChatGPT, DeepSeek, Gemini, or other AI tools.
BD1	Behavioral Dependence	I often copy suggestions from ChatGPT, DeepSeek, Gemini, or other AI tools without making changes.
BD2	Behavioral Dependence	I rarely write in English without first looking at suggestions from ChatGPT, DeepSeek, Gemini, or other AI tools.
BD3	Behavioral Dependence	I wait for input from ChatGPT, DeepSeek, Gemini, or other AI tools before completing my English writing.
BD4	Behavioral Dependence	I often change my ideas based on responses from ChatGPT, DeepSeek, Gemini, or other AI tools, even when I already have my own version.
LA1	Reduction of Learner Autonomy	I feel less motivated to practice writing in English without assistance from ChatGPT, DeepSeek, Gemini, or other AI tools.
LA2	Reduction of Learner Autonomy	Using ChatGPT, DeepSeek, Gemini, or other AI tools makes me less confident in my ability to write independently.
LA3	Reduction of Learner Autonomy	I rely on AI tools such as ChatGPT, DeepSeek, Gemini, or other AI tools more often than I practice my own writing skills.
LA4	Reduction of Learner Autonomy	I feel that my critical thinking skills in English writing have declined because I use ChatGPT, DeepSeek, Gemini, or other AI tools too frequently.
EV1	Perceived Educational Value	ChatGPT, DeepSeek, Gemini, or other AI tools help me learn new English vocabulary and grammar.
EV2	Perceived Educational Value	I use ChatGPT, DeepSeek, Gemini, or other AI tools to improve my writing, not to replace the content of my writing.
EV3	Perceived Educational Value	I feel that my English writing skills have improved since I started using ChatGPT, DeepSeek, Gemini, or other AI tools.
EV4	Perceived Educational Value	ChatGPT, DeepSeek, Gemini, or other AI tools provide useful feedback that I can learn from.

APPENDIX B. COMPLETE PATTERN MATRICES OF THE THREE- AND FOUR-FACTOR SOLUTIONS

Principal axis factoring with Oblimin rotation; every loading is shown. The four-factor solution is the provisional exploratory model discussed in the text; the three-factor solution is the alternative supported by parallel analysis.

Table B1. Pattern Matrices and Communalities of the Three- and Four-Factor Solutions

	Three-factor solution				Four-factor solution				
Item	**1**	**2**	**3**	h^2	**1**	**2**	**3**	**4**	h^2
F1	**.63**	.17	−.05	.48	**.79**	.08	−.12	.03	.60
F2	**.69**	.14	−.07	.53	**.81**	.05	−.06	−.02	.62
F3	**.62**	−.06	.13	.44	**.56**	−.10	.18	.06	.46
F4	**.56**	.00	.21	.47	**.53**	−.04	.18	.14	.49

FR1	.21	**.63**	−.06	.51	.13	**.63**	.09	−.08	.51
FR2	.34	**.51**	−.12	.47	.16	**.53**	.22	−.21	.49
FR3	.20	**.53**	.02	.41	.01	**.56**	.27	−.10	.44
FR4†	.26	**.38**	.12	.36	.10	**.397**	.29	−.01	.38
ED1	**.60**	.15	.10	.53	**.42**	.14	.33	−.03	.53
ED2	**.50**	−.04	.38	.56	.18	−.01	**.63**	.07	.60
ED3	**.40**	.00	.35	.43	.02	.04	**.71**	.00	.53
ED4	.39	.05	**.40**	.49	.08	.08	**.61**	.10	.54
BD1	.22	−.12	**.49**	.37	.07	−.11	**.43**	.25	.37
BD2†	.30	.06	**.39**	.38	.20	.06	**.31**	.23	.38
BD3†	**.45**	.19	.19	.44	.29	.19	**.32**	.05	.44
BD4	.21	.00	**.41**	.29	−.01	.03	**.47**	.16	.31
LA1	.14	.01	**.60**	.47	.10	.01	.31	**.43**	.47
LA2	−.16	.04	**.74**	.46	−.03	.02	.06	**.71**	.55
LA3	.14	−.01	**.58**	.43	.11	−.01	.29	**.41**	.43
LA4	−.09	.11	**.52**	.26	.06	.09	−.02	**.53**	.32
EV1	−.04	**.72**	.05	.52	.05	**.71**	−.10	.11	.52
EV2	−.20	**.75**	.06	.50	−.08	**.74**	−.16	.14	.50
EV3	.15	**.62**	−.06	.45	.21	**.59**	−.09	.01	.46
EV4	−.09	**.69**	.09	.46	−.10	**.70**	.04	.06	.46
% var.		44.7					47.4		

Note. N = 453. Bold = highest loading in the row; † = primary loading below .40 in at least one solution (three-factor: FR4 .38, BD2 .39; four-factor: FR4 .397, BD2 .31, BD3 .32). h² = communality after extraction. Three-factor solution (by primary loading): Factor 1 = F1–F4, ED1–ED3, and BD3; Factor 2 = FR1–FR4 and EV1–EV4; Factor 3 = ED4, BD1, BD2, BD4, and LA1–LA4 (ED4 loads .39 on Factor 1 and .40 on Factor 3). Factor correlations: .39 (1–2), .49 (1–3), .21 (2–3). Four-factor solution: Factor 1 = Habitual Use, 2 = Instrumental Reliance and Educational Value, 3 = Emotional–Behavioral Dependence, 4 = Reduced Learner Autonomy; factor correlations: .43 (1–2), .59 (1–3), .31 (1–4), .26 (2–3), .12 (2–4), .43 (3–4). In the text, BD2 and BD3 are not used to define Factor 3 of the four-factor solution.

***Venny Karolina (Corresponding Author)**
Department of Educational Technology, Faculty of Teacher Training and Education, Universitas Tanjungpura,
Jl. Profesor Dr. H. Hadari Nawawi, Pontianak, West Kalimantan 78124, Indonesia
Email: vennykarolina@fkip.untan.ac.id

Urai Salam
Department of Educational Technology, Faculty of Teacher Training and Education, Universitas Tanjungpura,
Jl. Profesor Dr. H. Hadari Nawawi, Pontianak, West Kalimantan 78124, Indonesia
Email: urai.salam@untan.ac.id

Farina Tazijan
Akademi Pengajian Bahasa, Universiti Teknologi MARA,
40450 Shah Alam, Selangor, Malaysia
Email: farina762@uitm.edu.my

Carla Queiroz
Instituto Superior de Administração e Finanças (ISAF), Academia BAI,
Talatona, Luanda, Angola
Email: carla.queiroz@isaf.co.ao

Toby Xaixanith
Faculty of Education, National University of Laos,
Dongdok Campus, Vientiane, Laos
Email: toby.x@nuol.edu.la

Lindsay M. Ozarraga
Department of Humanities, Education, and Social Sciences,
Mindanao State University–Maigo College of Education, Science and Technology,
Maigo, Lanao del Norte, Philippines
Email: lindsay.ozarraga@msumain.edu.ph

Qonita Indah Arrusli
Department of Educational Technology, Faculty of Teacher Training and Education,
Universitas Tanjungpura,
Jl. Profesor Dr. H. Hadari Nawawi, Pontianak, West Kalimantan 78124, Indonesia
Email: qonitaarrusli@gmail.com