Jurnal Pendidikan Progresif
e-ISSN: 2550-1313 | p-ISSN: 2087-9849
URL: https://jpp.fkip.unila.ac.id/index.php/jpp
Jounal's DOI: http://doi.org/10.23960/jpp

Vol. 16, No. 2, 2026
Article ID: 1653
Pages: 747-772

Development of a Multidimensional Psychometric Scale for Assessing Artificial Intelligence Dependency in Higher Education Task Completion

Wicha Amalia Putri[1], Aunurrahman[1,*], Venny Karolina[1], Halida[1], Afandi[1], & Farina Tazijan[2]
[1]Department of Educational Technology, Tanjungpura University, Indonesia
[2]Academy of Language Studies, Universiti Teknologi MARA, Malaysia

*Corresponding email: aunurrahman@fkip.untan.ac.id

Received: 07 February 2026 *Accepted: 09 April 2026* *Published: 24 April 2026*

Abstract: Recent large-scale surveys indicate that the use of generative Artificial Intelligence (AI) in higher education has become nearly ubiquitous. A survey in the United Kingdom reported that 92% of university students use generative AI tools in their studies. In contrast, a national Indonesian survey involving 1,501 respondents found that 86.21% use AI to assist with academic tasks at least once a month. Preliminary institutional data further revealed that 68.6% of students reported using AI in nearly every assignment. Such high prevalence suggests that AI use has shifted from occasional assistance to habitual reliance, raising concerns about potential dependency and reduced independent cognitive engagement. Although existing instruments, such as the Cognitive AI Dependence and Interaction Scale (CAIDS), assess attitudes and general interaction patterns toward AI, they do not specifically measure functional and cognitive dependency in academic task completion. Therefore, this study aimed to develop and validate an AI task completion dependency scale for university students. A psychometric scale development design was employed involving 500 students from several universities in Indonesia who reported using AI to complete academic assignments. Data were analyzed using exploratory factor analysis (EFA), confirmatory factor analysis (CFA), Rasch modeling, and differential item functioning (DIF) analysis based on gender. The findings revealed a stable three-factor structure comprising functional dependency on AI, reflective attitude, and independent use, and regulation and critical evaluation of AI, yielding a final 10-item scale. The results indicated preliminary structural support for the three-factor model, acceptable reliability, strong item functioning based on Rasch analysis, and acceptable gender invariance. Overall, the developed instrument provides a valid and reliable measure of students' dependency on AI in academic task completion and offers a practical tool for evaluating and managing responsible AI use in higher education contexts.

Keywords: academic task completion , AI dependency , indonesian students , rasch model , scale development.

Article's DOI: https://doi.org/10.23960/jpp.v16i2.pp747-772

INTRODUCTION

The development of Artificial Intelligence (AI) technology has brought significant changes across various areas of life, including education. AI is no longer limited to task automation but has evolved into intelligent systems capable of understanding context, generating text, and providing solutions to complex problems (Gruetzemacher & Whittlestone, 2022). The application of AI effectively enhances personalized learning in higher education by adapting to how students learn (Penuela et al., 2025), think, and complete academic tasks (Alifah & Hidayat, 2025). The emergence of text-based Generative AI (GenAI), such as ChatGPT and DeepSeek, represents a concrete example of AI utilization in academic activities, ranging from summarizing materials and writing essays to preparing research reports.

Despite its significant benefits, the use of AI also raises concerns about academic dependency. Excessive use of AI may shift students' thinking patterns from active to passive (Miranda & Yambao, 2025) and reduce reflective and independent problem-solving abilities (Shah & Asad, 2024). Gökbulut (2019) has argued that excessive use of digital technology can lead to addictive behavioral patterns, while Ulfah (2024) has shown that students who frequently use AI for task completion tend to rely more on system outputs than on their own reasoning. Lin & Chen (2024) emphasize that intensive AI usage may trigger dependency, a condition in which individuals lose confidence in completing academic tasks without AI assistance, posing a serious challenge for higher education that emphasizes critical thinking, independence, and creativity.

The impact of AI dependency is particularly evident in the weakening of students' critical thinking, as they tend to rely on system outputs rather than engage in independent reflection and analysis (Wahyuningsih, 2024; Fauzi et al., 2025). Substitutive use of generative AI, where thinking processes are replaced by ready-made answers, may encourage cognitive offloading, namely the transfer of cognitive load from the individual to the system, causing students to skip essential stages of analysis, reflection, and evaluation (Murray, 2024; Karny et al., 2024). This impact not only reduces critical thinking but is also associated with decreased autonomy in learning and self-regulation (Foo & Prihadi, 2021). Empirical studies indicate that AI can support self-regulated learning when used as reflective scaffolding; however, when AI replaces cognitive and metacognitive regulation functions, it is associated with reduced learning autonomy (Lin & Chen, 2024; Zhou et al., 2024).

In addition, excessive dependency on AI also affects students' creativity. Although generative AI can assist in the early stages of idea generation, excessive dependence may lead to the homogenization of ideas and reduced originality in academic work if not balanced with deep exploration and elaboration (Karny et al., 2024; Dwivedi et al., 2023). Students who treat AI as a final solution provider tend to be less engaged in intellectual experimentation and personal perspective development, resulting in shallow and convergent creativity (Zhai et al., 2024). Higher education literature emphasizes that AI brings transformational benefits to learning and research processes but also introduces challenges and ethical issues that must be managed through evaluative approaches and responsible usage policies (Moukhliss et al., 2024). In the long term, this condition may hinder the achievement of graduate competencies requiring innovation, independent thinking, and academic integrity.

The increasing phenomenon of students' AI dependency may indicate a shift in learning patterns in the digital era and is also associated with psychological aspects such as anxiety and a sense of academic meaning (Fawwaz et al., 2022). Murray (2024) states that habitual reliance on AI for thinking and writing strengthens cognitive offloading practices, which reduce critical thinking, independent learning motivation, and cognitive engagement. Karny et al. (2024) also highlight that uncontrolled AI dependency may decrease learning quality and academic integrity. Research by Hakimi et al. (2025) shows that acceptance and trust toward AI-based chatbots positively influence students' academic self-efficacy and their perception of social meaningfulness. However, the increasing integration of AI tools into academic task completion raises concerns about excessive dependence, indicating that AI dependence is not only conceptually relevant to examine but also requires systematic measurement through the development of valid and reliable psychometric instruments.

Several AI dependence measurement tools have been developed, including the CAI Dependence Scale (CAIDS) by Chen et al. (2025) and the Artificial Intelligence Chatbot

Dependence Scale by Zhang et al. (2025). These instruments are primarily grounded in psychological and human–technology interaction frameworks, focusing on users' attachment, emotional responses, usage preferences, and perceived reliance on AI as technological entities. In this sense, they conceptualize dependency as an affective or attitudinal orientation AI systems.

In contrast, the present study adopts an educational and behavioral framework that positions AI dependency as a learning behavior rather than merely a psychological attachment. The construct is rooted in theories of self-regulated learning, cognitive offloading, and academic autonomy, emphasizing how the use of AI may substitute for, reduce, or reshape students' cognitive engagement during task completion. Thus, while previous instruments measure how individuals feel about or relate to AI, the current scale measures how AI is functionally integrated into academic task processes and how this integration affects independent thinking, effort regulation, and creative engagement.

This theoretical distinction shifts the focus from technology-centered psychological dependence to learning-process-centered behavioral dependence within academic contexts. Although these instruments share a common terminology of "dependency," they differ substantially in theoretical grounding and measurement focus. To make this distinction explicit, Table 1 provides a comparative analysis between the present scale and previously developed instruments.

Table 1. Comparative theoretical frameworks underlying AI dependency measurement instruments

Aspect	**CAIDS / Chatbot Dependence Scale** (Chen et al., 2025)	**AI Task Completion Dependency Scale**
Theoretical basis	Psychological attachment, human–AI interaction	Self-regulated learning, cognitive offloading
Unit of analysis	User–technology relationship	Student–task interaction process
Focus of measurement	Emotional attachment, usage preference	Substitution of cognitive effort in academic tasks
Context specificity	General AI use	Academic task completion
Educational implication	Technology acceptance	Learning autonomy and critical thinking

As shown in Table 1, the present instrument diverges from prior measures by situating AI dependency within the domain of academic task execution rather than affective attachment to technology. This positioning allows for a more precise examination of how the use of AI shapes students' cognitive engagement and learning autonomy. Based on this framework, this study develops an AI dependency scale in academic task completion among university students grounded in three factors: functional dependency on AI, reflective attitude and independent use, and regulation and critical evaluation of AI. To further strengthen the conceptual foundation of the proposed dimensions, each factor is explicitly grounded in established theoretical frameworks in educational psychology and human-technology interaction. The dimension of functional dependency on AI aligns with the concept of cognitive offloading, which refers to the reliance on external tools to reduce cognitive effort and optimize task performance (Risko & Gilbert, 2016). In academic contexts, the use of AI to complete assignments reflects a shift in how cognitive processes are distributed between learners and technological systems.

The dimensions of reflective attitude and independent use are closely related to the framework of Self-Regulated Learning (SRL), particularly regarding learners' autonomy, decision-making, and control over learning strategies (Zimmerman, 2002). Students who demonstrate reflective attitudes toward AI use are more likely to regulate their engagement, including choosing when to rely on AI and when to maintain independent effort. Meanwhile, the dimension of regulation and critical evaluation of AI reflects higher-order metacognitive processes, such as monitoring, evaluation, and critical judgment. These processes are consistent with SRL theory, which emphasizes the monitoring and control of cognitive processes (Zimmerman, 2002) and metacognitive awareness (Flavell, 1979). In addition, perspectives from human–computer interaction highlight the importance of active and critical engagement with technological outputs rather than passive acceptance (Shneiderman, 2020). Together, these dimensions illustrate that AI dependency in academic contexts involves not only functional reliance but also reflective and regulatory processes.

These factors are constructed based on the concept of compulsive behavior and dependency in the DSM-5 (American Psychiatric Association, 2013) and on psychological aspects of digital dependency (Gökbulut, 2019; Lin & Chen, 2024), and are integrated with a preliminary mapping of students' AI usage patterns obtained from field survey results. The identified factors were conceptually informed by the broader literature on behavioral dependency and over-reliance in technology use, rather than by clinical diagnostic criteria. While the DSM-5 provides a structured understanding of compulsive and dependent behavioral patterns, this study does not conceptualize AI use as a clinical disorder. Instead, the construct is framed within educational and psychological perspectives on digital over-reliance (Gökbulut, 2019; Lin & Chen, 2024), integrated with a preliminary mapping of students' AI usage patterns based on field survey results.

This approach allows the scale to capture functional and cognitive dependency in academic contexts without pathologizing students' learning behaviors. Unlike clinical dependency frameworks, cognitive learning theories primarily explain how students process information, while socio-cultural theories emphasize contextual interaction. However, neither framework explicitly addresses patterns of over-reliance and reduced autonomous engagement in technology-mediated task completion. Therefore, this study adopts a behavioral dependency perspective situated within educational psychology rather than clinical pathology. To ensure measurement accuracy, the study employs Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) within the Structural Equation Modeling (SEM) framework, and the Rasch Model to test construct validity, reliability, and measurement invariance across gender (Hair et al., 2021; Hairida et al., 2023).

Based on the above rationale, this study aims to develop and validate the AI Task Completion Dependency Scale for university students. The research questions are: (1) what is the factor structure of the AI task completion dependency scale, (2) how is the validity and reliability of the scale, and (3) does the scale achieve measurement invariance across gender. To address these questions, the study proposes the following hypotheses:

H1: The AI task completion dependency Scale will exhibit a stable multidimensional factor structure with acceptable goodness-of-fit indices.

H2: The scale will demonstrate satisfactory construct validity, reflected in significant factor loadings and appropriate model fit.

H3: The scale will demonstrate adequate reliability, including internal consistency and Rasch-based reliability indices.

H4: The scale will achieve measurement invariance across gender, indicating equivalence in measurement between male and female students.

METHOD

Participants

Participants were recruited using quota sampling with relatively balanced gender proportions to support measurement invariance testing (Comrey & Lee, 2013; Rouquette et al., 2011; Byrne, 2012; Kyriazos, 2018). Data were collected online through a Google Form questionnaire, and participation was voluntary and anonymous. A total of 532 university students initially participated in the study. However, the primary inclusion criterion was the active use of AI tools (e.g., ChatGPT and DeepSeek) to complete academic tasks. Therefore, only respondents who reported using AI were included in the main analysis. Based on this criterion, 500 respondents (94%) were retained for further analysis, while 32 respondents (6%) who reported not using AI were excluded.

The study did not differentiate participants by field of study or academic major, as the primary objective was the psychometric validation of the scale rather than a comparative analysis across academic disciplines. The focus was on ensuring that participants had sufficient experience using AI in academic task completion to meaningfully respond to the instrument items. The sample consisted of 305 female students (57.3%) and 227 male students (42.7%). In terms of educational level, 407 students (76.5%) were enrolled in bachelor's degree programs (S1), 18 students (3.4%) in Diploma IV (D4), and 107 students (20.1%) in master's degree programs (S2). The demographic characteristics of participants are presented in Table 2.

Table 2. Sociodemographic characteristics of participants (N = 532)

Variable	Category	Frequency (n)	Percentage (%)
Gender	Female	305	57.3
	Male	227	42.7
Educational Level	Bachelor's Degree (S1)	407	76.5
	Diploma IV (D4)	18	3.4
	Master's Degree (S2)	107	20.1
Use of AI in Task Completion	Yes	500	94.0
	No	32	6.0

For factor analysis purposes, the final sample of 500 respondents was randomly divided into two independent subsamples. The first subsample (n = 200) was used for Exploratory Factor Analysis (EFA), and the second subsample (n = 300) was used for Confirmatory Factor Analysis (CFA). This sample-splitting strategy ensured analytical independence and strengthened the cross-validation of the measurement model.

Research Design and Procedures

This study used a quantitative approach employing a psychometric scale development design to develop and validate the Artificial Intelligence (AI)-Based Task Completion Dependency Scale for university students. The study was conducted through nine stages (Figure 1), beginning with an exploratory survey to map AI use in academic contexts (Vaus, 2013) and continuing through the stages of psychometric scale development for AI dependency in task completion (DeVellis, 2012). In Stage 1, the open-ended questionnaire survey results identified three primary dimensions of AI dependency. Subsequently, in Stage 2, the scale was constructed with 13 items representing these dimensions: (1) functional dependency on AI (6 items), (2) reflective attitude and independent use (4 items), and (3) regulation and critical evaluation of AI (3 items). Each dimension was operationalized through specific behavioral indicators derived from the preliminary qualitative findings. Table 3 presents the detailed mapping between dimensions, indicators, conceptual meanings, and representative items.

Table 3. Operationalization of dimensions, indicators, and representative items of the AI task completion dependency scale

Dimension	Indicator	Conceptual Meaning of Indicator	Representative Item
Functional Dependency on AI	Routine integration of AI in academic tasks	Habitual use of AI in completing assignments	Using AI regularly (Item 1); Using AI whenever there is an assignment (Item 6); AI has become part of my routine (Item 9)
	Perceived efficiency enhancement	Using AI to complete tasks more quickly and efficiently	AI helps complete tasks more quickly (Item 3)
	Reliance for comprehension support	Depending on AI to simplify difficult academic material	AI makes difficult material easier to understand (Item 4)
	General reliance on task execution	Relying on AI-generated outputs in assignment completion	Relying on AI when completing assignments (Item 11)
Reflective Attitude and Independent Use	Preference for independent completion	Choosing to complete tasks without AI assistance	Preferring to complete assignments independently (Item 7)
	Resistance under academic pressure	Avoiding AI use despite deadlines	Not using AI even when facing assignment deadlines (Item 10)
	Skepticism toward AI efficiency	Questioning AI's effectiveness in accelerating work	AI does not help speed up task completion (Item 5)
	Broader technological awareness	Recognizing that development is not solely AI-dependent	Technological development does not depend on AI (Item 12)
Regulation and Critical Evaluation of AI	Deliberate consideration before use	Thinking carefully before deciding to use AI	Thinking carefully before using AI (Item 2)
	Strategic idea development	Using AI to expand ideas rather than replace thinking	AI helps develop ideas more broadly (Item 8)
	Critical review of AI output	Evaluating AI-generated responses before use	Reviewing AI-generated answers (Item 13)

This structure ensured conceptual clarity and alignment between theoretical constructs and empirical item representation prior to psychometric testing. Content validity was then evaluated by seven experts, including psychometric, subject-matter, language, and educational technology experts (Stage 3). A pilot test was conducted with 30 students from a university in West Kalimantan (Stage 4), followed by field testing (Stage 5) involving 532 students who used AI in completing academic tasks across several universities in Indonesia, reflecting the target measurement characteristics in accordance with psychometric scale development principles (DeVellis, 2012).

After data collection, Stage 6 involved descriptive data analysis to assess data quality, including examination of missing data, score distribution, and corrected item–total correlations (Kline, 1998). The respondent sample was then divided for analysis in the EFA and CFA stages. In instrument development, EFA and CFA should ideally be conducted on different samples to ensure analytical independence and avoid capitalization on chance (Sairitupa-Sanchez et al., 2024). Therefore, the total of 500 respondents was randomly divided into two independent subsamples to ensure analytical independence between exploratory and confirmatory stages. The first subsample (n = 200) was used for EFA, yielding a subject-to-item ratio of approximately 15:1 for the 13-item scale, which exceeds the recommended minimum ratios of 5–10 participants per item (Hair et al., 2021; Kyriazos, 2018). The second subsample (n = 300) was used for CFA to independently test the factor structure using Maximum Likelihood (ML) estimation. Methodological guidelines indicate that sample sizes of 200-300 are generally adequate for CFA models with moderate complexity and normally distributed data (Kline, 2023; Wolf et al., 2013). Comrey & Lee (2013) further classify a sample size of 300 as "good" for factor analytic procedures. Given that the proposed model comprised three latent factors and 13 observed indicators, the CFA subsample size met the recommended adequacy criteria for ML-based estimation. This sample-splitting strategy is widely recommended in scale development research to enable cross-validation and reduce the risk of capitalization on chance, thereby strengthening construct validity and generalizability (Lorenzo-Seva, 2022; Worthington & Whittaker, 2006).

EFA was conducted using the Maximum Likelihood extraction method with Varimax rotation after data suitability was verified through the KMO test and Bartlett's Test of Sphericity (Kaiser, 1974). The factor structure obtained from EFA was subsequently tested through CFA within the Structural Equation Modeling (SEM). Construct validity and reliability were further examined using Rasch Model analysis (Stage 7) (Prayoga et al., 2024), followed by measurement invariance testing by gender based on Differential Item Functioning using WINSTEPS version 5.5.3 (Stage 8) (Hairida et al., 2023; Bond & Fox, 2013). The integration of EFA, CFA, and the Rasch Model enabled a comprehensive psychometric evaluation. It ensured the fairness and precision of scale measurement in the context of AI-supported higher education. Stage 9 involved finalization of the scale based on the overall analysis results.

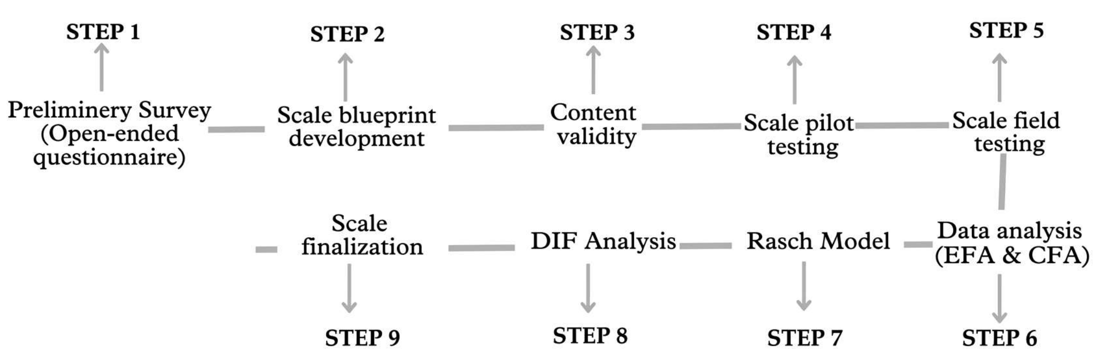

Figure 1. Research flow (Adapted from Vaus (2013) & DeVellis (2012)

Instruments

The AI task completion dependency Scale was developed to measure students' dependency on AI in completing academic tasks. The initial version consisted of 13 items distributed across three dimensions: functional dependency on AI (6 items), reflective attitude and independent use (4 items), and regulation and critical evaluation of AI (3 items).

All items were measured on a four-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree), with no neutral midpoint. This forced-choice format was selected to reduce central tendency bias and encourage clearer attitudinal positioning. Negatively worded items were reverse-scored prior to analysis to ensure consistent interpretation, with higher total scores indicating greater AI dependency. The instrument was designed to capture behavioral reliance, cognitive substitution processes, and self-regulatory control related to AI use in academic contexts.

Content validity was evaluated by seven experts representing psychometrics, educational technology, subject matter, and language to ensure item relevance and clarity. The scale was subsequently pilot-tested with 30 university students to assess comprehensibility and preliminary reliability prior to large-scale administration.

Data Analysis

Data analysis was conducted in several stages to ensure a comprehensive psychometric evaluation. First, descriptive statistics were examined to assess data quality, including the number of missing values, the score distribution, and corrected item–total correlations (Kline, 2023). Exploratory Factor Analysis (EFA) was conducted on the first subsample ($n = 200$) using Maximum Likelihood extraction with Varimax rotation. Sampling adequacy was assessed using the Kaiser–Meyer–Olkin (KMO) measure, with values e" .60 considered acceptable, and Bartlett's Test of Sphericity ($p < 0.05$) to confirm factorability of the correlation matrix (Kaiser, 1974; Hair et al., 2021). Factor loadings of e" .40 were used as the minimum criterion for item retention, as this threshold is commonly recommended for exploratory factor analysis to ensure meaningful associations between items and latent constructs (Hair et al., 2021; Stevens, 2002).

Confirmatory Factor Analysis (CFA) was subsequently conducted on the second subsample ($n = 300$) using Structural Equation Modeling (SEM) with Maximum Likelihood estimation in IBM AMOS version 26. Model fit was evaluated using multiple goodness-of-fit indices: Comparative Fit Index (CFI e" 0.90), Tucker–Lewis Index (TLI e" 0.90), Root Mean Square Error of Approximation (RMSEA d" 0.08), and Standardized Root Mean Square Residual (SRMR d" 0.08) (Hu & Bentler, 1999; Kline, 2023; Hair et al., 2021).

To further examine construct validity and reliability, Rasch Model analysis was conducted using WINSTEPS version 5.5.3. Item fit was evaluated using Infit and Outfit Mean Square (MNSQ) statistics, with values within the acceptable range of 0.5–1.5 (Bond & Fox, 2013; Linacre & Wright, 2000). Person reliability, item reliability, and separation indices were examined to assess measurement precision. Finally, measurement invariance across gender was tested using Differential Item Functioning (DIF) analysis within the Rasch framework. DIF contrast values exceeding ±0.5 logits with statistically significant p-values were considered indicative of potential item bias (Bond & Fox, 2013; Linacre, 2009). The integration of EFA, CFA, and Rasch modeling provided a comprehensive and robust psychometric evaluation of the developed scale.

RESULT AND DISCUSSION

Factor Structure of the Scale

Descriptive Statistics

Table 4 presents the descriptive statistics for the 13 items of the Artificial Intelligence (AI)

Task Completion Dependency Scale. The highest mean was observed for Item 8 (M = 3.36; SD = 0.54), followed by Item 4 (M = 3.28; SD = 0.58) and Item 3 (M = 3.19; SD = 0.56), reflecting students' positive perceptions of the ease and efficiency of using AI in learning. In contrast, the lowest mean scores were observed for Item 13 (M = 1.62; SD = 0.64) and Item 2 (M = 1.79; SD = 0.69), indicating a low tendency to reflect on and verify AI-generated outputs. Skewness values ranged from “0.58 to 0.78, while kurtosis ranged from “0.85 to 1.57. All values fell within the tolerance limit of ±3, indicating that the data distribution was relatively normal and showed no extreme deviations. The presence of response variation across items indicated adequate data quality to proceed with exploratory factor analysis (EFA) (Pérez & Medrano, 2010).

Table 4. Descriptive statistics of the AI task completion dependency scale items (N = 500)

Dimension	Item Statement	Mean	SD	Skewness	Kurtosis
Functional Dependency on AI	Using AI regularly (Item 1)	2.87	0.64	−0.07	−0.16
	AI helps complete tasks more quickly (Item 3)	3.19	0.56	−0.18	0.85
	AI makes difficult material easier to understand (Item 4)	3.28	0.58	−0.32	0.45
	Using AI whenever there is an assignment (Item 6)	2.62	0.68	0.20	−0.37
	AI has become part of my routine (Item 9)	2.71	0.67	−0.01	−0.25
	Relying on AI when completing assignments (Item 11)	2.74	0.65	−0.05	−0.17
Reflective Attitude and Independent Use	AI does not help speed up task completion (Item 5)	2.90	0.65	−0.43	0.68
	Preferring to complete assignments independently (Item 7)	2.70	0.61	−0.58	0.49
	Not using AI even when facing assignment deadlines (Item 10)	2.94	0.56	−0.44	1.57
	Technological development does not depend on AI (Item 12)	2.36	0.69	0.37	0.05
Regulation and Critical Evaluation of AI	Thinking carefully before using AI (Item 2)	1.79	0.69	0.67	0.57
	AI helps develop ideas more broadly (Item 8)	3.36	0.54	−0.02	−0.85
	Reviewing AI-generated answers (Item 13)	1.62	0.64	0.78	0.64

Construct Validity Based on Internal Structure

Construct validity of the scale was examined using Exploratory Factor Analysis (EFA) with the Maximum Likelihood Extraction method and Varimax rotation. The data suitability test showed a KMO value of 0.742, indicating very good sampling adequacy for factor analysis (Kaiser, 1974). Bartlett's Test of Sphericity was also significant ($\div^2 = 572.623$; df = 78; $p < 0.001$), demonstrating sufficient inter-item correlations (Worthington & Whittaker, 2006).

The number of factors was determined based on the Scree Plot and Parallel Analysis (Figure 2), which indicated an elbow point at the third factor, with eigenvalues greater than 1 for the first three factors (3.311, 1.986, and 1.448). The three-factor structure cumulatively explained 38.27% of the total variance. While this proportion of explained variance is moderate, similar values are frequently observed in exploratory analyses of newly developed psychological and educational instruments that measure complex behavioral constructs (Hair et al., 2021; Costello & Osborne, 2005). Constructs related to learning behaviors and technology use are inherently multifaceted and influenced by contextual and individual factors, which may limit the proportion of variance captured at the exploratory stage. Importantly, exploratory factor analysis serves as an initial structure-identification procedure rather than a final validation step. The adequacy of the three-factor model was further supported by confirmatory factor analysis, which demonstrated acceptable fit indices, and by Rasch model analysis, which indicated satisfactory item functioning and measurement precision. Taken together, these findings suggest that the extracted factors meaningfully represent the underlying construct despite the moderate explained variance. Therefore, the three-factor solution was considered the most parsimonious and empirically representative structure for describing the measured construct.

However, the moderate proportion of variance explained (38.27%) also indicates that the current model does not fully capture the complexity of AI dependency. This suggests that additional factors beyond the three identified dimensions may influence students' responses. Future research is therefore encouraged to explore additional relevant variables, such as learning styles, academic motivation, and digital literacy, to enhance the model's explanatory power and comprehensiveness.

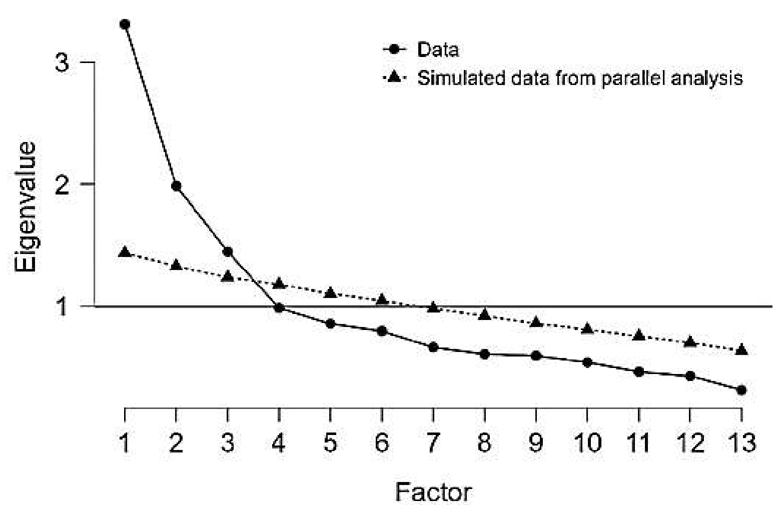

Figure 2. Scree plot and parallel analysis

The exploratory factor analysis (EFA) results indicated that the empirically derived structure consisted of three factors, consistent with the initial conceptual framework used during item development. In addition to examining eigenvalues and the scree plot, Parallel Analysis (Horn, 1965) was conducted to provide a more rigorous basis for factor retention. Parallel Analysis is widely regarded as one of the most accurate methods for determining the number of factors to retain because it compares observed eigenvalues with those generated from random data (Hayton et al., 2004). The results showed that only the first three observed eigenvalues exceeded the corresponding eigenvalues obtained from randomly simulated datasets, whereas the fourth factor fell below the parallel analysis threshold. This finding provides strong empirical support for retaining a three-factor solution.

The convergence of evidence from the scree plot and Parallel Analysis strengthens confidence in the scale's structural validity. The correspondence between the empirical structure and the theoretical framework further indicates that the dimensions formulated during the conceptualization and preliminary survey stages were adequately operationalized into measurable items. In other words, respondents were able to differentiate experiences and behaviors related to AI task-completion dependency along the predefined dimensions. This alignment between theory-driven item construction and empirically derived factor structure is consistent with fundamental principles of psychometric scale development, which emphasize that systematically developed instruments are expected to yield factor structures that meaningfully reflect their underlying theoretical constructs (DeVellis, 2012; Worthington & Whittaker, 2006).

Based on the factor loading pattern, the exploratory factor analysis identified three primary dimensions within the AI task completion dependency scale. The first factor represented functional dependency on AI, characterized by routine, habitual, and integrated reliance on AI in completing academic tasks. Although items such as "AI helps complete tasks more quickly" may appear to reflect functional efficiency rather than problematic use, dependency in behavioral psychology is not defined solely by dysfunction but by patterns of over-reliance and the substitution of independent effort (American Psychiatric Association, 2013; Griffiths, 2005).

In educational contexts, functional reliance may indicate dependency when AI is consistently adopted as the primary or default strategy for task completion, potentially reducing opportunities for autonomous cognitive engagement and self-regulated learning. Research on cognitive offloading suggests that repeated delegation of thinking processes to external tools can reshape cognitive habits and diminish independent problem-solving efforts (Risko & Gilbert, 2016). Therefore, the functional dimension captures not mere efficiency, but the extent to which AI becomes embedded as an indispensable tool in academic task execution, signaling behavioral reliance rather than simple instrumental use.

The second factor reflected a reflective and evaluative dimension of AI use, characterized by students' tendencies to critically assess AI outputs, deliberate before using AI, and employ AI as a supportive cognitive tool rather than as a substitute for independent thinking. This dimension aligns with theoretical perspectives on self-regulated learning, which emphasize metacognitive monitoring and strategic resource management in academic tasks (Zimmerman, 2002). From this perspective, reflective AI use represents an adaptive regulatory mechanism that may buffer against excessive reliance by maintaining cognitive engagement and evaluative control.

The third factor captured independent use and resistance to AI dependence, as indicated by students' preference for completing tasks without AI assistance and their ability to refrain from using AI even under academic pressure. This dimension corresponds to constructs of effort regulation and independent learning, which are central components of self-regulated learning in academic contexts (Pintrich, 2004). The presence of this factor suggests that AI dependency is not merely about frequency of use, but also about the capacity to regulate and limit AI integration when necessary.

Taken together, the three-factor structure suggests that AI task completion dependency is multidimensional, encompassing functional dependency on AI (Factor 1), reflective attitude and independent use (Factor 2), and regulation and critical evaluation of AI (Factor 3). The convergence between the empirically derived factors and the theoretically informed framework indicates preliminary evidence of structural validity. This alignment supports the conceptual grounding of the scale and provides a strong methodological basis for subsequent confirmatory factor analysis to examine model stability and accuracy (Kline, 2023).

After the three-factor structure was empirically identified, item retention decisions were guided by established criteria in exploratory factor analysis. Items were required to demonstrate primary factor loadings of at least |0.30|, consistent with recommendations for early-stage scale development (Costello & Osborne, 2005; Lloret-Segura et al., 2014). In addition, items exhibiting cross-loadings (i.e., secondary loadings e” 0.30) were evaluated based on the difference between primary and secondary loadings to ensure adequate factor interpretability (Hair et al., 2021). The complete rotated factor matrix for the initial 13 items is presented in Table 5.

Table 5. Rotated factor matrix from exploratory factor analysis

Item	Factor 1	Factor 2	Factor 3
Using AI whenever there is an assignment	0.757		
Using AI regularly	0.690		
AI has become part of my routine	0.664		
Relying on AI when completing tasks	0.619		
AI helps complete tasks more quickly	0.411	0.303	
Thinking carefully before using AI		-0.641	
Reviewing AI-generated answers		-0.616	
AI helps develop ideas more broadly		0.595	
AI makes difficult material easier to understand		0.436	
AI does not help speed up task completion			0.577
Not using AI even when facing deadlines			0.575
Preferring to complete assignments independently			0.484
Technological development does not depend on AI	<0.30	<0.30	<0.30

Loadings < 0.30 are not displayed.

As shown in Table 5, most items demonstrated primary loadings above 0.40, indicating satisfactory associations with their respective latent constructs. Although one item (“AI helps complete tasks more quickly”) showed a secondary loading (0.303) on another factor, the difference between its primary and secondary loading exceeded .20, and the item was retained due to its stronger conceptual alignment with Factor 1. Some negative loadings were observed; however, in exploratory factor analysis, the sign of a loading reflects directionality rather than adequacy (Costello & Osborne, 2005). Therefore, item evaluation was based on the absolute magnitude of loadings.

One item (Item 12: “Technological development does not depend on AI”) did not demonstrate a salient loading (|ë| < 0.30) on any extracted factor and therefore failed to meet the minimum retention threshold. Because the item did not meaningfully represent any latent dimension and contributed minimally to factor clarity, it was removed. The EFA was subsequently re-run after item removal to confirm the stability and coherence of the three-factor solution. The refined structure showed improved clarity and interpretability.

Confirmatory Factor Analysis (CFA) and Internal Reliability

Based on the latent structure obtained through EFA, confirmatory factor analysis (CFA) was conducted on an independent subsample to test model fit and cross-validate the factor structure identified at the exploratory stage. The initial confirmatory model comprised 12 items, following the removal of 1 item during EFA due to the absence of salient factor loadings. Model fit was evaluated using multiple goodness-of-fit indices representing absolute, incremental, and residual-based fit (CMIN/DF, CFI, TLI, RMSEA, and SRMR) (Hair et al., 2021; Kline, 2023). The initial three-factor model did not achieve a satisfactory fit ($\div^2$ = 253.783; CMIN/DF = 4.976; RMSEA = 0.115; CFI = 0.857; TLI = 0.815; RMR = 0.037), indicating that certain items may not have optimally represented their respective latent constructs.

Examination of standardized factor loadings revealed that two items (Item 2, ë = 0.22; Item 13, ë = 0.16) exhibited loadings substantially below the recommended threshold of 0.50 for confirmatory models (Hair et al., 2021; Kline, 2023). In scale development research, it is common for broader initial item pools, designed to capture the breadth of the construct, to undergo empirical refinement during confirmatory testing (DeVellis, 2012; Worthington & Whittaker, 2006). Low factor loadings indicate limited shared variance between an item and its intended construct, which may reduce convergent validity and measurement precision if retained (Fabrigar et al., 1999).

Therefore, Items 2 and 13 were removed based on both statistical criteria and conceptual evaluation of their alignment with the intended dimensions. CFA was subsequently re-estimated using the refined item set to assess improvements in model fit and structural coherence. This iterative process reflects recommended best practices in psychometric scale development, where exploratory breadth is followed by confirmatory precision to enhance structural validity (DeVellis, 2012).

Final CFA Model

After eliminating the two items with low factor loadings, the three-factor confirmatory factor analysis model was re-estimated and showed improvements across several fit indices ($\div^2$ = 133.104; CMIN/DF = 4.160; RMSEA = 0.103; CFI = 0.923; TLI = 0.891; RMR = 0.024). Compared with the initial model, the reduction in chi-square and improvements in incremental fit indices indicate that removing poorly performing items enhanced internal consistency and reduced measurement noise (Hair et al., 2021; DeVellis, 2012). A CFI value exceeding 0.90 suggests an acceptable incremental fit relative to the null model, and a low RMR indicates relatively small residual discrepancies between observed and predicted covariances (Hu & Bentler, 1999).

However, the RMSEA value (0.103) exceeds the commonly recommended upper threshold of 0.08 and falls within the range typically interpreted as poor absolute fit (Kline, 2023). The TLI value (0.891), while approaching 0.90, also does not fully meet conventional criteria. These findings indicate that although the model demonstrates reasonable incremental fit, the absolute fit of the proposed three-factor structure remains suboptimal. In practical terms, this suggests that the latent structure, while theoretically grounded, may not yet fully capture the covariance relationships among observed indicators.

To address this limitation, further model refinement was conducted by examining the modification indices (MI) provided by AMOS

(Arbuckle, 2014). Modification indices identify potential improvements in model fit by estimating the decrease in chi-square when certain parameters, particularly correlations between error terms, are estimated freely. Based on both statistical evidence and theoretical considerations, several covariances were introduced between the error terms (residuals) of specific indicators, rather than between the items themselves.

Specifically, correlated error terms were added between e1 and e6 (Item 1–Item 6; MI = 25.912), e3 and e10 (Item 3–Item 10; MI = 12.930), and e5 and e7 (Item 5–Item 7; MI = 9.905), as these pairs showed the highest modification indices among residuals. These modifications indicate that the associated items share variance not fully explained by the latent constructs. The covariance between the error terms of Item 1 and Item 6 is theoretically justified as both items assess habitual, frequent use of AI in academic tasks and thus reflect repetitive behavioral patterns of AI dependency. The covariance between the error terms of Item 3 and Item 10 can be explained by their shared focus on decision-making under time constraints, particularly regarding the perceived efficiency of AI versus deliberate avoidance under deadlines. Meanwhile, the covariance between the error terms of Item 5 and Item 7 reflects a shared skeptical attitude toward AI and a preference for independent task completion. These conceptual overlaps may introduce shared measurement error beyond the latent constructs, thereby justifying the inclusion of correlated error terms (Byrne, 2016; Kline, 2023). Importantly, these correlations were specified in a limited and theoretically grounded manner to preserve model parsimony and avoid overfitting (Brown, 2015).

Following these theoretically justified modifications, the model fit improved substantially. The re-specified model demonstrated good fit indices (÷² = 77.293; CMIN/DF = 2.665; RMSEA = 0.075; CFI = 0.963; TLI = 0.943; RMR = 0.019). The RMSEA value fell below the recommended threshold of 0.08, indicating acceptable absolute fit, while the CFI and TLI values exceeded 0.90, demonstrating strong incremental fit (Hu & Bentler, 1999; Kline, 2023). These results suggest that the refined three-factor model provides an convee representation of the observed data.

Importantly, the evaluation of construct validity in this study was not based solely on CFA results. Converging evidence from exploratory factor analysis, Rasch modeling (demonstrating acceptable item fit and measurement reliability), and gender invariance testing collectively supports the instrument's structural coherence. Although initial CFA results indicated the need for refinement, the final model demonstrates that the proposed three-factor structure is empirically supported after theoretically justified modifications. Therefore, the scale can be considered to have achieved an acceptable level of structural validity, while still allowing room for further refinement in future studies.

Validity and Reliability Evidence

Construct Reliability and Convergent Validity (CR and AVE)

After the measurement model fit was evaluated through CFA, the analysis proceeded at the construct level by examining reliability and convergent validity. Composite reliability (CR e" 0.70) was used as a model-based indicator of internal consistency, and average variance extracted (AVE e" 0.50) was used to measure the proportion of indicator variance explained by the latent construct (Siregar et al., 2021; Fornell & Larcker, 1981). In instrument development, both indices are important because a reliable construct does not necessarily demonstrate adequate convergent validity, and vice versa (DeVellis, 2012).

For Factor 1 (functional dependency on AI), the CR value was 0.889, and the AVE value

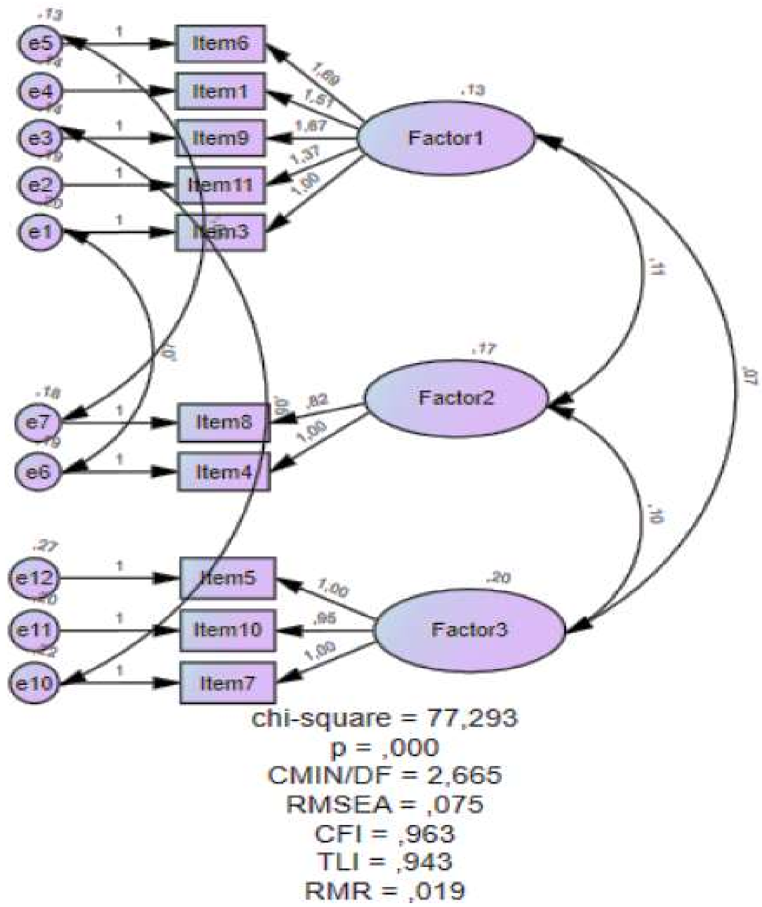

Figure 3. Final three-factor CFA model of AI-based academic task dependency scale (standardized estimates)

was 0.618, indicating very good internal consistency and that the construct explained the majority of indicator variance. This suggests that the construct was strongly and stably operationalized at the latent measurement level. According to Fornell & Larcker (1981), AVE e” 0.50 indicates that more than half of the indicator variance is attributable to the construct rather than to measurement error. Therefore, the first factor can be considered the core dimension of the scale, as it demonstrates adequate reliability and convergent validity. In scale development, such a dimension typically serves as the primary basis for construct interpretation because it exhibits the most consistent empirical representation (DeVellis, 2012).

Factor 2 (reflective attitude and independent use) yielded a Composite Reliability (CR) value of 0.657 and an Average Variance Extracted (AVE) value of 0.491. Although the CR value is slightly below the conventional threshold of 0.70, it remains within an acceptable range for early-stage scale development (Hair et al., 2021). The AVE value is marginally below the recommended cutoff of 0.50, indicating that the construct explains slightly less than half of the variance in its indicators. However, this condition does not necessarily indicate inadequate convergent validity. According to Fornell & Larcker (1981), convergent validity can still be considered acceptable when composite reliability approaches recommended levels. In addition, this dimension reflects attitudinal and reflective processes, which are inherently more heterogeneous and may lead to greater variability across indicators. Therefore, Factor 2 can be interpreted as demonstrating acceptable but not optimal convergent validity, suggesting that the construct is conceptually coherent while still allowing room for further refinement.

Factor 3 (regulation and critical evaluation of AI) demonstrated a Composite Reliability (CR) value of 0.715, exceeding the commonly recommended threshold of 0.70 (Hair et al., 2021), indicating adequate internal consistency. The Average Variance Extracted (AVE) value of 0.456 is slightly below the conventional 0.50 criterion (Fornell & Larcker, 1981), indicating that the construct explains a moderate proportion of the variance in its indicators. Nevertheless, the combination of acceptable composite reliability and a near-threshold AVE indicates that the construct still demonstrates adequate convergent validity for complex behavioral constructs (Kline, 2023). Given that this dimension involves regulatory and evaluative processes, which are inherently multifaceted, some dispersion in indicator variance is expected. Thus, Factor 3 can be interpreted as demonstrating satisfactory but improvable convergent validity.

Although Factors 2 and 3 did not meet the recommended AVE e” 0.50 threshold, their values were close to the cutoff, indicating marginal but acceptable levels of convergent validity. When considered alongside the composite reliability results and the overall model fit obtained from CFA, these findings suggest that the constructs demonstrate adequate preliminary convergent validity. Such variability across dimensions is not uncommon in early-stage scale development and reflects the complexity of the constructs being measured (DeVellis, 2012; Kline, 2023). Therefore, rather than indicating a fundamental limitation, these results highlight areas for targeted refinement in future research, such as improving item wording, enhancing conceptual clarity, or expanding the number of indicators.

Overall, the CR and AVE results indicate that the instrument demonstrates acceptable construct reliability and convergent validity, with Factor 1 showing strong psychometric properties, and Factors 2 and 3 demonstrating adequate but developing measurement quality. The interpretation of validity in this study is therefore cumulative, supported by consistency of the factor structure across EFA and CFA, as well as Rasch model evidence of item functioning and measurement stability. Taken together, the instrument provides sufficient preliminary structural support while allowing opportunities for further refinement to strengthen its psychometric robustness.

Despite the generally acceptable results, the slightly below-threshold AVE values for Factors 2 and 3 indicate a limitation in variance convergence, suggesting that a proportion of the variance convergence, suggesting that a proportion of indicator variance is still influenced by measurement error. This implies that the convergent validity of these dimensions, while adequate, has not yet reached an optimal level and may benefit from further refinement. Therefore, future research is recommended to refine item wording, enhance conceptual specificity, and develop additional indicators to improve the variance explained by the latent constructs. Such efforts are expected to strengthen the instrument's convergent validity and overall psychometric quality.

Rasch Model Analysis: Scale Validity and Reliability

Rasch analysis was conducted on the 10 items retained after CFA to evaluate measurement quality at both the item and respondent levels, including item fit to the model, measurement reliability, and construct unidimensionality. The Rasch results showed that all items met model fit criteria: outfit MNSQ (0.5 < MNSQ < 1.5), outfit ZSTD ("2.0 < ZSTD < 2.0), and point–measure correlation (0.40 < Pt Mean Corr < 0.85). These ranges indicate that participants' responses were consistent with model expectations and that items functioned effectively in measuring the construct (Bond & Fox, 2013; Boone et al., 2014). Within the Rasch framework, items falling within these ranges are considered to function normally without causing measurement distortion, thus providing evidence of item-level validity. Accordingly, students' responses to each item aligned with the measured latent ability level, and no indication of misconceptions or of measuring different constructs was observed.

In addition to model fit, item difficulty analysis showed logit values ranging from "2.08 to 1.40. This variation indicates that the construct was not measured dichotomously (present vs. absent) but emerged gradually across different intensity levels. Some behaviors were easier for respondents to endorse, whereas others required higher levels of dependency to appear. From a Rasch perspective, a range of item difficulties suggests that the instrument maps stages of construct manifestation hierarchically rather than merely detecting the presence of the construct (Bond & Fox, 2013). Therefore, these findings support construct validity because the items were not only consistent with the model but also represented a spectrum of the measured construct.

Rasch analysis yielded a person reliability of 0.82 and an item reliability of 0.99. Person reliability reflects consistency in respondents' response patterns, while item reliability indicates the stability of item difficulty ordering within the measurement construct (Bond & Fox, 2013; Sumintono & Widhiarso, 2015). The high person reliability indicates that the instrument distinguished respondents based on their levels of the measured construct. In contrast, the very high item reliability indicates that the hierarchy of item difficulty was stable and not dependent on the research sample. Thus, the scale demonstrated strong measurement consistency from both respondent and item perspectives. A Cronbach's alpha of 0.83 also indicated strong internal consistency between items and respondents overall. In the Rasch context, this coefficient reflects not only item homogeneity but also the appropriateness of the interaction between respondents and items within a single measurement construct. Furthermore, a person separation index of 2.13 indicated that the instrument classified respondents into three strata of dependency levels, high, moderate, and low (Linacre, 2009). The item separation value of 12.41 indicated a highly stable item difficulty hierarchy independent of the sample (Wright & Masters, 1982). Therefore, the scale was not only consistent but also sufficiently sensitive in mapping levels of the measured construct.

Instrument unidimensionality was tested using Principal Components Analysis (PCA) of Rasch residuals to ensure that the instrument was dominated by a single latent construct. Based on recommended Rasch criteria, a good proportion of the raw variance explained by the measures is e" 40%, indicating strong construct dominance (Linacre, 2009; Boone et al., 2020). The analysis showed that the raw variance explained by the measures was 44.7%, exceeding the recommended threshold and indicating that most of the variance in the data was accounted for by the primary construct, namely AI task completion dependency.

The principal component analysis of residuals showed an unexplained variance of 13.3% in the first contrast, with an eigenvalue of approximately 2.4. In Rasch modeling, a first contrast eigenvalue in the range of 2–3 is still considered weak and does not indicate the presence of a substantive second dimension (Bond & Fox, 2013; Linacre, 2009). This indicates that the residual variation was insufficient to form another construct beyond the primary measured construct. Therefore, the instrument satisfied the unidimensionality assumption, meaning that all items collectively measured the same latent construct: students' dependency on AI in completing academic tasks.

Beyond evaluating model fit, reliability, and unidimensionality, the Rasch model also allows examination of measurement targeting between respondent characteristics and item characteristics. Therefore, the distribution of respondents' abilities and item difficulties was analyzed using a Wright map.

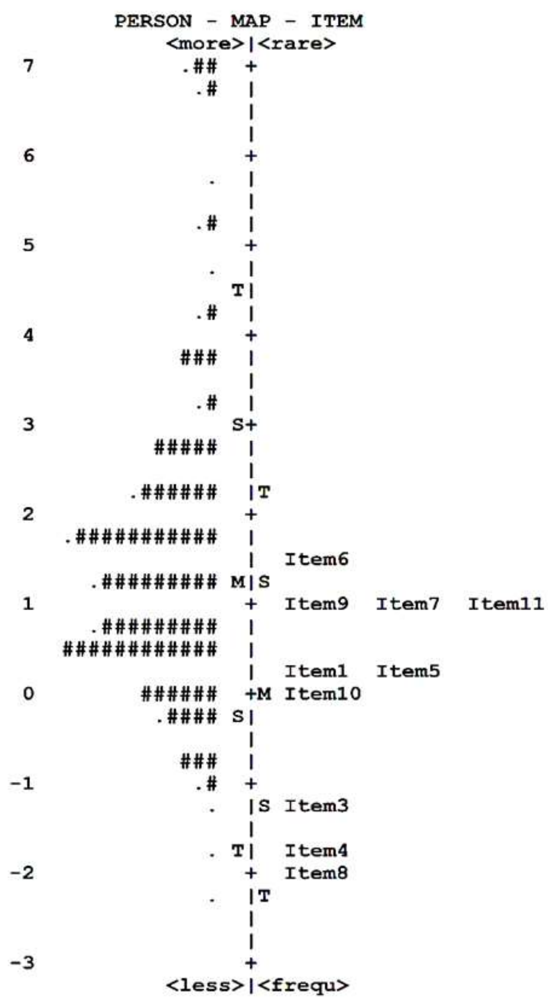

Figure 4. Wright map of the AI task dependency scale

The Wright map revealed a hierarchical ordering of items from easier to more difficult to endorse, reflecting increasing levels of AI dependency (Sumintono & Widhiarso, 2015). Beyond its statistical function, this hierarchy provides substantive insight into the qualitative progression in students' reliance on AI. Items positioned at lower difficulty levels, such as Item 3 and Item 4, which reflect the use of AI to speed up task completion and simplify difficult material, indicate instrumental and efficiency-driven engagement. behaviors suggest that students primarily perceive AI as a functional tool to suppo task completion rather than as a cognitive partner.

In contrast, items located at higher difficulty levels, particularly Item 8 ("AI helps develop ideas more broadly"), represent a more complex level of cognitive engagement. Endorsing this item implies that students attribute a role to AI in idea generation and conceptual development, which involves higher-order cognitive processes. The relatively greater difficulty in endorsing this item suggests that students are less inclined to perceive AI as a collaborator in creative thinking than as a facilitator of efficiency.

This pattern can be explained through the lens of creativity and cognitive processing. Creative thinking involves originality, idea elaboration, and internal cognitive restructuring (Runco & Jaeger, 2012). Students may be more cautious in attributing these processes to AI, as doing so may imply a reduced role of personal cognitive effort in knowledge construction. Empirical studies indicate that while generative AI can support ideation, excessive reliance on it may reduce originality and homogenize ideas (Dwivedi et al., 2023; Zhai et al., 2024). Therefore, students may recognize AI as useful for efficiency but remain hesitant to fully accept it as a source of creative cognition.

From a critical thinking perspective, this distinction is also meaningful. Critical thinking requires active analysis, evaluation, and reasoning (Facione, 1990). When AI is used primarily for efficiency (e.g., summarizing or simplifying tasks), students remain cognitively engaged in evaluating outputs. However, when AI is perceived as generating ideas, there is a greater risk of passive acceptance, which may weaken analytical engagement. This may explain why students are less likely to endorse AI as a tool for idea development than for in-task facilitation.

Furthermore, this finding aligns with metacognitive theory and self-regulated learning. Metacognition involves awareness and control of one's own thinking processes (Flavell, 1979), while Self-Regulated Learning emphasizes planning, monitoring, and evaluating learning activities (Zimmerman, 2002). Students who maintain metacognitive control are more likely to critically evaluate AI outputs rather than rely on them as primary sources of ideas. Thus, the greater difficulty of Item 8 may reflect students' efforts to preserve cognitive agency and avoid overreliance on AI in higher-order thinking processes.

From a cognitive offloading perspective, this distinction is particularly important. Cognitive offloading involves delegating cognitive processes to external tools to reduce mental effort (Risko & Gilbert, 2016). Early stages of offloading typically involve routine or low-level processes, such as information retrieval or task simplification. In contrast, delegating higher-order processes such as idea generation represents a deeper level of cognitive substitution. The Wright map suggests that most students remain at the level of instrumental offloading rather than fully externalizing generative thinking processes to AI.

Overall, the Wright map indicates that students are more willing to accept AI as a tool for efficiency than as a partner in creative cognition. This asymmetry suggests that AI dependency develops progressively, beginning with functional reliance and potentially extending toward deeper cognitive integration. However, the current pattern indicates that such a deeper dependency is not yet dominant, as students still retain some control over higher-order cognitive processes. This finding reinforces the conceptualization of AI task completion dependency as a gradual, multilevel construct that encompasses both functional use and evolving cognitive engagement.

Gender Invariance Analysis

Differential Item Functioning (DIF) analysis was used to evaluate whether items in the AI Task Completion Dependency Scale functioned equivalently for male and female students. In the Rasch model, two main indicators were used: the difference in item difficulty between groups (DIF contrast/logit) and statistical significance (p-value). A p-value < 0.05 indicates differences in item functioning between groups, while the magnitude of the DIF contrast reflects the strength of the difference (Bond & Fox, 2013; Linacre, 2009). In general, a logit difference of less than 0.5 is considered small and does not indicate substantial measurement bias (Sumintono & Widhiarso, 2015).

The analysis results (Table 6) showed that most items had p-values > 0.05 with small logit differences, indicating no DIF. This means the items were interpreted similarly by male and female students. Therefore, the probability of endorsing an item was influenced only by the level of the measured construct and not by respondents' group membership. This condition indicates that the instrument demonstrates measurement fairness (measurement invariance) across gender.

Table 6. Differential item functioning (DIF) results by gender (rasch model)

No	Item Statement	DIF Contrast (logit)	p-value	Interpretation
1	Using AI regularly (Item 1)	0.16	0.412	No DIF

2	AI helps complete tasks more quickly (Item 3)	0.07	0.753	No DIF
3	AI makes difficult material easier to understand (Item 4)	-0.14	0.508	No DIF
4	AI does not help speed up task completion (Item 5)	0.22	0.254	No DIF
5	Using AI whenever there is an assignment (Item 6)	0.00	1.000	No DIF
6	Preferring to complete assignments independently (Item 7)	-0.43	0.018	small DIF
7	AI helps develop ideas more broadly (Item 8)	0.49	0.019	small DIF
8	AI has become part of my routine (Item 9)	0.24	0.192	No DIF
9	Not using AI even when facing assignment deadlines (Item 10)	-0.31	0.119	No DIF
10	Relying on AI when completing assignments (Item 11)	0.20	0.273	No DIF

Based on the DIF analysis, Item 7 ("Preferring to complete assignments independently") and Item 8 ("AI helps develop ideas more broadly") showed statistically significant DIF. However, the magnitude of DIF for both items remained below 0.50 logits, indicating small DIF according to Rasch measurement guidelines (Bond & Fox, 2013; Boone et al., 2014). Small DIF suggests minor differences in response tendencies between groups rather than substantial measurement bias.

The pattern of DIF observed in these items may reflect differences in learning approaches and technology use between male and female students. Item 7 relates to preference for independent task completion, which is closely associated with self-regulated learning behaviors. Previous studies suggest that students may differ in their learning regulation strategies and technology use, with some research indicating that female students often report stronger tendencies toward self-regulated learning behaviors such as persistence and effort regulation, whereas male students may show greater willingness to experiment with technological tools during problem solving (Zimmerman, 2008; Venkatesh & Morris, 2000). Such differences could lead to slightly different response patterns when students evaluate their preference for completing assignments independently versus using AI assistance.

Meanwhile, Item 8 ("AI helps develop ideas more broadly") reflects the perception of AI as a cognitive partner in idea generation. The small DIF observed in this item may relate to differences in how students conceptualize AI as a creative or generative tool. Within the technology acceptance literature, some studies suggest that male students tend to report higher perceived usefulness of emerging digital technologies for exploratory or innovative purposes, whereas female students may evaluate such tools more cautiously and critically (Venkatesh & Morris, 2000; Ong & Lai, 2006). These differences in perceptions of technology may influence how students interpret AI's role in supporting idea generation, thereby contributing to minor variations in response patterns across gender groups. Consequently, the perception that AI contributes to idea development may be endorsed differently across gender groups.

Importantly, the magnitude of DIF for both items remained small and did not exceed recommended thresholds for problematic bias. Therefore, these differences likely reflect variations in learning strategies or perceptions of technology rather than a failure of the items to measure the same underlying construct of AI dependency. Consistent with scale development guidelines, items showing small DIF may be retained when their conceptual relevance is strong and when the overall measurement structure remains stable (DeVellis, 2012). Accordingly, both items were retained in the final instrument.

■ CONCLUSION

This study developed an Artificial Intelligence (AI) task completion dependency scale for university students and identified a three-factor structure consisting of functional dependency on AI, reflective attitude, independent use, regulation, and critical evaluation of AI. The integration of exploratory factor analysis, confirmatory factor analysis, and Rasch modeling provided a comprehensive psychometric examination of the proposed instrument. The Wright Map analysis further indicated that the items formed a meaningful hierarchical pattern, suggesting that AI dependency in academic task completion may progress from instrumental use to deeper cognitive reliance.

However, the psychometric evaluation also revealed several limitations. While the functional dependency dimension demonstrated satisfactory reliability and convergent validity, the reflective attitude and independent use dimension and the regulation and critical evaluation dimension did not fully meet the recommended threshold for convergent validity. These findings indicate that although the proposed structure shows preliminary empirical support, several indicators require further refinement to improve variance convergence and measurement robustness. Therefore, the present instrument should be interpreted as providing initial evidence of structural validity rather than a finalized measurement model, and additional scale development is required in future research.

Despite these limitations, the study contributes an initial conceptual and measurement framework for examining AI task completion dependency in higher education. The developed instrument has potential value as an evaluative tool in both research and educational practice. In research contexts, the scale may facilitate investigation of how AI usage relates to cognitive engagement, self-regulated learning, and academic outcomes. In instructional settings, lecturers may use the instrument as a diagnostic tool to identify patterns of AI reliance among students and to design learning activities that encourage reflective, responsible, and balanced AI use. At the institutional level, the instrument may also provide a preliminary basis for developing policies or guidelines related to AI literacy and responsible technology use in higher education.

This study also has several limitations that should be considered when interpreting the findings. The sample consisted only of university students in Indonesia, which may limit the generalizability of the results to other educational systems or cultural contexts. In addition, the study did not collect detailed background variables, such as academic discipline or the intensity of AI use, which may influence patterns of AI dependency. Furthermore, the data relied on self-reported responses, which may not fully reflect actual AI usage behavior. Future studies are therefore recommended to test the scale across more diverse populations, including different countries, educational levels, and academic disciplines, and to integrate behavioral data or longitudinal designs to better capture the dynamics of AI dependency over time. Taken together, the findings suggest that the proposed instrument offers a promising starting point for understanding how AI integration may shape students' learning behaviors and cognitive engagement in higher education. Continued refinement and validation of this scale will be essential for developing more robust tools to examine the evolving relationship between students and AI-supported learning environments.

DECLARATION OF GENERATIVE AI USAGE IN THE WRITING PROCESS

During the drafting and revision of this manuscript, the author utilized several AI-assisted tools to support the writing process. ChatGPT (OpenAI) was used for brainstorming and idea refinement; DeepL for translating text into English; Scopus AI for identifying relevant scholarly references; and QuillBot for paraphrasing selected sentences to improve clarity and readability. All AI-generated outputs were carefully reviewed, critically evaluated, and substantially revised by the author to ensure academic accuracy, coherence, and integrity. The author assumes full responsibility for the content, interpretation, and conclusions presented in this manuscript.

REFERENCES

Alifah, N., & Hidayat, A. R. (2025). Effectiveness of artificial intelligence-based learning analytics tool in supporting personalized learning in higher education. *Jurnal Pendidikan Progresif, 15*(1), 74–84. https://doi.org/10.23960/jpp.v15i1.pp

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Arbuckle, J. L. (2014). *IBM SPSS Amos 23 User's Guide*. IBM Corp.

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Psychology Press.

Boone, W. J., & Staver, J. R. (2020). Principal component analysis of residuals (PCAR). In W. J. Boone, J. R. Staver, & M. S. Yale (Eds.), *Advances in Rasch analyses in the human sciences* (pp. 13–24). Springer.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Byrne, B. (2016). *Structural Equation Modeling with AMOS/ : Basic Concepts, Application and Programming*, Third Edition. Routledge. https://doi.org/https://doi.org/10.4324/9781315757421

Byrne, B. M. (2012). Choosing structural equation modeling computer software: Snapshots of LISREL, EQS, AMOS, and Mplus.

Chen, Y., Wang, M., & Yuan, S. (2025). Development and validation of the conversational AI dependence scale for Chinese college students.

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology Press.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, *10*(1).

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). SAGE Publications.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., … Wright, R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

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), 272–299.

Facione, P. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report).

Fauzi, C., Rahmani, E. F., & Utimadini, N. J. (2025). Beyond technology acceptance: An interplay of self-efficacy, language proficiency, and ChatGPT adoption from a TAM perspective. *Jurnal Pendidikan Progresif, 15*(3). https://doi.org/10.23960/jpp.v15i3.pp19

Fawwaz, A. N., Prihadi, K. D., & Purwaningtyas, E. K. (2022). Studying online from home and social anxiety among university students: The role of societal and interpersonal mattering. *International Journal of Evaluation and Research in Education, 11*(3), 1338–1345. https://doi.org/10.11591/ijere.v11i3.23464

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34(10)*, 906.

Foo, Z., & Prihadi, K. D. (2021). Happiness of university students in the new normal Malaysia: The role of mattering, optimism, and social support. *International Journal of Evaluation and Research in Education, 10*(2), 448–454. https://doi.org/10.11591/ijere.v10i2.21138

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Gökbulut, B. (2019). The relationship between sense of belonging and technology addiction of high school students. *International Journal of Eurasian Education and Culture, 4*(7), 281–297. https://doi.org/10.35826/ijoecc.46

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

Gruetzemacher, R., & Whittlestone, J. (2022). The transformative potential of artificial intelligence. *Futures, 135*, 102884. https://doi.org/10.1016/j.futures.2021.102884

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *An introduction to structural equation modeling*. Springer. https://doi.org/10.1007/978-3-030-80519-7_1

Hairida, H., Benõ, C., Soeharto, S., Charalambos, C., Rasmawan, R., Martono, M., Arifiyanti, F., Winarti, A., & Enawaty, E. (2023). Evaluating digital literacy of pre-service chemistry teachers: Multidimensional Rasch analysis. *Journal of Science Education and Technology, 32*(5), 643–654. https://doi.org/10.1007/s10956-023-10070-z

Hakimi, A., Li, R., Yue, M., Muhsin, M. S., & Bakar, M. A. (2025). The social impact of artificial intelligence chatbots on college students. *International Journal of Evaluation and Research in Education, 14*(1), 10–16. https://doi.org/10.11591/ijere.v14i1.29469

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.

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

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.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31–36.

Karny, S., Mayer, L. W., Ayoub, J., Song, M., Su, H., Tian, D., Moradi-Pari, E., & Steyvers, M. (2024). Learning with AI assistance: A path to better task performance or dependence? In *Proceedings of the ACM Collective Intelligence Conference* (pp. 10–17). https://doi.org/10.1145/3643562.3672610

Kline, R. B. (1998). Software programs for structural equation modeling: AMOS, EQS, and LISREL. *Journal of Psychoeducational Assessment, 16*(4), 343–364.

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology, 9*(8), 2207–2230. https://doi.org/10.4236/psych.2018.98126

Lin, H., & Chen, Q. (2024). Artificial intelligence-integrated educational applications and college students' creativity and academic emotions: Students and teachers' perceptions and attitudes. *BMC Psychology, 12*(1). https://doi.org/10.1186/s40359-024-01979-0

Linacre, J. M. (2009). *A user's guide to Winsteps Rasch-model computer programs*. MESA Press.

Linacre, J. M., & Wright, B. D. (2000). *Winsteps Rasch measurement computer program*.

Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). Exploratory item factor analysis: A practical guide revised and updated. *Anales de Psicología, 30*(3), 1151–1169.

Lorenzo-Seva, U. (2022). SOLOMON: A method for splitting a sample into equivalent subsamples in factor analysis. *Behavior Research Methods, 54*(6), 2665–2677.

Miranda, J. P. P., & Yambao, J. A. (2025). Assessing novice programmers' perception of ChatGPT: Performance, risk, decision-making, and intentions. *Journal of Education and Learning (EduLearn), 19*.

Moukhliss, G., Lahyani, K., & Diab, G. (2024). The impact of artificial intelligence on research and higher education in Morocco. *Journal of Education and Learning (EduLearn), 18*(4), 1292–1300.

Murray, J. (2024). The influence of education and self-perceived tech savviness on AI reliance: The role of trust.

Ong, C.-S., & Lai, J.-Y. (2006). Gender differences in perceptions and relationships among determinants of e-learning acceptance. *Computers in Human Behavior, 22*(5), 816–829.

Penuela, A., Habaña, M. P., Oducado, R. M. F., Pedroso, J. E., Bonzo, E., & Tamdang, K. (2025). Correlates of ChatGPT use intention among freshmen health sciences students. *Journal of Education and Learning (EduLearn), 19*(3), 1615–1622.

Pérez, E. R., & Medrano, L. A. (2010). Exploratory factor analysis: Conceptual and methodological foundations. *Revista Argentina de Ciencias del Comportamiento, 2*(1), 58–66.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385–407.

Prayoga, K. P., Suryana, D., Supriatna, M., & Budiman, N. (2024). *Penggunaan Rasch model untuk menganalisis konstruk* *instrumen kontrol diri pada siswa sekolah menengah* [The use of the Rasch model to analyze the construct of self-control instruments in high school students]. *G-Couns: Jurnal Bimbingan dan Konseling, 9*(1), 367–381. https://doi.org/10.31316/gcouns.v9i1.4459

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688.

Rouquette, A., Falissard, B., & De Pinho, R. (2011). Sample size requirements for the internal validation of psychiatric scales. *International Journal of Methods in Psychiatric Research, 20*(4), 235–249. https://doi.org/10.1002/mpr.352

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96.

Sairitupa-Sánchez, 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*. https://doi.org/10.3389/feduc.2024.1323898

Shah, S. S., & Asad, M. M. (2024). Impact of critical thinking approach on learners' dependence on innovative transformation through artificial intelligence. In *The evolution of artificial intelligence in higher education: Challenges, risks, and ethical considerations* (pp. 161–182). Emerald Publishing.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504.

Siregar, Z. M. E., Parlauangan, A., Supriadi, Y. N., Ende, & Pristiyono. (2021). *Structural equation modeling*. Deepublish.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Lawrence Erlbaum Associates.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Trim Komunikata Publishing House.

Ulfah, M. (2024). *Dampak ketergantungan pada artificial intelligence terhadap kemampuan analitis dan kreatif mahasiswa* [The impact of dependence on artificial intelligence on students' analytical and creative abilities]. *Vox Edukasi: Jurnal Ilmiah Ilmu Pendidikan, 15*, 120–130.

Vaus, D. de. (2013). *Surveys in social research* (6th ed.). Routledge. https://doi.org/10.4324/9780203519196

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly, 24*(1), 115–139.

Wahyuningsih, S. (2024). Does artificial intelligence play roles in enhancing academic writing? Unravelling lecturers' voices in Indonesian higher education. *Jurnal Pendidikan Progresif, 14*(1), 489–503. https://doi.org/10.23960/jpp.v14.i1.202436

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913–934.

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.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press.

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). https://doi.org/10.1186/s40561-024-00316-7

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.

Zhou, X., Teng, D., & Al-Samarraie, H. (2024). The mediating role of generative AI self-regulation on students' critical thinking and problem-solving.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological development, and future prospects. *American Educational Research Journal, 45*(1), 166–183.