International Journal of Instruction
e-ISSN: 1308-1470 ● www.e-iji.net

April 2026 ● Vol.19, No.2
p-ISSN: 1694-609X
pp. 749-766

Article submission code: 20250910122226

Received: 10/09/2025
Revision: 01/12/2025

Accepted: 09/12/2025
OnlineFirst: 07/01/2026

From Prompting to Understanding: Factorial Structure and Network Psychometrics of a Deep-Learning Scale in Generative AI Context

Eman Salah Daha
Asst. Prof., corresponding author, Department of Educational Psychology, Faculty of Education, Damanhour University, Egypt, *Eman_Daha@edu.dmu.edu.eg*

Aml Altelwany Altelwany
Lecturer, Department of Educational Psychology, Faculty of Education, Damanhour University, Egypt, *dr.aml.telwany@edu.dmu.edu.eg*

This study aimed to develop and validate a scale for assessing deep learning approaches in their own learning among university students in the context of generative AI tools. Most existing measures were originally designed for traditional educational settings and do not reflect the substantial changes introduced by AI in learning practices. To address this gap, a 16-item scale was designed across four dimensions—search for meaning, relating ideas, analysing evidence, and attention to concepts—after adapting the items to AI-supported environments. The scale was administered to 765 Damanhour University students with prior experience using generative AI. Data were examined through exploratory and confirmatory factor analyses, convergent and discriminant validity tests, and network analysis. Results indicated strong psychometric properties, with validity and reliability at high levels. The network analysis revealed clear interconnections among items and highlighted central nodes representing core aspects of deep learning, which shaped other learning behaviours and reinforced the adoption of deep learning approaches. Overall, findings confirm the scale as a modern, reliable tool for capturing students' deep learning strategies in AI-enhanced contexts, offering implications for future research and for improving higher education practices in light of ongoing digital transformations.

Keywords: deep learning scale, generative artificial intelligence, psychometric validation, convergent validity, discriminant validity, network psychometrics

INTRODUCTION

Higher education has recently undergone major shifts due to rapid advances in artificial intelligence (AI), which introduced a variety of smart tools increasingly embedded in learning environments. These tools provide personalized content and instant feedback, reshaping student and teacher roles and changing interaction with knowledge. The impact is most visible in digital platforms that integrate generative AI to promote modern learning and offer interactive, customized experiences. Students now rely on

Citation: Daha, E. S., & Altelwany, A. A. (2026). From prompting to understanding: Factorial structure and network psychometrics of a deep-learning scale in generative AI context. *International Journal of Instruction*, *19*(2), 749-766. https://doi.org/10.29333/iji.2026.19240a

such tools to complete academic tasks more flexibly and efficiently, fostering self-directed learning and enhancing active participation (Iqbal et al., 2025).

Despite these opportunities, challenges remain. Studies warn that excessive reliance on generative AI without activating students' cognitive processes may lead to passive information use, limited understanding, and even unethical practices affecting academic integrity (Abbas et al., 2024; Zhang et al., 2025). This highlights the need for deeper learning approaches, where students engage in higher order thinking to connect ideas and construct meaning. Deep learning, unlike surface approaches based on rote memorization, enables learners to pursue clear cognitive goals and ensure meaningful outcomes, positioning AI tools as supports for thinking rather than substitutes (Abbas et al., 2024; Zhang et al., 2025).

Given these concerns, it is essential to examine deep learning as a key pathway for understanding how students interact with generative AI tools. This involves not only educational outcomes but also the cognitive processes and strategies that sustain meaningful learning (Ogunleye et al., 2024).

Deep learning focuses on grasping meaning and linking concepts, unlike surface learning that relies on memorization. Research shows that deep learners develop more durable knowledge, better skill transfer, and stronger critical and creative thinking, which are essential for higher education and future careers (Biggs & Tang, 2007; Entwistle & Peterson, 2004; Tamire et al., 2022). Their engagement enables evidence analysis, idea generation, and logically grounded conclusions, supporting sustainable academic achievement."

Given its importance, there is a growing need for precise tools to assess deep learning and learners' orientations. While instruments like the SPQ (Biggs, 1987) and ASSIST (Entwistle & Ramsden, 1983; Entwistle et al., 2000) have shown validity and reliability in higher education, their design for traditional contexts raises questions about their adequacy in today's digital learning environments.

In recent years, studies have begun examining how generative AI tools relate to deep and surface learning. Duffy (2025) showed that students may use these tools to deepen understanding, whereas Gezgin (2024) found that others rely on them for ready-made answers, reflecting surface approaches. Despite such insights, standardized instruments to measure learning patterns in AI-based contexts remain lacking.

Given the limitations of traditional instruments that overlook the dynamics of AI-supported learning, and the scarcity of research addressing deep learning in this context, there is a clear need for an updated tool. Such a scale should build on established dimensions of deep learning (seeking meaning, connecting ideas, analyzing evidence, and focusing on concepts), while adapting them to generative AI environments in higher education. Developing and validating this instrument through advanced psychometric analyses will help bridge the existing gap and provide educators and researchers with a reliable means to assess how students engage with these technologies in ways that foster genuine, deep learning. To ensure the theoretical soundness of the instrument, construct validity was considered a central principle guiding the scale's development. This involves confirming that each dimension accurately represents the underlying aspects of deep learning thereby providing a robust foundation for subsequent psychometric validation procedures.

Accordingly, this study aims to develop and validate a Deep Learning Scale to assess students' adoption of deep learning approaches when using generative AI tools in higher education.

CONTEXT AND REVIEW OF LITERATURE

Generative AI and Higher Education

Generative AI has transformed higher education by enabling students to interact with knowledge through dynamic methods such as generating personalized explanations, formulating questions to enhance understanding, and providing instant feedback. Major companies have also introduced AI-driven educational applications, including Coursera Coach, Khanmigo, and ongoing updates in Microsoft Teams for Education. Recent studies highlight that structured use of these tools supports critical thinking, expands interactive learning opportunities, and fosters environments that promote deep understanding (Essien et al., 2024; Urban et al., 2024; Hakiki et al., 2023).

Despite its potential, uncritical reliance on generative AI raises concerns. Studies warn that easy access to ready-made answers may reduce students' independent effort and weaken deep cognitive processing unless they use these tools consciously (Fan et al., 2025; Farrokhnia et al., 2023). Overall, these studies indicate a range of outcomes, reflecting both the potential benefits and the concerns associated with students' engagement with generative AI. Thus, generative AI offers significant opportunities for higher education but also poses the challenge of ensuring it supports deep rather than superficial learning.

The Importance of Deep Learning with AI

The concept of deep learning emerged in the 1970s as researchers distinguished between surface and deep processing of information ,Deep learning involves critical thinking and connecting new knowledge with prior concepts and experiences, while surface learning focuses on rote memorization for exams without grasping underlying relationships (Marton & Säljö, 1976; Entwistle & Ramsden, 1983; Entwistle, 2017).

Deep learning is reflected in seeking meaning, connecting ideas, analyzing evidence, and active engagement with concepts, which foster intrinsic motivation (Harackiewicz et al., 2016). It also promotes reflective thinking and self-regulation, supports information evaluation and sound conclusions, and enhances academic achievement and knowledge transfer (Biggs & Tang, 2007; Entwistle & Peterson, 2004; Tamire et al., 2022; Dohlmans et al., 2016; Nhat& Van Le,2023).

Deep learning differs from surface learning by emphasizing understanding, analysis, and learner autonomy, while surface learning relies on memorization (Marton & Säljö, 1976; Entwistle & Peterson, 2004), and it is also influenced by several factors, such as self-efficacy, motivation, and overall student satisfaction (Yating et al., 2025). Given its role in fostering critical thinking and idea integration, deep learning is especially valuable when using AI tools, enabling students to apply them consciously and effectively for independent learning and complex problem-solving (Biggs & Tang, 2007; Tamire et al., 2022).

Theoretical Models and Measurement Tools of Deep Learning

Researchers have proposed several models to explain deep learning. Biggs' 3P model (1987) views learning as an interaction between learner characteristics (inputs), adopted approach (surface or deep), and outcomes, highlighting that intrinsic motivation and supportive environments foster deep learning and high-quality results. Similarly, Entwistle & Peterson (2004) linked learning approaches to students' perceptions: surface when assessment stresses memorization, deep when the environment promotes understanding. Hattie & Donoghue (2016) distinguished surface, deep, and transfer strategies, emphasizing that self-explanation and critical questioning enhance knowledge transfer and long-term retention.

To assess students' adoption of deep learning, Biggs first developed the Study Process Questionnaire (SPQ) in the late 1970s, later revised in 1987, which measures learning motives and strategies as deep, surface, or achieving. Subsequently, the ASSIST scale (Entwistle et al., 2013) offered a broader framework with three main approaches—deep, surface, and strategic—detailing the deep approach into four sub-dimensions: seeking meaning, relating ideas, analyzing evidence, and focusing on concepts. Empirical studies, such as Brown et al. (2015) with university chemistry students, confirmed the validity and reliability of these dimensions.

Classical measures such as SPQ and ASSIST proved effective in distinguishing learning approaches across cultures (Entwistle & McCune, 2013). However, most existing models and tools were developed in traditional or general digital contexts, which may not fully capture students' deep learning behaviors in AI-enhanced environments. This gap underscores the importance of developing a specialized instrument tailored to these contemporary learning contexts.

The Need for a Deep Learning Measure in Generative AI Environments

The widespread use of generative AI in universities shapes students' learning approaches, ranging from surface copying to deep critical inquiry (Yang et al., 2024; Rahimi & Mosalli, 2025). Factors such as usage patterns, purpose, and intensity influence the depth of learning more than gender or academic year (Gezgin, 2024), while instructional guidance and question formulation steer AI use from mere information consumption toward critical thinking (Duffy, 2025). These emerging behaviors go beyond what traditional measures like SPQ and ASSIST capture.

This highlights the need to develop modern measures tailored to generative AI environments, as traditional items do not accurately capture students' practices in this context. For instance, an item like "I look for real-life examples to illustrate concepts" may need revision to assess how students integrate AI outputs with prior knowledge or other external sources. Similarly, "I do not accept information without verifying it" gains new dimensions when the source is an intelligent system requiring students to check its credibility. Therefore, designing a new or adapted instrument is essential to monitor students' adoption of deep learning approaches while using generative AI tools. Based on this perspective, the present study relied on the ASSIST dimensions as a foundation, with items reformulated to align with AI interaction, aiming to examine its psychometric properties in a contemporary educational context.

Study Hypothesis

The Deep Learning Scale in AI contexts shows psychometric adequacy—reliability, validity, and applicability for use in educational contexts.

METHOD

Sample and Data Collection:

This study adopted a descriptive-analytical approach. The population comprised all first-year undergraduate students at the Faculty of Education, Damanhour University, registered in the 2024/2025 academic year. A proportional stratified sampling based on department, academic year, and gender was used to ensure balanced representation and minimize potential stratification bias in estimating the psychometric properties of the study instruments. This approach is recommended in scale development and construct validation studies to enhance external validity and generalizability (Boateng et al., 2018; DeVellis, 2017). The studies were conducted in the Egyptian higher education context; therefore, generalizability to other cultural or educational settings may be limited.

A preliminary survey on generative AI usage intensity for educational purposes was applied to screen participants based on sufficient exposure and experience. Students with a mean score ≥ 2.5/5 on the total usage index were included to ensure actual and diverse AI engagement before administering the Deep Learning Approach scale. This aligns with best practices for setting inclusion criteria in behavioral and educational scale development (DeVellis, 2017; Boateng et al., 2018).

The online invitation for the generative AI usage survey was sent to 1,200 students, with 1,128 responses (≈94%). After removing incomplete or invalid entries, 1,032 valid responses remained. Applying the inclusion criterion (mean usage ≥ 2.5/5) yielded 865 students (≈83.8%), from whom the final sample of 765 students was drawn for analysis.

The final sample (N = 765) ensured high statistical power and stable estimates for CFA and fit indices (CFI/TLI/RMSEA). Its size also supports subgroup analyses (e.g., by gender or major) and reduces the risk of improper solutions or biased SEM estimates (Kline, 2016; Wolf et al., 2013; MacCallum et al., 1999).

Participation was voluntary, for educational research only, with confidential, aggregated reporting to ensure data reliability (Boateng et al., 2018).

The sample used proportional stratified sampling by year (1–4), gender, and specialization, split into EFA (N = 250) for exploratory factor analysis and CFA (N = 515) for confirmatory factor analysis via structural equation modeling.

This EFA → CFA design enhances external validity and reduces overfitting, aligning with best practices in educational scale development (Boateng et al., 2018; DeVellis, 2017).

Table 1 shows the numeric distribution of EFA and CFA participants by year, gender, and specialization, with totals. The four-year distribution is balanced (190–193), females slightly predominate (~51%), and science vs. arts is nearly even (388 vs. 377), supporting adequate representation and reducing bias in measurement estimates (Kline, 2016).

Table 1
FA vs. CFA Samples by Year, Gender, and Specialization

Indicator / Sample		EFA (N=250)	CFA (N=515)	Total (N=765)
Year	First Year	61	129	190
	Second Year	61	129	190
	Third Year	63	129	192
	Fourth Year	65	128	193
Gender	Male	122	253	375
	Female	128	262	390
Specialization	Science	132	256	388
	Arts	118	259	377
Total		250	515	765

The mean age was 20.28 ± 1.19 years for the EFA sample and 20.25 ± 1.18 years for the CFA sample, with a total mean of 20.26 ± 1.19, consistent with typical undergraduate ages and providing adequate variance to detect structural differences without inflating age effects (Kline, 2016).

The study tools

Generative AI Usage Intensity Questionnaire (Screening Tool): A scale was developed to measure students' use intensity of generative AI tools (e.g., ChatGPT, Copilot, Gemini) in higher education. It comprised three dimensions—frequency, diversity, and educational purposes (e.g., explanation, question generation, writing, assessment preparation)—with responses on a five-point Likert scale (1 = never, 5 = always). A cutoff (≥ 2.5/5) was applied to include only participants with sufficient exposure before administering the main instrument.

Psychometric validation showed acceptable content validity (expert review), a three-factor structure (EFA), good model fit (CFA), satisfactory reliability (α = 0.72–0.84), and statistically significant internal consistency. Thus, the scale was deemed suitable for screening and ensuring inclusion of actual users only (DeVellis, 2017; Rosen et al., 2013; Boateng et al., 2018).

Deep Learning Scale during Generative AI Use

This primary study tool was adapted from prior literature (Marton & Säljö, 1976; Brown et al., 2015). It consists of 16 items across four dimensions: seeking meaning, relating ideas, analyzing evidence, and focusing on concepts.

Analysis Method

The Deep Learning Scale was developed from prior literature, reviewed by experts, administered to the sample, and validated through EFA followed by CFA. CFA models were estimated in IBM SPSS AMOS (Version 27) using Maximum Likelihood (ML) estimation. Convergent and discriminant validity, model fit, and reliability (Cronbach's α, CR) were assessed, and a psychometric network analysis was conducted to examine the internal structure and item relations.

FINDINGS

Study Hypothesis: The Deep Learning Scale in AI contexts shows psychometric adequacy—reliability, validity, and applicability in educational settings —verified through subsequent analyses.

1. The scale was evaluated by 11 experts in educational technology, educational psychology, machine learning, and AI to ensure item clarity, appropriate alignment with dimensions, and comprehensive coverage of deep learning with balanced weighting. Content validity was assessed using the Content Validity Ratio (CVR), with values ranging from 0.818 to 1, all exceeding the critical value of 0.62 for n=11 (Ayre, 2013), confirming adequate expert validity. Necessary revisions were made based on their feedback.

2. Exploratory factor analysis was conducted on the 16 items after confirming suitability: most inter-item correlations exceeded 0.30, the determinant of the correlation matrix was 0.001 (> 0.00001), the KMO value was 0.822 indicating sample adequacy, and Bartlett's test was significant ($\chi^2 = 5643.815$, df = 120, $p < 0.01$).

3. First-order EFA using Hotelling's principal components with eigenvalue > 1, Varimax rotation, and item loading ≥ 0.30 extracted four factors explaining 86.14% of total variance. Factor 1 "Conceptual Interest" explained 33.13% (eigenvalue = 5.3, loadings = 0.947–0.983); Factor 2 "Linking Ideas" 21.81% (3.489, 0.948–0.986); Factor 3 "Evidence Analysis" 17.70% (2.832, 0.941–0.983); and Factor 4 "Seeking Meaning" 13.51% (2.161, 0.754–0.834). All items were retained with loadings > 0.40. Figure 1 shows the scree plot of eigenvalues.

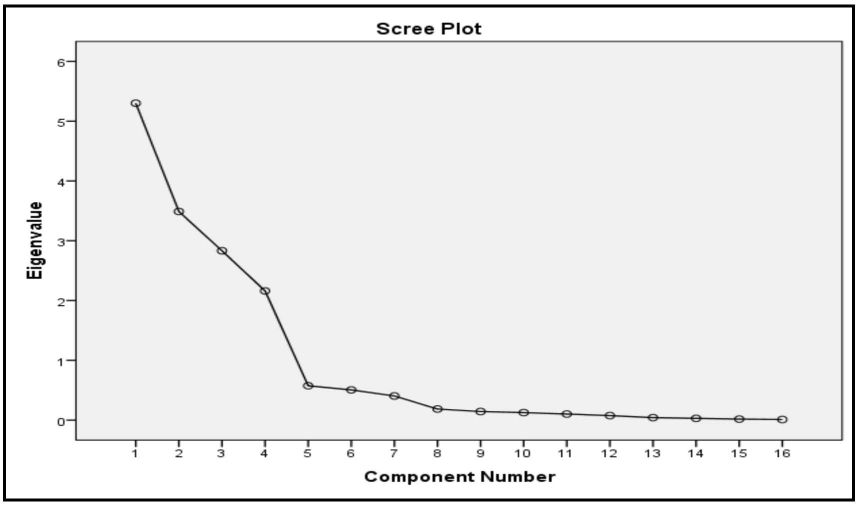

Figure 1
Eigenvalues of the Deep Learning Scale – Scree Plot.

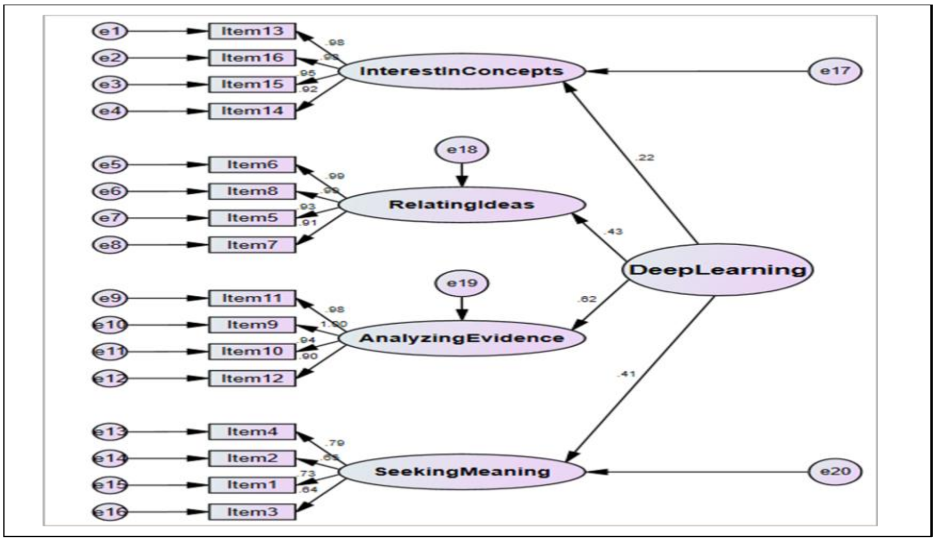

Figure 2
First- and Second-Order CFA Results of the Deep Learning Scale

The Scree Plot revealed a clear elbow at the fourth component, supporting the retention of four factors per Cattell's criterion. Their eigenvalues (>1) were 5.300, 3.489, 2.832, and 2.161, jointly explaining 86.143% of the total variance. Item loadings ranged from 0.754 to 0.986 without substantial cross-loadings, indicating well-differentiated factors and supporting their retention for subsequent CFA/SEM validation. After establishing the factor structure through EFA, the model was confirmed using CFA.

4. CFA (first- and second-order) was conducted using AMOS. Results showed that the 16 items loaded onto four primary factors, which in turn loaded onto a higher-order factor, "Deep Learning" (Figure 2). Fit indices (χ^2/df, GFI, AGFI, RMSEA, NFI ,CFI; Table 2) indicated acceptable model fit at both levels, supporting the factorial validity of the scale in AI contexts.

Table 2
Fit indices for the factorial structure of the Deep Learning Scale.

Index	First-order CFA index value	Second-order CFA index value	Goodness-of-fit criterion
χ^2/df	1.522	2.221	0 to 5
GFI	0.733	0.852	0 to 1
AGFI	0.711	0.808	0 to 1
RMR	0.070	0.154	0 to 1
RMSEA	0.051	0.079	0 to 0.1
NFI	0.876	0.911	0 to 1
CFI	0.954	0.949	0 to 1
RFI	0.871	0.897	0 to 1

To provide detailed justification of the statistical analysis and model fit, Table 2 presents key fit indices for the first- and second-order CFA models. The chi-square to degrees of freedom ratios (χ^2/df = 1.522 and 2.221) fall within the good-fit range (1–3) and well below the commonly accepted upper limit of 5. RMSEA values (0.051 and 0.079) indicate good to acceptable fit, while CFI values (0.954 and 0.949) demonstrate high comparative fit quality, approaching the ideal 0.95 threshold. NFI and RFI also reached acceptable levels (above 0.87 and exceeding 0.90 in the second-order model). Although GFI and AGFI were below 0.90 in the first-order model, their improvement in the second-order model (0.852 and 0.808) and sensitivity to model complexity suggest they remain acceptable. Overall, these indices support the factorial validity of the Deep Learning Scale, providing a solid foundation for subsequent reliability and network analyses. With the factorial validity confirmed, the scale's reliability and validity indices were subsequently examined.

5- Reliability**:** Scale reliability was assessed using Cronbach's alpha. Subscale coefficients ranged from 0.791 to 0.978, while the overall scale reached 0.848—indicating high reliability.

6- Convergent Validity**:** To verify the internal consistency of the four dimensions, convergent validity was assessed using Average Variance Extracted (AVE) and Composite Reliability (CR). CR values exceeded the 0.70 threshold (Hair et al., 2019), ranging from 0.835 to 0.998, indicating low random error and strong item–factor loadings. AVE values also surpassed the 0.50 benchmark (Fornell & Larcker, 1981), ranging from 0.561 to 0.991, confirming item homogeneity and substantial variance explained—thus supporting the scale's convergent validity.

7- Discriminant Validity: Discriminant validity was examined using the Fornell–Larcker criterion, requiring that the square root of AVE for each factor exceed its correlations with other factors (Fornell & Larcker, 1981). Results confirmed this condition; for example, the square root of AVE for the *Connecting Ideas* factor was 0.984, exceeding its highest correlation with another factor (0.36 with *Evidence Analysis*). In addition, the HTMT (heterotrait-monotrait ratio), which measures the degree of overlap among dimensions, was applied (Henseler et al., 2015), with values ranging from 0.09 to 0.47—well below the accepted threshold of 0.85. These findings confirm that the four factors are statistically distinct, supporting the scale's discriminant validity. To complement the traditional factor analyses and explore inter-item relationships, a psychological network analysis was conducted.

8- **Psychological Network Analysis:** As a complementary method to traditional factor analyses, a psychological network was used to represent the 16 items as nodes and their regularized partial correlations as edges (Borsboom & Cramer, 2013). To ensure transparency and replicability, ordinal Likert responses were first converted into a polychoric correlation matrix, followed by the estimation of a Gaussian Graphical Model using the EBICglasso algorithm with a tuning parameter of $\gamma = 0.50$, which retains only the most stable and interpretable edges (Epskamp et al., 2018). The final network was visualized using a Fruchterman–Reingold force-directed layout, allowing items with stronger associations to cluster naturally in the plotted structure.

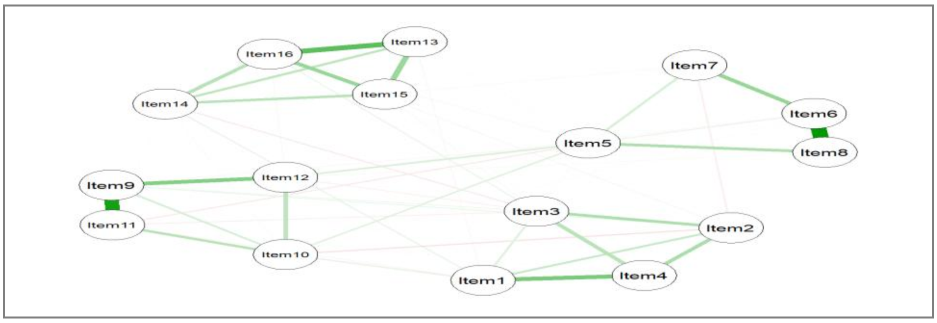

Figure 3
Psychological network of 16 items (EBICglasso)

Figure 3 illustrates the psychological network of the items, where circles represent nodes (items) and lines represent partial correlations; line thickness indicates correlation strength, while node arrangement reflects item proximity based on their interrelations. This visualization helps readers quickly see which items are most closely related and how the main groups of items form distinct clusters.

The network shows four distinct communities, each containing four items corresponding to the scale's dimensions (Seeking Meaning, Connecting Ideas, Analyzing Evidence, Focusing on Concepts), confirming clear and coherent dimensions consistent with previous factor analyses.

Despite the clear clustering of items within their four dimensions, the network analysis reveals some cross-dimensional links that may not appear in traditional factor analysis, such as connections between Analyzing Evidence items and those of Connecting Ideas or Seeking Meaning. For instance, Item 9 ("I do not accept information from AI without verifying its credibility") shows a direct link with Connecting Ideas items (5, 6), indicating that students who verify information also tend to integrate it with prior knowledge. These links reflect a natural interconnection among deep learning practices but remain weaker than within-dimension links, confirming the distinction of the main communities.

One advantage of network analysis is its ability to identify the most central nodes—items that play a key role in linking different parts of the network. Node strength, calculated as the sum of absolute edge weights connected to each item (Epskamp et al., 2018), indicates how strongly an item is connected to others and highlights potential anchor points in deep learning structure. Figure 4 shows strength values for all items, with peaks representing the most influential and widely connected items across dimensions, and troughs indicating less connected items. This allows readers to easily identify which items play key roles in linking different aspects of deep learning.

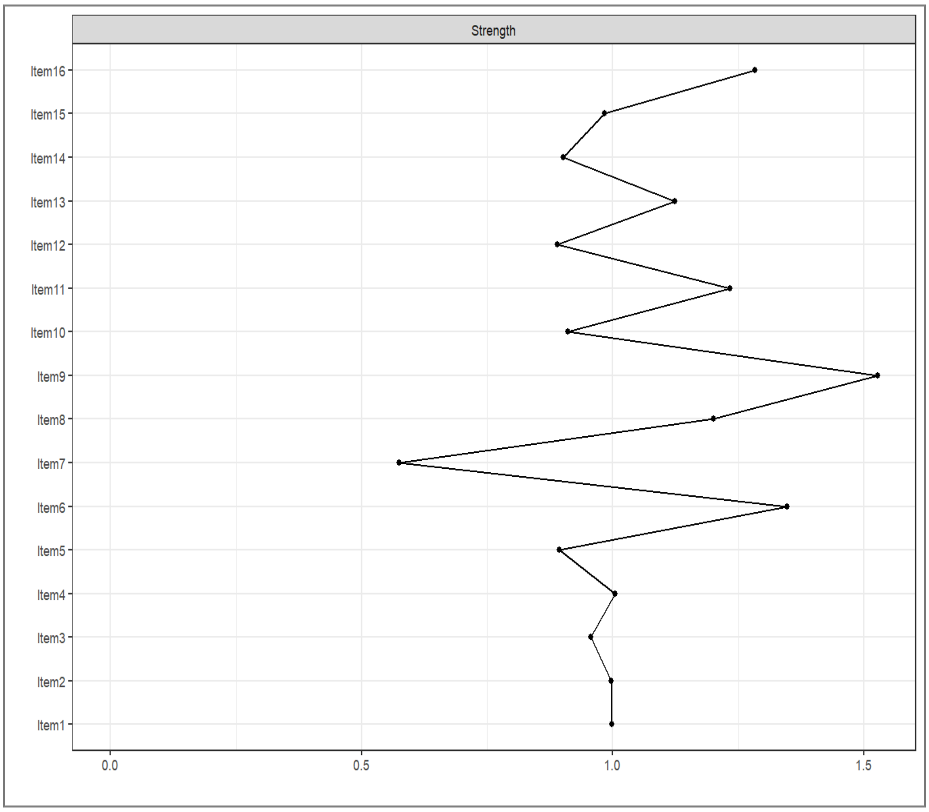

Figure 4
Items' strength centrality values ordered by number

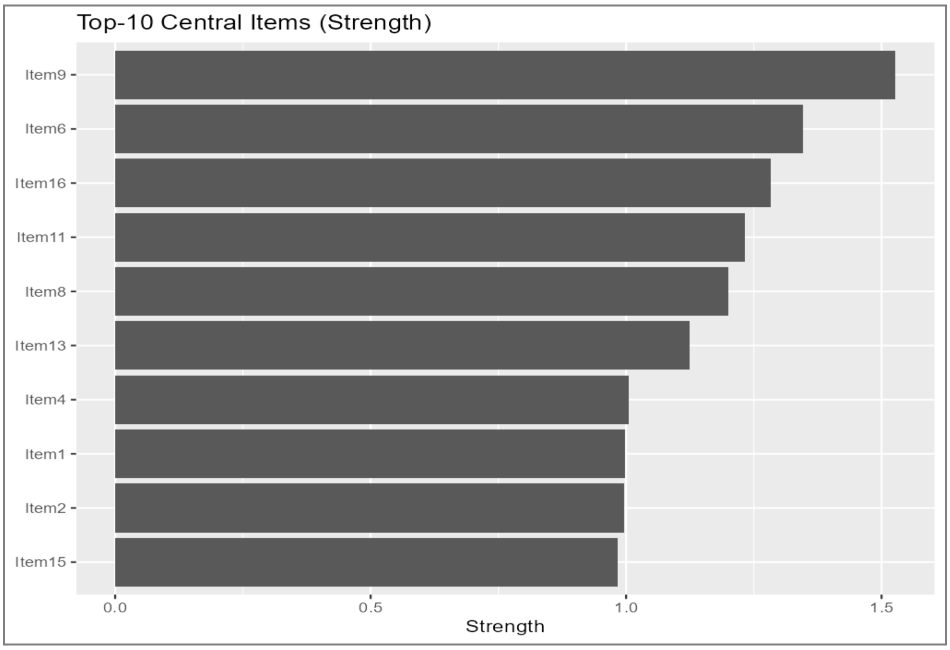

Figure 5
Top 10 items by strength centrality in the scale network

Figure 5 shows variation in strength centrality among the items in the network; some items (Item 9, Item 6, Item 16, and Item 11) exhibit high values, indicating strong connections with other items, whereas Item 7 shows notably low centrality. These results suggest that certain items act as pivotal nodes in the scale's structure, while others have limited influence in linking the deep learning dimensions.

Figure 5 displays the top ten items by strength centrality in the network. Item 9 (verifying the credibility of AI-provided information) is the most central, followed by Item 6 (connecting AI-provided ideas with prior knowledge), Item 16 (feeling enthusiastic when using AI to discover new concepts), and Item 11 (ensuring that AI-provided information aligns with scientific facts before relying on it). These results indicate that items related to analyzing evidence (9, 11), Connecting Ideas (6), and interest in concepts (13, 16) serve as pivotal nodes, strongly linking the deep learning behaviors in the network.

The present study developed a Deep Learning Scale aligned with the emerging behaviors and educational practices of students in AI-enhanced learning environments. The scale was rigorously validated using multiple approaches, including classical factor analyses (EFA and CFA) and psychological network analysis, which examined inter-item relationships and the structural coherence of the dimensions. These procedures ensure that the instrument reliably captures contemporary deep learning behaviors in response to generative AI tools, offering a novel contribution to educational assessment.

DISCUSSION

The study's results indicate that the developed scale for measuring deep learning during the use of AI possesses strong psychometric properties, with its validity verified using multiple complementary approaches. Exploratory and confirmatory factor analyses demonstrated that the scale's factor structure aligns with the theoretical framework, as the items clearly clustered into four distinct dimensions, confirming the scale's adequacy in accurately representing the facets of deep learning. Furthermore, convergent validity tests (e.g., AVE and CR) supported the strong inter-item correlations within each dimension, indicating that each dimension measures a coherent and unified construct, while discriminant validity tests (e.g., Fornell-Larcker and HTMT) confirmed that the four factors are relatively independent of one another, reinforcing the notion that deep learning consists of interconnected yet non-overlapping components. These findings align with prior international studies (e.g., Brown et al., 2015; Biggs, 2003; Entwistle, 2018), which similarly highlighted the multidimensional nature of deep learning, while extending previous work by capturing behaviors specific to AI-supported educational contexts.

Regarding reliability, the high Cronbach's alpha coefficients (ranging from 0.79 to 0.98) indicated that the scale is highly reliable, both at the level of each dimension and the overall scale, reflecting its suitability for confident use in various research and applied settings. Additionally, the psychological network analysis provided a dynamic view of the relationships among the items, showing clustering according to their theoretical dimensions and highlighting the most central items in the deep learning behavior network. Item 9 – related to critically verifying AI-provided information – emerged as the most central, being connected to multiple other deep learning behaviors, indicating that it serves as a key "link" in the deep learning approach. This was followed in importance by the ability to connect AI-provided ideas with prior knowledge and the enthusiasm for exploring new concepts when using AI, suggesting that these three traits may constitute core pillars for enhancing deep learning with AI support.

These core behaviors—critically verifying AI-provided information, connecting AI-provided ideas with prior knowledge, and actively exploring new concepts—reflect principles of self-regulated learning, constructivist knowledge construction, and metacognitive engagement. Item 9 illustrates students' monitoring and regulation of understanding (self-regulation and metacognition), while connecting ideas aligns with constructivist integration of new knowledge, and the enthusiasm for exploring concepts demonstrates proactive, goal-directed engagement typical of deep learning in AI-supported contexts.

Based on the above, these results indicate that the scale is not only a precise measurement tool but also contributes to explaining the dynamics of deep learning within AI-supported environments. Consequently, educators and researchers can benefit from this tool in designing interventions that focus on pivotal behaviors—such as critical thinking, connecting ideas, and active engagement with concepts—due to their cascading effect on other components of deep learning, thereby enhancing students' use of AI technologies in a deeper and more responsible manner.

In practical terms, the Deep Learning Scale can guide educators in designing AI-integrated curricula and targeted learning activities. By assessing students' abilities to critically verify AI-generated information, connect new ideas with prior knowledge, and actively explore concepts, teachers can identify areas for intervention and foster behaviors aligned with self-regulated learning, constructivist knowledge construction, and metacognitive engagement.

CONCLUSION

In conclusion, the developed scale (in its Arabic version) has demonstrated strong psychometric properties in terms of validity and reliability, making it a trustworthy tool for assessing students' adoption of a deep learning approach when using generative AI tools. Theoretically, this study extends the understanding of deep learning dynamics in AI-supported educational environments; practically, it provides educators and researchers with a validated instrument to assess and enhance students' engagement in deep learning behaviors and offers a context-specific contribution by providing one of the first validated measures of deep learning behaviors within AI-supported Arabic educational settings.

Its significance lies in providing researchers and educators with a practical means to examine the factors that support deep learning in the era of smart technology, whether through designing AI-based learning activities, evaluating the effectiveness of critical thinking development programs, or comparing learning patterns across diverse cultural samples and contexts. Utilizing this scale contributes to guiding the use of smart technologies toward deeper and more effective learning and fosters the development of critical and creative learners with sustainable understanding aligned with the objectives of modern education.

LIMITATIONS

Despite the strong psychometric properties demonstrated by the developed scale in terms of validity and reliability. The scale was validated only on university students, which may limit its generalizability to other educational levels or cultural contexts. Additionally, relying solely on self-report questionnaires could introduce bias in students' responses, potentially affecting the accuracy of measured deep learning behaviors.

SUGGESTIONS FOR FUTURE RESEARCH

Future studies should examine the scale's applicability across diverse populations and educational contexts. Incorporating qualitative methods, such as interviews or observations, could provide a deeper understanding of deep learning dynamics in AI-supported learning environments and further validate the instrument.

ACKNOWLEDGEMENT

The authors would like to express their sincere thanks to all the research participants.

FUNDING

The authors received no financial support for the research, authorship, and/or publication of this article.

DISCLOSURE STATEMENT

The authors report there are no competing interests to declare.

REFERENCES

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education, 21*(1), 10. https://doi.org/10.1186/s41239-024-00444-7

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods. *Measurement and Evaluation in Counseling and Development, 47*(1), 79–86. https://doi.org/10.1177/0748175613513808

Biggs, J. B. (1987). *Student approaches to learning and studying*. Hawthorn, Victoria: Australian Council for Educational Research. (ISBN: 9780864310327)

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university* (3rd ed.). Maidenhead, UK: Open University Press.

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. https://doi.org/10.3389/fpubh.2018.00149

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Brown, S., White, S., Wakeling, L., & Naiker, M. (2015). Approaches and Study Skills Inventory for Students (ASSIST) in an introductory course in chemistry. *Journal of University Teaching & Learning Practice, 12*(3), Article 6. https://doi.org/10.53761/1.12.3.6

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Thousand Oaks, CA: SAGE. (ISBN: 978-1506341569)

Duffy, M. (2025). *AI vs learning: A deeper look at ChatGPT in academic writing.* Chartered Association of Business Schools. https://charteredabs.org/blogs/ai-vs-learning-a-deeper-look-at-chatgpt-in-academic-writing/

Entwistle, N. J., & McCune, V. (2004). The conceptual bases of study strategy inventories. *Educational Psychology Review, 16*(4), 325–345. https://doi.org/10.1007/s10648-004-0003-0

Entwistle, N. J., & Peterson, E. R. (2004). Conceptions of learning and knowledge in higher education: Relationships with study behaviour and influences of learning environments. *International Journal of Educational Research, 41*(6), 407–428. https://doi.org/10.1016/j.ijer.2005.08.009

Entwistle, N. J., & Ramsden, P. (1983). *Understanding student learning*. London, UK: Croom Helm.

Entwistle, N. J., Tait, H., & McCune, V. (2000). Patterns of response to an approaches to studying inventory across contrasting groups and contexts. *European Journal of Psychology of Education, 15*(1), 33–48. https://doi.org/10.1007/BF03173165

Entwistle, N., McCune, V., & Tait, H. (2013). *Approaches and Study Skills Inventory for Students (ASSIST): Report of the development and use of the inventories* (3rd ed.). University of Edinburgh. (Report; available via institutional/author repositories.) Example access: https://www.researchgate.net/publication/50390092

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Essien, A., Bukoye, O. T., O'Dea, C., & Kremantzis, M. (2024). The influence of AI text generators on critical thinking skills in UK business schools. *Studies in Higher Education*. Advance online publication. https://doi.org/10.1080/03075079.2024.2363151

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*. Advance online publication. https://doi.org/10.1111/bjet.13544

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 61*(5), 460–474. https://doi.org/10.1080/14703297.2023.2193790

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

Gezgin, D. M. (2024, April 4–7). Chatbot use in education: A study on deep and surface learning approaches of university students. In *11th Balkan International Conference on Social Sciences (Balkan ICOSS)*, Skopje, North Macedonia. Academy Global Publishing House. (ISBN: 978-625-6830-98-1)

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Andover, UK: Cengage. (ISBN: 978-1473756540)

Hakiki, M., Fadli, R., Samala, A. D., Fricticarani, A., Dayurni, P., Rahmadani, K., Astiti, A. D., & Sabir, A. (2023). Exploring the impact of using Chat-GPT on student learning outcomes in technology learning: The comprehensive experiment. *Advances in Mobile Learning Educational Research, 3*(2), 859–872. https://doi.org/10.25082/AMLER.2023.02.013

Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: A synthesis and conceptual model. *npj Science of Learning, 1*, 16013. https://doi.org/10.1038/npjscilearn.2016.13

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

Iqbal, J., Hashmi, Z. F., Asghar, M. Z., & Abid, M. N. (2025). Generative AI tool use enhances academic achievement in sustainable education through shared metacognition and cognitive offloading among preservice teachers. *Scientific Reports, 15*, 16610. https://doi.org/10.1038/s41598-025-16610-8

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: The Guilford Press. (ISBN: 978-1-4625-2334-4)

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology, 46*(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Nhat, T. N. M., & Van Le, T. (2023). Effects of reflective thinking on deep learning in theoretical linguistics classes. International Journal of Instruction, 16(4), 101-120. https://doi.org/10.29333/iji.2023.1647a

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). Higher education assessment practice in the era of generative AI tools. *Journal of Applied Learning & Teaching, 7*(1), 46–56. https://doi.org/10.37074/jalt.2024.7.1.28

Rahimi, A. R., & Mosalli, Z. (2025). Language learners' surface, deep, and organizing approaches to ChatGPT-assisted language learning: What contextual, individual, and ChatGPT-related factors contribute? *Smart Learning Environments, 12*, 16. https://doi.org/10.1186/s40561-025-00368-3

Rosen, L. D., Whaling, K., Carrier, L. M., Cheever, N. A., & Rokkum, J. (2013). The Media and Technology Usage and Attitudes Scale: An empirical investigation. *Computers in Human Behavior, 29*(6), 2501–2511. https://doi.org/10.1016/j.chb.2013.06.006

Tamire (Negash), T., Eshete, M. T., & Ataro, G. (2022). Students' learning approaches as a factor of academic achievement at selected public universities: A cross-sectional study. *Frontiers in Education, 7*, 806454. https://doi.org/10.3389/feduc.2022.806454

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Švacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education, 215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

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. https://doi.org/10.1177/0013164413495237

Yang, Y., Luo, J., Yang, M., Yang, R., & Chen, J. (2024). From surface to deep learning approaches with Generative AI in higher education: An analytical framework of student agency. *Studies in Higher Education, 49*(5), 817–830. https://doi.org/10.1080/03075079.2024.2327003

Yating, Z., Nasri, N. M., & Jamaludin, K. A. (2025). Factors influencing deep learning in tourism consumer behaviour courses: A study based on constructivism and reciprocal determinism. International Journal of Instruction, 18(2), 69-84. https://doi.org/10.29333/iji.2025.1825a

Zhang, Z., Geldsetzer, P., Chen, S., & Bärnighausen, T. (2025). Effects of generative artificial intelligence on cognitive effort and task performance: A study protocol for a randomized controlled experiment among college students. *Trials, 26*, 244. https://doi.org/10.1186/s13063-025-06524-1