Computers & Education 254 (2026) 105700

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Educational prompt engineering self-efficacy scale (Ed-PESS): Development and psychometric validation

Fatih Karatas [a,*], Recep Gür [b], Barış Erïçok [c], Fatma Başarır [a], Lokman Tanrikulu [d]

[a] Nevsehir Haci Bektas Veli University, School of Foreign Languages, Nevsehir, Türkiye
[b] Osman Gazi University, Faculty of Education, Eskişehir, Türkiye
[c] Ordu University, Faculty of Education, Department of Educational Sciences, Ordu, Türkiye
[d] Nevsehir Haci Bektas Veli University, Faculty of Education, Nevsehir, Türkiye

ARTICLE INFO

Keywords:
Prompt engineering
Educational prompt engineering
Prompt engineering self-efficacy
AI
Scale development

ABSTRACT

Pre-service teachers' confidence in designing pedagogically purposeful prompts for generative AI tools remains an underexamined construct in teacher education. This study developed and validated the Educational Prompt Engineering Self-Efficacy Scale for Pre-service Teachers (Ed-PESS) to address the absence of a psychometrically sound instrument measuring this domain. Scale development followed Boateng et al.'s three-phase framework, grounded in a systematic literature review and expert review procedures. Two independent samples of pre-service teachers were recruited: an exploratory factor analysis (EFA) sample (n = 300) and a confirmatory factor analysis (CFA) sample (n = 305). EFA using principal axis factoring with Promax rotation produced a four-factor structure explaining 64.60% of total variance. CFA with Yuan-Bentler robust correction confirmed acceptable model fit. The final 26-item scale comprises four dimensions: Basic Prompt Engineering, Pedagogical Prompt Engineering, Ethical Prompting Practice, and Continuous Professional Development. Internal consistency was high across all subscales. The Ed-PESS provides a valid, reliable instrument for assessing prompt engineering self-efficacy in pre-service teacher education. It supports formative curriculum evaluation and targeted intervention design.

1. Introduction

With the widespread adoption of AI applications, fundamental aspects of teaching including lesson planning, content development, personalized feedback, and interactive learning experiences are being reshaped (Serra & Oliveira, 2025). Within this context, AI integration is increasingly recognized as a strategic priority for educational institutions (Pratschke, 2024), gaining growing attention at the institutional level in higher education (Lee & Palmer, 2025).

Despite this growing integration, AI's pedagogical value depends not on the technology itself but on the quality of user-AI interaction. Across teacher education and higher education settings, studies have demonstrated that the pedagogical quality of AI outputs depends directly on the quality of the prompts users provide (Celik et al., 2026; Jacobsen & Weber, 2025; Moorhouse et al., 2025). A

* Corresponding author.
E-mail addresses: fatih.karatas@nevsehir.edu.tr (F. Karatas), math.recepgur@gmail.com (R. Gür), barisericok@odu.edu.tr (B. Erïçok), basarirfatma@nevsehir.edu.tr (F. Başarır), ltanrikulu@nevsehir.edu.tr (L. Tanrikulu).

https://doi.org/10.1016/j.compedu.2026.105700
Received 10 March 2026; Received in revised form 14 May 2026; Accepted 23 June 2026
Available online 10 July 2026

prompt is a text-based instruction provided to an AI system to generate a response (Arocha, 2025; Correia et al., 2025). Crosthwaite et al. (2025) indicate that effective prompts are aligned with learning objectives, contextually rich, and well structured, whereas ineffective prompts consist of vague or overly simplistic commands. Consequently, the ability to craft effective prompts has emerged as a critical competency for AI use in education, and the conceptual framework for purposefully structuring this interaction is referred to as educational prompt engineering (Agirdag, 2025).

Contemporary generative AI tools such as ChatGPT, Claude, and Gemini are built on large language models (LLMs) trained on extensive text corpora through deep learning architectures. These models do not retrieve pre-stored responses; rather, they generate output by predicting the most statistically probable sequence of tokens given the input provided by the user (Kang et al., 2025; Park & Choo, 2024). This technical characteristic explains why prompt design known also as prompt engineering has emerged as a distinct competency rather than a peripheral usage skill. Recognizing this distinction, researchers have sought to formalize the conceptual boundaries of prompt engineering by distinguishing it from broader AI literacy frameworks and situating it within systematic definitional structures (Federiakin et al., 2024; Qian, 2025). In the literature, prompt engineering is defined within a rigorous computational and semantic framework as the systematic structuring and refinement of user inputs to obtain desired outputs from AI systems (Lee & Palmer, 2025; Pratschke, 2024). Although prompt engineering is closely related to AI literacy, which encompasses general competencies such as understanding and evaluating AI technologies (Long & Magerko, 2020; Ng et al., 2021), it represents a distinct, practice-oriented application of this broader competence, enabling teachers to take concrete actions toward their pedagogical goals and obtain purpose-aligned outputs from AI tools (Celik et al., 2026; Kang et al., 2025). While the concept was initially developed in technical contexts, the need for domain-specific strategies (Ognibene et al., 2025) has led to the emergence of a new competency area termed educational prompt engineering, which intersects with pedagogical expertise in education (Correia et al., 2025; Tümen Akyıldız, 2026).

Prompt engineering is a teachable and developable skill; therefore, its pedagogy can be empirically investigated (Lee & Palmer, 2025). Recent empirical research demonstrates that structured prompt engineering training significantly influences learners' prompting strategies and learning outcomes. Woo et al. (2025) found that students used prompt engineering not merely as a technical operation but as a tool supporting cognitive processes such as overcoming writer's block and developing story plots. In the context of teacher education, Sigot and Tassoti (2025) observed that through a structured prompting framework, students transitioned from simple search-engine-like usage to strategic and iterative prompting practices. Similarly, Hwang et al. (2025) reported that students without guidance focused on superficial linguistic features and struggled to construct goal-aligned prompts. Moreover, AI-supported instructional practices and prompt engineering interventions have been shown to strengthen task-specific self-efficacy beliefs (Andewi et al., 2025; Huang et al., 2024; Woo et al., 2025), support self-regulated learning (Mzwri & Turcsányi-Szabó, 2025), foster creativity and critical thinking (Kabeer et al., 2025), and reduce technology anxiety (Davila-Moran et al., 2025). Taken together, these findings suggest that prompt engineering skills do not develop spontaneously but require intentional, pedagogically grounded instruction. This highlights the importance of systematically integrating such training into teacher education curricula.

However, the effectiveness of such training also depends on teachers' confidence in applying these skills (Celik et al., 2026; Davila-Moran et al., 2025; Tümen Akyıldız, 2026). This points to the need for a construct that captures teachers' confidence in designing and refining pedagogically grounded prompts. Given that these competencies and beliefs are predominantly formed during pre-service education, teacher education curricula should move beyond technical usage to also address the development of educational prompt engineering self-efficacy (Moorhouse et al., 2025; Serra & Oliveira, 2025). Accordingly, scholars argue that pedagogically productive interaction with AI requires the holistic integration of pedagogical content knowledge, AI knowledge, and prompting skills (Correia et al., 2025; Moorhouse et al., 2025). Drawing on Bandura's (2006) framework, which conceptualizes self-efficacy as task- and context-specific, pre-service teachers' educational prompt engineering self-efficacy can be defined as their subjective belief in their capability to create pedagogically meaningful prompts, integrate these into teaching and learning processes, and sustain this interaction within an evaluation-refinement cycle (Crosthwaite et al., 2025).

For future teachers, prompt engineering self-efficacy is emerging as an increasingly important dimension of professional belief that develops alongside established competencies such as lesson planning and assessment design (Celik et al., 2026; Davila-Moran et al., 2025; Tümen Akyıldız, 2026). Within the professional roles that pre-service teachers will assume, traditional teaching competencies such as lesson planning, assessment, and instructional material development are expected to be carried out in integration with AI tools (ElSayary et al., 2025; Park & Choo, 2024). This integration, however, is not straightforward. This suggests that pre-service teachers need to develop not only technical familiarity with AI tools but also confidence in directing them toward pedagogical purposes. Correia et al. (2025) further frame teachers' evolving roles as ‘prompt engineers' within UNESCO's AI competency framework, underscoring the importance of cultivating these competencies during initial teacher training.

In the specific context of teacher education, other studies examined the use of prompt engineering strategies within the Intelligent TPACK framework (Celik et al., 2026), the mediating role of disciplinary literacy (Kang et al., 2025), and the relationship with pedagogical creative thinking (Tümen Akyıldız, 2026). While these studies reveal the relationship between prompt engineering and pedagogical practices, they have predominantly been framed around general AI self-efficacy or perceptions toward technology, rather than directly focusing on a distinct, task- and context-specific self-efficacy construct unique to prompt engineering in pedagogical settings. As a learnable and developable capacity, this construct is also directly measurable, which in turn necessitates valid and reliable instruments for its systematic assessment.

1.1. Need for Educational Prompt Engineering Self-Efficacy Scale (Ed-PESS)

The growth of generative AI in educational contexts has led to the development of various measurement instruments targeting

related constructs. Existing instruments include the AI Literacy Scale (AILS; Wang et al., 2023), the AI Self-Efficacy Scale (Wang & Chuang, 2024), the Teacher AI Competence Self-Efficacy Scale (TAICS; Chiu et al., 2025), and the Intelligent-TPACK framework (Celik, 2023), which together measure general AI literacy, general AI use confidence, broad teacher AI competence, and AI-extended pedagogical content knowledge. Despite these contributions, significant limitations constrain their applicability to measuring prompt engineering or prompt engineering self-efficacy among pre-service teachers. First, most instruments were developed before the emergence of generative AI and prompt engineering as distinct constructs, focusing instead on general technology acceptance and use (Laupichler et al., 2023). Second, existing measures emphasize general AI interaction capabilities rather than the specific belief structures underlying confidence in prompt design for pedagogical purposes. Third, none of these instruments capture the iterative, dialogic nature of prompt engineering confidence, which requires educators to engage in cycles of prompt formulation, output evaluation, and refinement to achieve instructionally appropriate responses (Celik et al., 2026; Tümen Akyıldız, 2026).

Besides these studies, most recently and most closely related to the present study, Gibreel and Arpaci (2025) developed the Prompt Engineering Competence Scale (PECS), a unidimensional nine-item instrument measuring users' general proficiency in prompt engineering. The PECS represents a significant contribution by establishing prompt engineering as a measurable construct. However, critical distinctions limit its applicability to education. First, the PECS measures competence rather than self-efficacy, yet self-efficacy beliefs are more predictive of behavioral engagement than objective skill levels (Bandura, 2006). Second, the instrument was developed with general AI users, lacking pedagogical specificity essential for understanding how teachers design prompts to achieve instructional objectives. These limitations reveal a clear methodological gap, specifically the absence of a psychometrically validated instrument designed to measure educational prompt engineering self-efficacy for pre-service teachers. The present study addresses this gap through the development and validation of the Educational Prompt Engineering Self-Efficacy Scale for Pre-service Teachers (Ed-PESS). While educational policy increasingly requires AI integration in teacher preparation (Kang et al., 2025; Tümen Akyıldız, 2026), no valid instrument exists to determine whether teacher candidates believe they can effectively engineer prompts to bring out pedagogically appropriate AI responses. The consequences of this problem are far-reaching, as teacher training curricula cannot systematically identify students with low prompt engineering self-efficacy, evaluate the impact of instructional interventions, or detect prompt-related anxiety that may slow down effective AI utilization.

Self-efficacy theory (Bandura, 2006) provides the theoretical foundation for the Ed-PESS. Bandura (2006) argues that perceived self-efficacy is not a global trait but a differentiated set of beliefs tailored to the domain of functioning, and that sound efficacy scales must align with the cognitive, behavioral, and self-regulatory demands of that domain. Because behavior is better predicted by beliefs in the full range of capabilities required to succeed than by beliefs in a single aspect of the domain (Bandura, 2006), the Ed-PESS operationalizes educational prompt engineering self-efficacy as a multifaceted construct whose dimensions correspond to the distinct task demands that pre-service teachers face when using generative AI for instructional purposes. Accordingly, Basic Prompt Engineering captures confidence in the operational syntax of prompting that develops through mastery experiences and supplies the procedural base for higher-order applications (Bandura, 2006); Pedagogical Prompt Engineering captures domain-bound, task-specific self-efficacy for translating AI outputs into instructionally appropriate resources such as lesson plans, learning materials, and differentiated assessments; Ethical Prompting Practice captures confidence in activating self-regulatory safeguards in contexts where responsibility may be diffused onto the AI tool, grounded in Bandura's (2001) account of moral agency; and Continuous Professional Development captures self-efficacy for sustained adaptation and self-directed learning under rapidly changing tool conditions, consistent with self-regulation theory (Zimmerman, 2002). Consistent with Bandura's (2006) "can do" formulation, every Ed-PESS item was phrased as a capability judgment, and the four dimensions were derived a priori from theory rather than extracted solely from data, which strengthens the construct validity foundation of the instrument.

This theoretical foundation, combined with the measurement gap identified above, defines the contribution of the present study. The Ed-PESS offers the first psychometrically validated instrument designed to measure educational prompt engineering self-efficacy among pre-service teachers, extending the unidimensional conceptualization of prompt engineering competence offered by the PECS (Gibreel & Arpaci, 2025) to a multidimensional self-efficacy construct that encompasses operational, pedagogical, ethical, and developmental facets. Accordingly, the purpose of this study is to develop the Educational Prompt Engineering Self-Efficacy Scale for Pre-service Teachers (Ed-PESS) and to provide evidence of its validity and reliability. To ensure theoretical validity, the instrument's development was grounded in a systematic literature review and thematic analysis, with every item derived from and supported by direct evidence from the literature. To this end, the following research questions were addressed.

(i) What is the factor structure of the Ed-PESS based on exploratory factor analysis?
(ii) To what extent does confirmatory factor analysis support the factor structure of the Ed-PESS?
(iii) What is the evidence for the reliability and convergent validity of the Ed-PESS and its subscales?

2. Methods

2.1. Research design

This study employed an instrument development and validation design to develop the Educational Prompt Engineering Self-Efficacy Scale for Pre-service Teachers (Ed-PESS). The scale development process was guided by Boateng et al.'s (2018) three-phase, nine-step framework: item development (domain identification and item generation; content validity), scale development (question pre-testing, survey administration and sampling, item reduction, and latent factor extraction), and scale evaluation (tests of dimensionality, reliability, and validity). In Phase 1, the construct domain was defined and an initial item pool was created based on a

systematic review of the prompt engineering and AI in teacher education literature. Expert review and pilot testing with the target group were then used to improve item relevance and clarity. In Phase 2, the draft scale was administered to Sample 1 of pre-service teachers, where item reduction procedures and exploratory factor analysis were used to identify the latent structure. In Phase 3, the retained items were tested in an independent Sample 2 (n = 305) using confirmatory factor analysis.

2.2. Participants

Given the inherent requirements of the scale development process, this study was conducted with two independent participant groups. Participant recruitment was shaped by time and cost constraints and by the researchers' accessibility to pre-service teachers across multiple institutions; accordingly, a convenience sampling approach was adopted. To mitigate the risk of clustering within a single subgroup and to obtain a configuration more representative across key demographic variables, deliberate efforts were made to reach pre-service teachers from a range of year levels, ages, genders, program types, and universities, so that both study groups would display the greatest feasible heterogeneity within the limits of the sampling frame. The samples used for the exploratory factor analysis (EFA) and the confirmatory factor analysis (CFA) were recruited independently of one another, consistent with Boateng et al.'s (2018) recommendation that the two phases draw on separate participant groups.

In the first phase, data were collected from 327 pre-service teachers to examine the construct validity of the 32-item draft scale through EFA. Following univariate (z-score, ±3) and multivariate (Mahalanobis distance) outlier screening, the final EFA sample comprised $n_1 = 300$ participants. During the EFA-based item reduction procedure, six items were removed based on two pre-specified psychometric criteria: (a) factor loadings below .40 (Items 1, 13, and 14) and (b) cross-loadings on more than one factor with a loading difference smaller than .10 (Items 7, 25, and 29). Items were removed iteratively rather than simultaneously, with EFA re-run after each removal. Items 1, 13, and 14 were removed first in order of ascending factor loading, followed by Items 7, 25, and 29 in order of decreasing cross-loading severity. The resulting 26-item, four-factor structure was subsequently administered to an independent second sample for CFA. For the CFA phase, a separate sample of 341 pre-service teachers was recruited after the completion of EFA-based item reduction, ensuring that the revised 26-item structure was tested on data fully independent of the exploratory phase. The CFA data set contained no missing values; five univariate outliers ($|z| > 3$) and 31 multivariate outliers (based on Mahalanobis distance) were removed, yielding a final CFA sample of $n_2 = 305$ participants. The participant-to-item ratio exceeding 10:1 (305:26) provides evidence for adequate sample size for CFA. Detailed demographic characteristics of both groups are presented in Table 1.

As seen in Table 1, Both samples showed a gender imbalance reflecting Turkish teacher education demographics (~73% female). Limited prior AI training (<13% in Group 1, <7% in Group 2) suggests the scale was validated under ecologically representative conditions, though future studies should examine whether self-efficacy scores are attenuated by limited prior exposure.

Table 1
Demographic characteristics of study group 1 (EFA) and study group 2 (CFA).

Variable	Category	Study Group 1 (EFA, $n_1 = 300$)	Study Group 2 (CFA, $n_2 = 305$)
Gender			
	Female	219 (73.0%)	224 (73.4%)
	Male	81 (27.0%)	81 (26.6%)
Age			
	17–18		—
	19		66 (21.6%)
	20	56 (18.7%)	74 (24.3%)
	21	79 (26.3%)	
	22 and above		
	Range	18–40	17–38
Year of Study			
	1st Year		101 (33.1%)
	2 nd Year	83 (27.7%)	127 (41.6%)
	3rd Year	113 (37.7%)	
	4th Year		
Academic Program (Top Programs)			
	English Language Teaching	24.0%	9.2%
	German Language Teaching	16.0%	—
	Social Studies Teaching	—	9.5%
Primary AI Usage Purpose			
	Homework & Presentation Prep	69 (23.0%)	31.1%
AI Usage Duration			
	1–2 years	36.4%	30.5%
AI Training Status			
	Received prior AI training	37 (12.3%)	20 (6.6%)
	No training; willing to receive	181 (60.3%)	140 (45.9%)
	No training; not interested	79 (26.3%)	141 (46.2%)

Note. EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis; HBV = Hacı Bektaş Veli. Percentages may not sum to 100 due to rounding. Dashes (−) indicate data not separately reported for that group.

2.3. Instrument development and procedure

Data were collected via an online survey administered to pre-service teachers between May and November 2025. The research team conducted a preparatory meeting to standardize the administration procedure and reviewed the participant briefing, procedural steps, and key administration considerations in detail. The research team then contacted each participating institution, shared the ethics committee approval document, and obtained the necessary institutional permissions prior to recruitment. At each institution, members of the research team conducted brief face-to-face information sessions with pre-service teachers in which the purpose of the study, the voluntary and anonymous nature of participation, the absence of right or wrong answers, and the importance of completing the instrument in an environment free from distractions (e.g., excessive noise, inadequate lighting) were explained. The data collection instrument was subsequently shared with participants through an online form link. Participation was voluntary, anonymous, and preceded by informed consent, with no incentives offered. Following a systematic literature review and researcher discussions, an initial pool of 38 items was drafted in accordance with 5 Likert-type scale development conventions, with ambiguous expressions deliberately avoided. Items were reviewed by three Measurement and Evaluation specialists, two Computer Education and Instructional Technology specialists, and two Turkish Language specialists for language, content, and psychometric quality; applying Lawshe's (1975) CVR criterion of .80 at $p < .05$, six items falling below this threshold were removed, and the remaining items were revised to yield a 32-item draft instrument comprising four subscales: Basic Prompt Engineering (9 items), Pedagogical Prompt Engineering (7 items), Ethical Prompting Practice (9 items), and Continuous Professional Development (7 items).

2.4. Data analysis

Data analysis was carried out in three stages: data screening, construct validity analyses, and reliability analyses. Prior to conducting the analyses, the normality assumption of the dataset was evaluated using Mardia's coefficients. Additionally, z-scores (±3 threshold) were employed for univariate outlier detection, and Mahalanobis distance values were used for multivariate outlier identification. To provide evidence of construct validity and to determine the factor structure of the scale, EFA was performed as the initial analytical step. The suitability of the dataset for factorization was assessed using the Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's Test of Sphericity. The KMO value was found to be .942, indicating excellent sampling adequacy, and Bartlett's test results were statistically significant, χ^2 (325) = 6173, $p < .001$. Principal Axis Factoring was selected as the extraction method, and given the presence of inter-factor correlations (r = .507 to .645), Promax rotation, an oblique rotation method, was employed. The number of factors to retain was determined through convergent evaluation of multiple complementary criteria: the Kaiser criterion (eigenvalues >1), visual inspection of the scree plot, Horn's parallel analysis, optimal parallel analysis, Velicer's Minimum Average Partial (MAP) test, and the Hull method. When these criteria yielded discrepant suggestions, candidate solutions were compared empirically through model fit indices (RMSEA, TLI, BIC) and explained variance to identify the most appropriate factor structure (Lorenzo-Seva et al., 2011; Preacher et al., 2013). CFA was subsequently applied to confirm the structure identified through EFA. Because the data did not conform to a multivariate normal distribution, the Maximum Likelihood (ML) estimation method was used in conjunction with Yuan-Bentler (robust) correction. Model fit was evaluated using the following indices: Robust χ^2/df, RMSEA (Root Mean Square Error of Approximation), SRMR (Standardized Root Mean Square Residual), CFI (Comparative Fit Index), and TLI (Tucker-Lewis Index). To provide evidence of reliability, Cronbach's alpha (α), McDonald's omega (ω), and the composite reliability (CR) coefficient were computed to assess the internal consistency of both the total scale and its subscales. Average Variance Extracted (AVE) values were examined as evidence of convergent validity. All analyses were performed using the JAMOVI statistical software.

2.5. Measures: item generation and development

The item development process followed a theoretically rigorous approach supported by a systematic literature review to identify the construct's dimensional structure. Subsequently, thematic analysis of core papers established the framework for item generation, ensuring that each item was based on specific empirical or theoretical evidence. This literature-grounded methodology ensures initial content validity. Following Boateng et al.'s (2018) scale development framework, the systematic search was conducted in “Web of Science” and “Scopus” databases to establish the theoretical foundation for item development. The review identified core papers addressing prompt engineering in educational contexts, from which principal themes were extracted through thematic analysis. Based on this synthesis, educational prompt engineering self-efficacy was operationally defined as a multidimensional construct comprising pre-service teachers' confidence in designing, implementing, and evaluating prompts for generative AI tools in pedagogically appropriate ways. This conceptualization integrates Bandura's (2006) self-efficacy theory with domain-specific competencies identified in the literature.

The systematic literature review revealed four recurring themes that formed the theoretical architecture for item development: (1) technical prompt construction skills, (2) pedagogical integration of AI outputs, (3) ethical considerations in educational AI use, and (4) ongoing professional learning for AI integration. These themes were marked as candidate dimensions, with each theme's basic elements serving as the framework for item generation. Items were written as first-person self-efficacy statements (e.g., “I can … ") using clear language appropriate for pre-service teachers, following established guidelines (Boateng et al., 2018). Each item was supported by direct citations from the literature to ensure theoretical validity (see Appendix 1 for all items and literature evidence). An initial pool of 32 items was generated, distributed across the four thematically-derived dimensions: Basic Prompt Engineering (9 items), Pedagogical Prompt Engineering (7 items), Ethical Prompting Practice (9 items), and Continuous Professional Development (7 items). The pilot study was conducted with a diverse sample of 22 participants to evaluate the reliability and structural integrity of the

'Pedagogical Prompt Engineering Self-Efficacy Scale.' The participant group was primarily composed of female students (77.3%) and represented a wide spectrum of ten distinct teacher education programs, including English Language Teaching, Preschool Education, and Psychological Counseling, thereby ensuring the instrument's applicability across varied pedagogical contexts. Statistical analysis confirmed the scale's robust internal consistency, yielding a Cronbach's Alpha coefficient of .961. Detailed item-total correlation analysis provided critical insights for instrument refinement; specifically, four items (Items 1, 6, 9, and 33) were found to have correlation coefficients below .30. These items were subsequently modified to improve linguistic clarity and conceptual alignment with the overall construct. These statistical insights and the resulting minor modifications ensured that the final version of the scale is both highly reliable and valid for the main study. Building on this thematic architecture, the four dimensions of the Ed-PESS are described below in terms of their conceptual scope, theoretical grounding, and item coverage.

Basic Prompt Engineering (BPE). This dimension represents the technical capacity to structure clear, context-aware instructions using specific prompt components and to iteratively refine inputs. The theoretical grounding derives from the CLEAR framework (Lo, 2023), the PARTS framework (Park & Choo, 2024), and empirical work on prompt literacy (Hwang et al., 2025). Seven items capture competencies including prompt construction, component specification, role assignment, and systematic documentation.

Pedagogical Prompt Engineering (PPE). This dimension represents the capacity to design prompts aligned with learning objectives, differentiate instruction, and critically evaluate AI outputs. It is grounded in the TPACK framework (Mishra & Koehler, 2006) and its AI-specific extension, Intelligent-TPACK (Celik, 2023). The items address learning objective alignment, differentiated content, critical evaluation, and engagement strategies.

Ethical Prompting Practice (EPP). This dimension represents the capacity to protect student data privacy, mitigate algorithmic bias, and model ethical AI use. Walter (2024) positions ethical discussions as integral to AI-integrated learning, while Hsu (2025) documents data privacy concerns in GenAI interactions. Mzwri and Turcsányi-Szabó (2025) articulate that prompt engineering guides outputs toward fairness and inclusivity. Nine items address privacy protection, bias mitigation, source verification, and ethical modeling for students.

Continuous Professional Development (CPD). This dimension represents the disposition to monitor AI advancements and seek ongoing skill improvement. Foundationally grounded in Schön's reflective practice framework and supported by ElSayary et al. (2025), this dimension recognizes that rapid AI evolution requires continuous updating (Lee & Palmer, 2025; Walter, 2024). Woo et al. (2025) demonstrated that targeted workshops significantly improve prompt engineering ability. The items address awareness, adaptation, feedback-driven refinement, and training participation.

3. Results

3.1. Item analysis results

Item analysis was conducted to evaluate the descriptive properties and discrimination indices of the scale items. Item means and standard deviations for each item, together with the upper 27 percent versus lower 27 percent comparisons, are provided in Appendix 2. To test item discrimination, item scores of the upper 27 percent (n = 81) and the lower 27 percent (n = 81) of the total

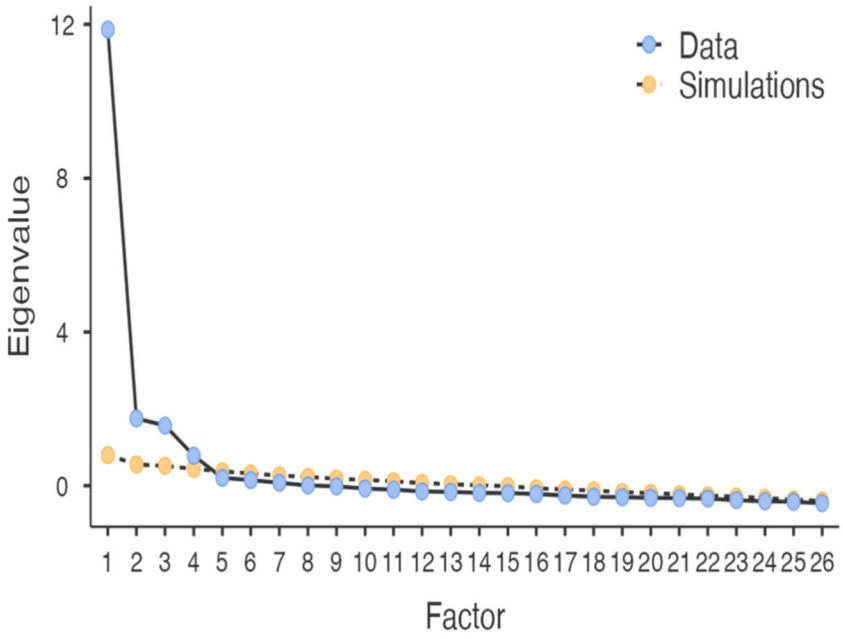

Fig. 1. Parallel analysis.

score distribution were compared via independent-samples t-tests. All 32 items showed statistically significant differences in favor of the upper group ($p < .05$), indicating high item discrimination.

3.2. Exploratory factor analysis results

As the initial step in the scale development process, the multivariate normality assumption of the dataset was tested using Mardia's coefficients. The statistically significant skewness (117) and kurtosis (.888) values ($p < .001$) indicated that the data did not conform to a multivariate normal distribution. Accordingly, robust estimation methods were employed throughout the analyses.

Factor retention was evaluated through convergent assessment of six complementary criteria. The Kaiser criterion (eigenvalue >1), Horn's parallel analysis (see Fig. 1), and the scree plot (see Fig. 2) supported a four-factor structure, whereas Velicer's MAP test (average partial = .02408) and optimal parallel analysis (real-data variance = 7.02%; 95th percentile of random variance = 6.22%) indicated a three-factor solution, and the Hull method (GFI = .231, df = 464, scree test value = 1.904) suggested a unidimensional structure. Because the criteria yielded divergent suggestions, the three- and four-factor solutions were empirically compared. The three-factor solution showed inadequate fit (RMSEA = .097, TLI = .834, BIC = −509, χ^2 (273) = 1048, $p < .001$; 60.00% explained variance), whereas the four-factor solution showed acceptable fit, higher explained variance, and a lower BIC (RMSEA = .076, TLI = .902, BIC = −670, χ^2 (227) = 624, $p < .001$; 64.60% explained variance), with the lower BIC value (−670 < −509) providing further support for retaining the four-factor structure (Lorenzo-Seva et al., 2011; Preacher et al., 2013). The EFA findings for the four-factor scale are presented in Table 2.

The KMO value of .942 and Bartlett's Test of Sphericity, χ^2 (325) = 6173, $p < .001$, confirmed that the data were suitable for factor analysis. EFA conducted using principal axis factoring and Promax rotation yielded a four-factor structure in which all retained factors had eigenvalues exceeding 1. These four factors collectively accounted for 64.60% of the total variance. The individual contributions of each factor to the total variance were 22.20% (Basic Prompt Engineering), 15.30% (Continuous Professional Development), 15.00% (Ethical Prompting Practice), and 12.10% (Pedagogical Prompt Engineering), respectively. Item factor loadings ranged from .46 to .977, and inter-factor correlation coefficients ranged from .51 to .65. For the four-factor solution in the EFA sample (n = 300), the EFA-derived model fit indices, RMSEA = .076 [.069, .084], TLI = .902, and χ^2/df = 2.75 (χ^2 (227) = 624, $p < .001$), indicated an acceptable level of fit to the observed correlation matrix (Kline, 2011; Tabachnick & Fidell, 2013). These EFA-stage indices were computed within the exploratory sample to evaluate how well the extracted four-factor structure reproduced the observed item correlations; they are distinct from the CFA-derived robust fit indices reported in Section 3.3, which evaluate the same factor structure as a constrained measurement model in the independent confirmatory sample (n = 305).

3.3. Confirmatory factor analysis results

The 26-item, four-factor structure identified through EFA (Basic Prompt Engineering, Pedagogical Prompt Engineering, Ethical Prompting Practice, and Continuous Professional Development) was subsequently tested via CFA in the independent confirmatory sample (n = 305; see Fig. 3). Examination of the robust fit indices derived from Robust Maximum Likelihood (RML) estimation yielded the following values: χ^2/df = 2.03 (595/293), RMSEA = .064 [95% CI: .057, .071], SRMR = .057, CFI = .929, and TLI = .921, collectively indicating that the model demonstrated good fit to the data (Brown, 2015; Kline, 2011). Standardized factor loadings (β) ranged from .575 to .882, all statistically significant at $p < .001$ (see Table 3).

To examine whether the Ed-PESS measures the same construct equivalently across female and male pre-service teachers, in Table 4 measurement invariance was tested via multi-group confirmatory factor analysis (MG-CFA). Four nested models, configural, metric,

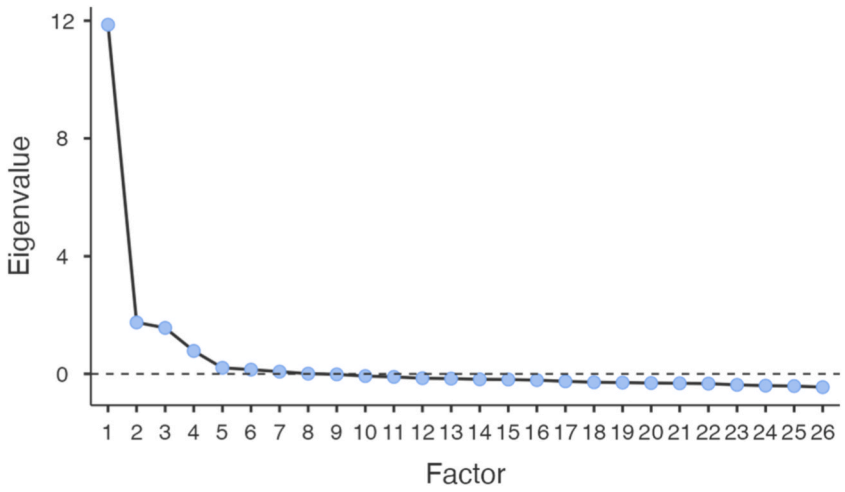

Fig. 2. Scree plot.

Table 2
Exploratory factor analysis results of the prompt engineering scale.

Items	Dimension Loadings				Factor Statistic		
	Factor 1	Factor 2	Factor 3	Factor 4	SS Loading	% of Variance	Cumulative %
Basic Prompt Engineering					3.91	15.00	52.50
item_2			.65				
item_3			.80				
item_4			.86				
item_5			.70				
item_6			.82				
item_8			.59				
item_9			.50				
Pedagogical Prompt Engineering					3.15	12.10	64.60
item_10				.83			
item_11				.97			
item_12				.77			
item_15				.70			
item_16				.46			
Ethical Prompting Practice					5.78	22.20	22.20
item_17	.69						
item_18	.66						
item_19	.86						
item_20	.85						
item_21	.98						
item_22	.81						
item_23	.83						
item_24	.70						
Continuous Professional Development					3.97	15.30	37.50
item_26		.79					
item_27		.81					
item_28		.76					
item_30		.77					
item_31		.82					
item_32		.77					

Inter-Factor Correlations	1	2	3
1. Basic Prompt Engineering			
2. Continuous Professional Development	.59		
3. Ethical Prompting Practice	.63	.51	
4. Pedagogical Prompt Engineering	.56	.59	.65

Note (see Table 3). N = 300. KMO = .942; Bartlett's Test of Sphericity $\chi^2(325) = 6173$, $p < .001$. Model fit measures of Exploratory Factor Analysis: RMSEA = .076 [.069, .084], TLI = .902, $\chi^2(227) = 624$, $p < .001$. Promax oblique rotation and principal axis factoring were used as the factor extraction and rotation methods, respectively. Factor loadings above .40 are bolded.

scalar, and strict, were tested sequentially, with invariance evaluated using the ΔCFI <.010 criterion (Cheung & Rensvold, 2002). As shown in Table 4, all four models demonstrated acceptable fit. The configural model provided baseline evidence that the four-factor structure held in both groups (χ^2 (586) = 1293.827, RMSEA = .070, SRMR = .056, Robust CFI = .921). ΔCFI values across successive constraints remained well below the .010 threshold (ΔCFI = .000 for metric and scalar; ΔCFI = .001 for strict), indicating that configural, metric, scalar, and strict invariance were all supported. These results provide evidence that the scale measures the same construct equivalently for female and male participants and that latent mean comparisons across gender are interpretable.

3.4. Reliability analysis results

Reliability and convergent validity coefficients for the EFA (n = 300) and CFA (n = 305) samples are presented in Table 5. Across both samples and all four subscales, Cronbach's α, McDonald's ω, and composite reliability (CR) coefficients exceeded .80, with the Ethical Prompting Practice subscale yielding the highest values (α = .949 in the EFA sample; α = .938, ω = .939, CR = .939 in the CFA sample). At the total-scale level, α and ω values exceeded .95 in both samples, and CR reached .975 and .973, respectively. AVE values exceeded the .50 threshold for the Pedagogical Prompt Engineering, Ethical Prompting Practice, and Continuous Professional Development subscales in both samples. For the Basic Prompt Engineering subscale, AVE values fell marginally below the threshold in both the EFA (.498) and CFA (.490) samples, whereas the corresponding α (.883/.865), ω (.889/.871), and CR (.871/.870) coefficients exceeded conventional thresholds (Fornell & Larcker, 1981; Hair et al., 2019).

The research findings collectively support the conclusion that the developed instrument possesses a valid and reliable structure. A KMO value exceeding .90 indicates that the sample size was at an "excellent" level for factorization. Item factor loadings ranging from .46 to .977, combined with the total variance explained by the scale exceeding the 60% threshold at 64.60%, provide compelling evidence that the factor structure effectively represents the measured construct. Turning to the CFA results, CFI and TLI values exceeding .90, alongside RMSEA and SRMR values falling below .08, confirm that the four-factor model is statistically congruent with

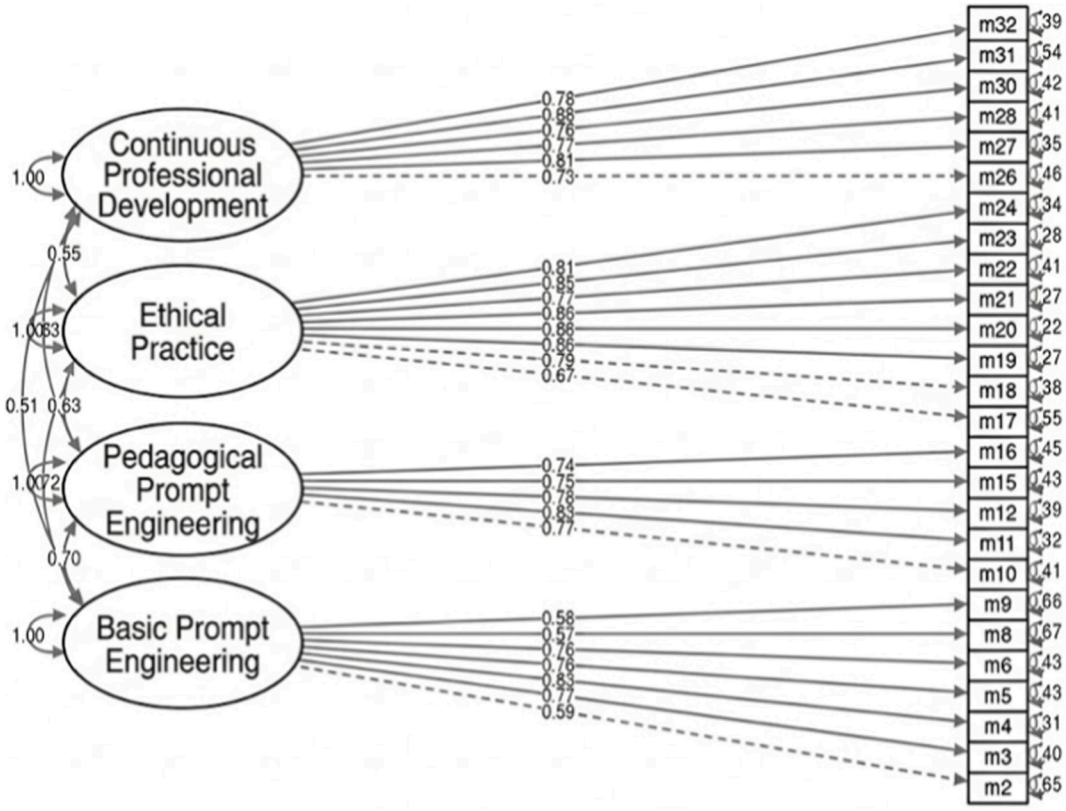

Fig. 3. Path diagram.

the theoretical structure. Inter-factor correlations ranging from .51 to .65 indicate that the subscales are interrelated yet measure conceptually distinct constructs. Examination of Cronbach's α, McDonald's ω, and CR values across all subscales further confirms that the scale is reliable. In conclusion, the 26-item, four-subscale 5-point Likert prompt engineering scale, comprising Basic Prompt Engineering, Pedagogical Prompt Engineering, Ethical Prompting Practice, and Continuous Professional Development, can be considered a valid and reliable instrument for measuring the relevant competencies of pre-service teachers.

4. Discussion

This study addressed a measurement gap in AI-integrated teacher education by developing and validating the Educational Prompt Engineering Self-Efficacy Scale for Pre-service Teachers (Ed-PESS). Overall, the findings provide initial validity evidence supporting the interpretation of Ed-PESS scores as a multidimensional measure of pre-service teachers' prompt engineering self-efficacy.

Converging EFA and CFA results supported a stable four-factor structure of the Ed-PESS: Basic Prompt Engineering, Pedagogical Prompt Engineering, Ethical Prompting Practice, and Continuous Professional Development. Reliability was high across dimensions, and convergent validity was generally adequate, although the Basic dimension showed borderline AVE, suggesting somewhat broader operational content. Moderate inter-factor correlations further support the interpretation of these dimensions as related but distinguishable facets of a broader prompt engineering self-efficacy construct.

Content-focused validity of the Ed-PESS was strengthened by a literature-grounded item development process, expert review, and pilot-based refinement. Together, these steps helped align the items with the pedagogical, ethical, and developmental demands of prompt engineering in teacher education.

Consistent with recent studies, the findings support prompt-related self-efficacy as a developable construct that can be strengthened through targeted instruction and meaningfully assessed in educational settings (e.g., Davila-Moran et al., 2025; Jang et al., 2025). At the same time, the four-factor structure suggests that prompt engineering self-efficacy in teacher education is not unitary but composed of related dimensions. Taken together, these results suggest that prompt engineering self-efficacy in teacher education should be understood not as a peripheral technical confidence, but as a pedagogically meaningful form of professional readiness because it concerns how future teachers structure AI-supported tasks, judge the instructional usefulness of outputs, and maintain ethical oversight in classroom-oriented use. In this respect, the Ed-PESS goes beyond the single-factor competence structure reported for general AI users by Gibreel and Arpaci (2025). This contribution is important because the Ed-PESS operationalizes prompt engineering self-efficacy as an education-specific construct, whereas prior AI-related measures discussed in this study have more often emphasized broader competence or general user-oriented dimensions rather than pedagogically purposeful prompting in teacher education.

Table 3
Confirmatory factor analysis results.

	B	*SE*(*B*)	*z*	*p*	*β*	*δ*
Basic Prompt Engineering						
item_2	1.00				.590	.652
item_3	1.34	.149	8.99	.000	.775	.400
item_4	1.49	.160	9.27	.000	.832	.309
item_5	1.42	.159	8.92	.000	.756	.428
item_6	1.30	.150	8.68	.000	.756	.429
item_8	1.05	.143	7.33	.000	.575	.669
item_9	1.14	.150	7.61	.000	.585	.658
Pedagogical Prompt Engineering						
item_10	1.00				.766	.413
item_11	1.12	.061	18.24	.000	.827	.317
item_12	.96	.067	14.40	.000	.779	.392
item_15	.91	.066	13.69	.000	.754	.432
item_16	.84	.068	12.32	.000	.744	.446
Ethical Prompting Practice						
item_17	1.000				.674	.545
item_18	1.179	.075	15.63	.000	.786	.383
item_19	1.271	.090	14.18	.000	.856	.267
item_20	1.319	.093	14.24	.000	.882	.221
item_21	1.311	.098	13.44	.000	.856	.268
item_22	1.182	.106	11.19	.000	.765	.414
item_23	1.199	.089	13.41	.000	.847	.282
item_24	1.155	.091	12.72	.000	.811	.342
Continuous Professional Development						
item_26	1.000				.733	.463
item_27	1.048	.058	18.14	.000	.809	.346
item_28	.923	.068	13.56	.000	.771	.406
item_30	1.003	.071	14.23	.000	.758	.425
item_31	.965	.076	12.68	.000	.676	.544
item_32	1.033	.073	14.16	.000	.779	.393

Note. B: Unstandardized factor loading; *SE*(*B*): Standard error of the unstandardized factor loading; *β*: Standardized factor loading; *δ*: Standardized error variance; AVE: Average Variance Extracted; *α*: Internal consistency coefficient; *ω*: McDonald's omega; CR: Composite Reliability.

Table 4
Measurement invariance results across gender.

Stage	χ^2	df	p	RMSEA	RMSEA [%90 CI]	SRMR	Robust-TLI	Robust-CFI	ΔCFI
Configural	1293.827	586	<.001	.070	.065- .075	.056	.913	.921	-
Metric	1318.978	608	<.001	.069	.064-.074	.059	.916	.921	.000
Scalar	1346.168	630	<.001	.068	.063-.073	.059	.919	.921	.000
Strict	1356.895	656	<.001	.066	.061-.071	.06	.922	.922	.001

Table 5
Reliability and convergent validity results for the EFA and CFA samples.

Subscale	EFA Sample (n = 300)				CFA Sample (n = 305)			
	α	ω	AVE	CR	α	ω	AVE	CR
Basic Prompt Engineering	.883	.889	.498	.871	.865	.871	.490	.870
Pedagogical Prompt Engineering	.901	.901	.598	.884	.880	.886	.605	.882
Ethical Prompting Practice	.949	.949	.666	.941	.938	.939	.659	.939
Continuous Professional Development	.912	.913	.586	.893	.887	.886	.566	.888
Total	.963	.964	.586	.975	.955	.956	.583	.973

Note. α = Cronbach's alpha; ω = McDonald's omega; AVE = Average Variance Extracted; CR = Composite Reliability.

4.1. Interpreting the hierarchy of Ed-PESS

Basic Prompt Engineering. The first dimension, Basic Prompt Engineering, accounted for the largest share of variance, indicating that operational prompt structuring forms a central component of the Ed-PESS. Although convergent validity was borderline (AVE = .490), reliability remained strong (α = .865; ω ≈ .87), suggesting a broader operational self-efficacy domain rather than a narrowly homogeneous one. Its moderate association with Pedagogical Prompt Engineering (r = .591) further indicates that confidence in structuring prompts supports, but does not collapse into, confidence in instructional integration.

Theoretically, this operational dimension can be read Cognitive Load Theory: when prompt syntax itself consumes working-

memory resources, it may generate extraneous load that competes with pedagogical reasoning and output evaluation (Sweller, 1988). Accordingly, self-efficacy in Basic Prompt Engineering reflects confidence in managing core prompt components with minimal unnecessary effort so that attention can shift to instructional planning. This interpretation is supported by recent work on reducing extraneous load in LLM-supported learning and by component-based prompt frameworks such as Input, Instruction, Output, and Context, while also suggesting that operational fluency is a foundation for, rather than a guarantee of, later pedagogical use (Jin et al., 2025; Mendes et al., 2025).

Pedagogical Prompt Engineering. In the Ed-PESS, Basic and Pedagogical Prompt Engineering were moderately correlated ($r = .591$), while overall inter-factor correlations ($r = .507$-$.645$) supported a coherent but multidimensional structure. Reliability was strong and convergent validity was adequate for Pedagogical Prompt Engineering (AVE = .605). Together, these findings support interpreting pedagogical prompting as a distinct self-efficacy facet centered on instructional integration rather than mere operational prompt construction.

Theoretically, this distinction aligns with Shulman's (1986) view of teaching as professional judgment, in which teachers must connect procedures, content, and pedagogical reasoning in response to situational demands. It also fits the TPACK argument that knowing how to use technology is not equivalent to knowing how to teach with it (Mishra & Koehler, 2006), a point further reinforced in AI-focused extensions that emphasize the centrality of pedagogical knowledge for the ethical and effective use of AI tools (Celik, 2023). From an instrumental genesis perspective, GenAI becomes an instructional instrument only when coupled with teachers' situation-specific utilization schemes; in this sense, higher self-efficacy in Pedagogical Prompt Engineering reflects confidence in turning AI outputs into pedagogically useable resources rather than merely generating them (Rabardel, 2002). Taken together, these findings suggest that teacher education should move beyond "chatting with AI" toward structured, case-based design experiences that cultivate instructional alignment and reflective judgment.

Ethical Prompting Practice. In the Ed-PESS, Ethical Prompting Practice formed a highly coherent subdomain of prompt engineering self-efficacy. Internal consistency was excellent ($\alpha = .938$; $\omega \approx .939$) and convergent validity was strong (AVE = .659), while inter-factor correlations ($r = .507$-$.645$) indicated that this dimension was meaningfully connected to the others yet remained empirically distinguishable. Interpreted through Bandura's (2001) account of moral agency, this dimension reflects pre-service teachers' confidence in maintaining ethical responsibility when using GenAI, particularly in situations where responsibility might otherwise be displaced onto the tool. In the scale, this is reflected in items concerning privacy and data protection, bias and age-appropriateness guardrails, reliability-oriented constraints, and the modeling of ethical prompting norms for students. This interpretation is also consistent with UNESCO's emphasis on human control and accountability in educational AI use (UNESCO, 2024) and with studies showing that ethical prompting involves noticing bias, revising prompts, and critically checking or rewriting outputs to ensure that they are safe, fair, and instructionally appropriate (Agirdag, 2025; Selwyn et al., 2025).

Continuous Professional Development (CPD). In the Ed-PESS, CPD emerged as a separate but connected dimension of prompt engineering self-efficacy. Internal consistency was strong ($\alpha = .887$; $\omega \approx .884$-$.886$) and convergent validity was acceptable (AVE = .566), while inter-factor correlations ($r = .507$-$.645$) indicated that this dimension was related to, yet distinct from, the others. These findings suggest that CPD reflects not merely a positive attitude, but self-efficacy for keeping prompt-related practices current and improving them as GenAI tools and classroom expectations evolve. This interpretation is consistent with work on self-efficacy and self-regulation in technology-mediated learning, which emphasizes ongoing cycles of planning, monitoring, evaluating, and adaptation (Artino, 2012; Zimmerman, 2002). It is also supported by recent studies portraying prompting as an experimental and continually changing practice that requires repeated updating, reflection, and refinement in classroom use (ElSayary et al., 2025; Tümen Akyıldız, 2026).

4.2. Theoretical implications: prompting as "teacher agency"

Taken together, the Ed-PESS supports viewing prompt engineering self-efficacy as a form of teacher agency in GenAI-supported instruction, in which teachers actively regulate how AI is used and for what pedagogical purposes. From an instrument-mediated activity perspective, GenAI is initially an artifact and becomes an instructional instrument only when linked to teachers' utilization schemes; in this sense, prompt engineering self-efficacy reflects confidence in building, adapting, and coordinating these schemes to support learning goals (Rabardel, 2002). This view is further clarified by moral agency: ethical prompting in educational use depends not on abstract reasoning alone, but on teachers' confidence that they can activate self-regulatory safeguards and maintain human accountability when responsibility might otherwise be displaced onto the tool (Bandura, 2006). Most importantly, this agency operates through an iterative cycle of evaluating whether outputs fit instructional goals and ethical expectations, identifying weaknesses, and revising prompts accordingly.

4.3. Practical implications for teacher education

General lectures that introduce AI are unlikely to build prompt engineering self-efficacy as teacher agency on their own (Bandura, 2001, 2006; Zimmerman, 2002). Teacher education curricula should therefore include Prompt Labs, where pre-service teachers practise writing prompts for clear instructional goals, embedding ethical and safety constraints, and iteratively refining prompts and outputs within realistic teaching tasks. Such experiences can provide mastery opportunities that strengthen self-efficacy (Bandura, 2006) while following guided cycles of planning, doing, and reflecting that help teacher candidates understand why an output failed and how the prompt should be revised (Zimmerman, 2002).

To support evidence-informed curriculum design, the Ed-PESS can be used as a pre-test and post-test in educational technology or

methods courses to identify teacher candidates' self-efficacy profiles and to examine whether specific instructional activities strengthen targeted dimensions of prompt engineering self-efficacy. For practical use, teacher educators may interpret Ed-PESS subscale profiles formatively to identify which dimension requires greater support, while researchers may use the scale to track dimension-specific change across courses or interventions, with scores treated as indicators of perceived capability rather than direct evidence of actual prompting performance. Teacher education should also include explicit modules on ethical use that train candidates to write prompts with safety and fairness constraints, such as requesting gender-neutral roles, child-appropriate content, and fair examples, while also teaching routines for checking and revising outputs. In this way, ethical prompting practice remains a teachable and assessable part of professional preparation rather than something assumed to emerge automatically from tool use.

4.4. Limitations and future directions

Although this study provides strong evidence for the internal structure and reliability of the Ed-PESS, several limitations should guide interpretation and future research. First, the Ed-PESS is a self-report instrument, and thus reflects pre-service teachers' perceived capability to design, adapt, and safeguard GenAI prompts rather than their actual performance in authentic prompting tasks. Future studies should therefore combine Ed-PESS scores with performance-based assessments, such as rubric-scored tasks requiring candidates to write prompts, evaluate outputs, and revise them under pedagogical and ethical constraints, to examine criterion-related validity evidence.

Second, the scale was validated within teacher education programs in a single national context, which limits generalizability. Although gender-based measurement invariance was supported in the present study, future research should test invariance across other relevant groups, such as year level, prior GenAI training, AI tool familiarity, and cross-cultural samples, before broader group comparisons are made. In addition, the AVE value for Basic Prompt Engineering remained slightly below the conventional threshold (AVE = .490), which may indicate a somewhat broader operational domain despite otherwise acceptable reliability; this should therefore be acknowledged as a minor measurement limitation.

Third, prompting is embedded in a rapidly changing technological environment. As models, interfaces, and interaction modes evolve, the conditions under which prompting is learned and evaluated may also shift. For this reason, the Ed-PESS should be interpreted as measuring generalizable self-efficacy for pedagogically purposeful, ethically responsible, and developmentally appropriate prompting rather than model-specific tactics. Regular monitoring and planned item review will be important for maintaining content validity as new classroom demands and risks emerge.

5. Conclusion

The main purpose of this study was to develop and validate the Educational Prompt Engineering Self-Efficacy Scale for Pre-service Teachers (Ed-PESS). This responds to the growing need for reliable measurement in AI integrated teacher education. Psychometric results supported a clear four factor structure. The EFA explained 64.60 percent of the total variance. The CFA, with a robust correction, showed good model fit (CFI = .929, RMSEA = .064, SRMR = .057). Reliability evidence was also strong across the scale. These results suggest that the Ed-PESS can support teacher education in a practical way. It can help prepare future teachers to stay human in the loop and guide GenAI use toward pedagogically sound and ethically responsible goals. This study also makes a theoretical and methodological contribution. To our knowledge, it is the first validated instrument designed to assess prompt engineering self-efficacy beliefs among pre-service teachers. In practice, the Ed-PESS addresses a gap by enabling systematic assessment and targeted development of this emerging self-efficacy construct in teacher education curricula.

Ethical approval

This research received ethical approval from the Scientific Research and Publication Ethics Committee at Nevsehir Hacı Bektas Veli University in Türkiye, as per decision number 2025.11.389.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this research, the authors used *Paperpal* to paraphrase their writing for more academic enhancement, and *Claude Opus* (AI tools) for language translation and text reduction. After using these AI tools, the author(s) reviewed and edited the content as necessary and take full responsibility for the content of the publication. Alongside using AI tools for language tasks, the author(s) thoroughly reviewed and accurately cited all references in this research. They verified each reference's authenticity, including DOI links. It is crucial to note that all data and findings come from properly cited sources, not AI-generated. The author(s) fully ensure the research's integrity and accuracy.

Funding

The author reports no funding.

CRediT authorship contribution statement

Fatih Karatas: Conceptualization, Investigation, Writing – review & editing. **Recep Gür:** Data curation, Methodology, Validation, Writing – original draft. **Barış Erïçok:** Methodology, Supervision, Writing – original draft. **Fatma Başarır:** Writing – original draft, Writing – review & editing. **Lokman Tanrikulu:** Data curation, Resources, Writing – review & editing.

Competing interests

The authors declare that they have no competing interests related to this research.

Appendix 1. Items pool and supporting literature evidence

Scale Item (Variable)	Direct Quote (Evidence)
Dimension 1: Basic Prompt Engineering	
1. I can explain how AI tools generate content.	1. "AI systems generate text (outputs) by making statistically informed predictions based on the patterns they have learned and responding to prompts (inputs) entered by users (Park & Choo, 2024)." 2. "The session included demonstrations of interactions with ChatGPT-4, showcasing its text generation and text-to-image functionalities. Particular emphasis was placed on the importance of crafting effective prompts, as well as demonstrating how refining prompts could elicit more accurate and contextually appropriate responses from the AI (Biberman-Shalev, 2025)." 3. "ChatGPT works by learning from a large amount of text data, which helps it understand grammar, vocabulary, and the meaning of words in different contexts. This learning process of ChatGPT is analogous to training ChatGPT's brain to understand language (Lee et al., 2024)."
2. I can select the AI tool (e.g., ChatGPT, Claude, Gemini) that best fits my needs.	1. "When prompt engineering, you will start by choosing a model. Prompts might need to be optimized for your specific model, regardless of whether you use Gemini language models in Vertex AI, GPT, Claude, or an open source model like Gemma or LLaMA (Boonstra, 2025)." 2. "I preferred ChaGPT. It is expensive, but it gives you value for money. I used my bursary and subscribed to this tool (Van Wyk, 2025)."
3. I can create simple, clear, and understandable prompts when using AI tools.	1. "The first component of the CLEAR Framework emphasizes the importance of conciseness in crafting prompts. A concise prompt removes superfluous information, allowing AI language models to focus on the most important aspects of the task, resulting in more pertinent and precise responses. Clarity is also crucial, as unclear or imprecise instructions may result in AI-generated content that does not meet the user's needs or expectations (Lo, 2023)." 2. "'Concise' prompts use clear, brief language (Park & Choo, 2024)." 3. "Lo (2023) suggested four guiding principles for prompt engineering: (a) clarity and precision, (b) contextual information, (c) desired format, and (d) verbosity control. These principles emphasize that prompts must be clear and precise to produce an accurate, relevant response (Park & Choo, 2024)."
4. I can clearly express necessary components in my prompts (instruction, context, output format).	1. " Federiakin et al. (2024) list four elements of a prompt. These elements, combined together, substitute a prompt, which formulates the problem, gives the model the necessary information to solve it, and contains output in the desired form: Instruction – a specific task that guides the model's behavior; Context – external information or additional context that provides background knowledge to the model; Input data – the content of the prompt that the model needs to solve; Output indicator – specifies the type or format of the desired output (Federiakin et al., 2024)." 2. "The most advanced practices treat the prompt as a quasi-formal specification of the desired behavior (Serra & Oliveira, 2025), based on four key principles: Role Delegation; Rich Contextualisation and Task Delimitation; Explicit and Structured Instructions; Definition of Constraints and Output Format (Serra & Oliveira, 2025)." 3. "Park and Choo (2024) suggest incorporating the following five components into prompts: Persona, Aim, Recipients, Theme, and Structure (PARTS) (Park & Choo, 2024)."
5. I can assign roles to the AI (e.g., "act as a math teacher") to get responses that match my goals.	1. "First, 'Persona' identifies the role. Including the persona in the prompt sets the context for the user's request. For example, educators can identify their roles … or assigning the AI's role (e.g., 'Act as if you are Martin Luther King Jr.,' 'As a special education teacher in fifth grade reading class' and 'You are __.') (Park & Choo, 2024)." 2. "Role Delegation: assigning the model an explicit persona (e.g., 'Act as an expert tutor … ') to guide its behavior, tone, and knowledge (Serra & Oliveira, 2025)." 3. "Role prompting is a technique in prompt engineering that involves assigning a specific role to the gen AI model. This can help the model to generate more relevant and informative output, as the model can craft its responses to the specific role that it has been assigned (Boonstra, 2025)."

(*continued on next page*)

(*continued*)

Scale Item (Variable)	Direct Quote (Evidence)
	4. "Role-Based/Persona Prompting. It helps AI adopt a pedagogically or professionally relevant voice. Prompts such as 'Act as a mentor helping with a project' can simulate teacher-student interaction, fostering reflection and personalization (Qian, 2025)."
6. I ensure the AI guides me in the best way by choosing the appropriate path for my needs.	1. "Goal-oriented and task-specific frameworks focus on designing prompts to fulfill clearly defined instructional or functional objectives. These frameworks align prompt structures with desired outcomes, such as eliciting student reasoning or performing a diagnostic task (Qian, 2025)." 2. "This alignment suggests that educators should not only consider what content they want AI to generate but also the nature of the learning task—whether it's cultivating meta-cognition, facilitating argumentation, or assessing understanding—and select or adapt prompting strategies accordingly (Qian, 2025)."
7. I can analyze and improve prompts that do not produce the expected results.	1. "Crafting prompts often requires iterative refinement (Park & Choo, 2024). Initial prompts are evaluated and verified based on the quality of AI responses. If the output is unsatisfactory, the prompt needs to be revised and retested. This process continues until the desired response is achieved, in a manner similar to trial and error (Park & Choo, 2024)." 2. "Iterative Refinement. It emphasizes prompt development as a cyclical process of experimentation and adjustment, mirroring the way humans naturally learn through trial, reflection, and revision (Qian, 2025)." 3. " Federiakin et al. (2024) define prompt literacy as the ability to generate precise prompts as input for AI tools, interpret the outputs, and iteratively refine prompts to achieve desired results (Federiakin et al., 2024)." 4. "The data indicates several specific strategies used in this refinement process … Nearly all participants modified their initial prompts to make them better (Moorhouse et al., 2025)."
8. I can reuse the same prompt for different purposes by leaving input slots blank (e.g., topic, tone, target audience).	1. "To reuse prompts and make it more dynamic use variables in the prompt, which can be changed for different inputs … Variables can save you time and effort by allowing you to avoid repeating yourself. If you need to use the same piece of information in multiple prompts, you can store it in a variable and then reference that variable in each prompt (Boonstra, 2025)." 2. "Moreover, the prompt can be used as a template for solving similar problems or generating new prompts for similar problems (template pattern) (Velásquez-Henao et al., 2023)." 3. "prompt patterns are essential to effective prompt engineering' (p. 1) and they offer 'reusable solutions to specific problems' (p. 2) (Lee & Palmer, 2025)."
9. I can systematically save prompts that yield effective results.	1. "We recommend creating a Google Sheet as a template. The advantages of this approach are that you have a complete record when you inevitably have to revisit your prompting work … If you're lucky enough to be using Vertex AI Studio, save your prompts (using the same name and version as listed in your documentation) and track the hyperlink to the saved prompt in the table (Boonstra, 2025)." 2. "A very important recommendation is to use the designed prompt in the chat and then collect and save the system output. At this point, it is imperative to preserve the history of the process design to realize ex-post evaluations of the performance of the designed prompts (Velásquez-Henao et al., 2023)." 3. "Customized prompts can be designed and saved as prompt templates in the chat, allowing for future use (Velásquez-Henao et al., 2023)."
Dimension 2: Pedagogical Prompt Engineering	
10. I can design prompts that generate content aligned with my learning objectives.	1. "Goal-oriented and task-specific frameworks focus on designing prompts to fulfill clearly defined instructional or functional objectives. These frameworks align prompt structures with desired outcomes, such as eliciting student reasoning or performing a diagnostic task (Qian, 2025)." 2. "For teachers to use GenAI tools effectively, they must craft prompts that encourage the AI to engage substantively with instructional goals and deliver responses aligned with intended learning outcomes (Celik et al., 2026)." 3. "Continuous refinement through iterative prompt writing ensures alignment with learning objectives and facilitates real-time lesson adaptation (Carl et al., 2024)." 4. "Each instruction must be 'initiated with a clear and precise directive' and maintain alignment with the learning objectives and educational goals (Lee & Palmer, 2025)."
11. I can develop prompts that generate differentiated content suitable for students' needs.	1. "Adaptive learning is an instructional approach in which learning pathways are adjusted to align with learners' needs, prior knowledge, and interests … instruction is tailored to students' progress and characteristics (Celik et al., 2026)." 2. "Crucially, the task emphasized designing adaptive or personalized instructional strategies by encouraging pre-service teachers to request support from the AI that considered pupil diversity, engagement, and differentiated learning needs (Celik et al., 2026)." 3. "Real-time AI assistance in classrooms enhances lesson delivery and enables personalized learning for diverse student needs (Carl et al., 2024)." 4. "Prompts emphasizing contextual clarity aid differentiation by adjusting outputs to learner needs (Qian, 2025)."

(*continued on next page*)

(*continued*)

Scale Item (Variable)	Direct Quote (Evidence)
12. I can develop prompts that generate content specific to my teaching field.	1. "The argument is that teachers must understand their content area to construct effective prompts (e.g., pedagogical content knowledge [PCK]), they need to be able to critically evaluate the responses generated in relation to their subject, task requirements (Moorhouse et al., 2025)." 2. "The study's approach to developing a financial literacy game serves as a blueprint for creating engaging, subject-specific educational games across disciplines (Carl et al., 2024)." 3. "Disciplinary literacy for pre-service teachers includes not only an understanding of subject content and the application of teaching strategies, but also mastery of educational technology (Kang et al., 2025)." 4. "The ability to generate subject-specific content aligns with […] the extensive applicability of AI in creating educational materials (ElSayary et al., 2025)."
13. I can use AI prompts to provide feedback to my students in a short time.	1. "LLMs have emerged as powerful tools due to their ability to provide adaptive, scalable, and timely feedback, addressing long-standing challenges such as time constraints and limited resources (Jacobsen & Weber, 2025)." 2. "Participants also emphasized the importance of immediate feedback and personalized learning experiences facilitated by AI. One participant articulated this well, stating, 'I think it is beneficial for providing immediate feedback and personalized learning experiences (Mzwri & Turcsányi-Szabó, 2025). '" 3. "In this phase, EnSmart evaluates student responses, providing immediate grades and personalized, human-like feedback (Mzwri & Turcsányi-Szabó, 2025)."
14. I can examine the suitability of AI outputs for my purpose with a critical perspective.	1. "for teachers to be prompt literate, they need to have three essential components: content knowledge, critical thinking, and the ability to engage in an iterative design process with LLMs (Moorhouse et al., 2025)." 2. "I can critically evaluate AI-generated responses for reliability, factual accuracy, and bias (Gibreel & Arpaci, 2025)." 3. "Training in high-order prompt engineering skills, including critical evaluation, is necessary to prepare students in higher education for industries engaging with AI technologies (Lee & Palmer, 2025)." 4. "Programs may include simple evaluation rubrics for AI-generated materials so that teacher candidates learn to judge accuracy, appropriateness, and creative value rather than accepting suggestions automatically (Tümen Akyıldız, 2026)."
15. I can design prompts that generate content reflecting different perspectives.	1. "In social studies, they may request multiple historical viewpoints or primary-source excerpts and critically inspect ideological bias before classroom use (Davila-Moran et al., 2025)." 2. "By directing the AI model's responses, prompt engineering enables learners to explore diverse perspectives and generate creative content (Kabeer et al., 2025)." 3. "The discussions should encourage students to debate different viewpoints, fostering an environment of critical analysis and ethical reasoning (Walter, 2024)."
16. I can develop creative prompt ideas that will increase student engagement.	1. "Prompt engineering, the process of crafting adequate inputs for AI models, has emerged as a powerful method to nurture engagement and originality (Kabeer et al., 2025)." 2. "Developing skills in prompt engineering improves students' abilities to transform AI from a 'mere repository of information into an interactive tool that stimulates deeper learning and understanding (Walter, 2024)."' 3. "Prompts can encourage students to explore genres, experiment with narrative styles, or visualize abstract concepts through AI-generated imagery. This process enhances their creative writing skills and fosters a deeper understanding (Kabeer et al., 2025)." 4. "Unstructured prompting fosters creativity, exploration, and divergent thinking (Qian, 2025)."
Dimension 3: Ethical Prompting Practice	
17. I can develop classroom rules for prompt creation (e.g., 'do not use your name') to ensure privacy and data security.	1. "Educate students about data privacy, including how their data is used by AI systems and ways to protect their digital footprint (Walter, 2024)." 2. "One practical approach is to develop a personal checklist for appropriate and ethical uses of AI (Park & Choo, 2024)." 3. "It should be clear how the expectations of the school look like so that students know exactly what they are allowed and what they are not allowed to do (Walter, 2024)."
18. I can protect the confidentiality of student data while creating prompts.	1. "Providers of LLM-based GenAI tools, such as OpenAI, use prompting data to train their AI models, which may raise concerns regarding data privacy and protection … To mitigate this, learners should avoid providing personal or identifiable information when interacting with LLM-based GenAI (Hsu, 2025)." 2. "AI systems frequently handle private user data without following open procedures for its storage or use … many AI applications gather personal data without express consent (Tümen Akyıldız, 2026)." 3. "Ethical AI development is essential, focusing on transparency, unbiased content, and privacy-respecting practices (Walter, 2024)."
19. I can use protective expressions (e.g., 'do not save this conversation') to ensure student safety when creating prompts.	1. "To mitigate this, learners should avoid providing personal or identifiable information when interacting with LLM-based GenAI. If avoiding the use of personal data is less feasible, opting for localised LLM-based GenAI tools … could be considered (Hsu, 2025)." 2. "Google for Education has also released a responsibility checklist. This checklist provides a list of guidelines which includes review of AI outputs, disclosure of the use of AI,

(*continued on next page*)

(*continued*)

Scale Item (Variable)	Direct Quote (Evidence)
	consideration of privacy and security implications, and thoughtful use of AI with consideration of personal judgment (Park & Choo, 2024)."
20. I can add expressions to my prompts (e.g., 'use examples including different cultures') to prevent biases related to gender, culture, age, etc.	1. "The fundamental technology behind LLMs relies on the strategy of generalization … existing biases embedded in the training data—whether related to race, class, gender, or culture—are algorithmically reinforced and normalized (Agirdag, 2025)." 2. "The potential for bias in AI algorithms was a recurring theme, with one participant warning that AI can inherit biases from its training data, potentially leading to inequitable outcomes in content generation and assessment (Mzwri & Turcsányi-Szabó, 2025)." 3. "Prompt engineering serves as a discipline that guides model outputs toward fairness, accuracy, and contextual relevance. […] furthermore its role in mitigating biases in transformer-based models by ensuring that AI-generated content remains neutral, inclusive, and ethically aligned (Mzwri & Turcsányi-Szabó, 2025)." 4. "Iteratively refining prompts to improve relevance and bias mitigation (Davila-Moran et al., 2025)."
21. I can add ethical instructions like 'please be neutral' or 'produce content suitable for children' to my prompts.	1. "The < target_age_group > component thus functions as a mechanism for communicative adaptation, aligning the discourse with the expectations and needs of the target audience (Serra & Oliveira, 2025)." 2. "To prevent the model from generating harmful or biased content or when a strict output format or style is needed … constraints are still valuable (Boonstra, 2025)." 3. "Current LLMs aim to produce human-like, helpful and safe content while avoiding harmful outputs (Woo et al., 2025)."
22. I can ensure more accurate information by adding expressions like 'answer based on reliable sources' to my prompts.	1. "Is the answer as accurate as expected?.. Does the answer have elements that may be factually incorrect (hallucinations)?.. requesting additional evidence, such as querying top-k information sources and having the language model rate the credibility of each source, is another strategy (Velásquez-Henao et al., 2023)." 2. "It is crucial for educators to be mindful of the imperfections inherent in emerging technologies … educators should evaluate, assess, and verify AI responses by being vigilant for hallucinations and other AI errors (Park & Choo, 2024)."
23. I can demonstrate ethical prompt writing skills to students using my own guidelines as examples.	1. "These sessions should focus on up-to-date AI developments, ethical considerations, and best practices. By regularly engaging with AI topics, the academic community can stay informed and proficient in managing AI tools and concepts (Walter, 2024)." 2. "Central to all these efforts is the fostering of a critical and ethical approach to AI. Ethical discussions should be an integral part of the learning process, encouraging students to contemplate AI's societal impact (Walter, 2024)." 3. "Educators played a crucial role in fostering ethical AI literacy by teaching students how to use AI responsibly, ensuring that AI is a tool for inspiration rather than a substitute for independent thought (Kabeer et al., 2025)."
24. I can explain to students what to pay attention to when writing prompts for safe interaction (e.g., 'do not show harmful content').	1. "The deliberate goal is to eventually lead students towards a responsible use of AI, but to do so, they need to understand how one can 'talk' to an AI so that it does what it is supposed to (Walter, 2024)." 2. "Framing prompting as a pedagogical responsibility signals the need to integrate AI training into educational design, enabling students to critically engage with AI in their learning (Andewi et al., 2025)." 3. "Students certainly need critical thinking skills to interact with GAI, both for prompting (input) and for evaluating the quality, accuracy, and relevance of its outputs (Oliveira et al., 2025)."
25. I can detect and correct parts of a prompt that require ethical fixing (e.g., ambiguous expressions).	1. "Disambiguation can be avoided by providing a detailed description or scope of the problem or the data (Velásquez-Henao et al., 2023)." 2. "Being aware of this bias, I prompted ChatGPT to reflect upon its translation and to explore what biases could have caused such a translation (Agirdag, 2025)."
Dimension 4: Continuous Professional Development	
26. I can follow current developments in AI tools and prompt engineering.	1. "Regular training and workshops for educators will ensure they stay updated with the latest AI technology advancements." (Walter, 2024) 2. "Provide workshops for career guidance that emphasize adaptability and the importance of continuous learning in an AI-evolving job landscape. Teach an agile mindset and provide sources to learn the newest developments." (Walter, 2024) 3. "The rapid evolution of AI technologies can render existing frameworks quickly outdated. As new AI capabilities evolve, educators may find themselves needing to continuously adapt their teaching strategies, frameworks and assessment processes, or risk passively developing potential gaps in knowledge and skills." (Lee & Palmer, 2025)
27. I strive to use current developments in AI prompt engineering in my teaching practices.	1. "Several teachers indicated a desire to investigate AI tools with specialized functionalities, highlighting the necessity of continuously adapting AI technologies to address the changing demands of education." (ElSayary et al., 2025) 2. "Integrating prompt engineering into teacher education curricula may therefore serve as a bridge between digital literacy and pedagogical creativity, preparing future teachers to use AI tools reflectively and effectively." (Tümen Akyıldız, 2026) 3. "The deliberate goal is to eventually lead students towards a responsible use of AI, but to do so, they need to understand how one can 'talk' to an AI so that it does what it is supposed to." (Walter, 2024)

(*continued on next page*)

(*continued*)

Scale Item (Variable)	Direct Quote (Evidence)
28. I can adapt my teaching practices according to current developments in AI and prompt engineering.	1. "Adapt teaching methods to cater to diverse learning styles influenced by AI and technology, including interactive and multimodal learning approaches. AI assistants and platforms can help teachers quickly adapt to new formats." (Walter, 2024) 2. "This dynamic landscape necessitates ongoing research as well as continuous professional and curriculum development driven by a commitment to staying current with technological advancements." (Lee & Palmer, 2025) 3. "AI output, even when useful, sometimes required refinement and retuning in an attempt to perfectly tailor it to specific individual classroom requirements. This underlines not only utilizing AI tools but also developing competencies for effective communication with such tools through effective prompt engineering." (ElSayary et al., 2025)
29. I can improve my prompts based on student feedback.	1. "The results confirm that effective engineering of prompts helps teachers make a more direct link between lesson content and educational objectives … Iterative refinement of prompts is an important part of getting high-quality, relevant output." (ElSayary et al., 2025) 2. "Teachers iteratively developed constraints, limitations, and refined their goals through the feedback loops provided by ChatGPT." (Carl et al., 2024) 3. "AI-driven feedback promotes SRL by identifying knowledge gaps and aligning insights with students' learning progress." (Mzwri & Turcsányi-Szabó, 2025) 4. "The pre-service teachers reflect on their prompts based on feedback from AI responses and refine their prompts as needed, with additional support from teacher educators or more knowledgeable peers when necessary." (Hsu, 2025)
30. I can determine my deficiencies regarding prompt engineering and create a professional development plan.	1. "Reflective practice is a critical component of teacher professional development. Schön's concepts of reflection-in-action and reflection-on-action provide a robust foundation for understanding teacher-reflective practices … This type of reflection facilitates a more thorough evaluation of teaching practices and outcomes, promoting continuous professional development." (ElSayary et al., 2025) 2. "Prompt engineering is becoming more and more important, yet it is still not taught enough in schools … This difference shows a big gap in talent and preparation that education needs to fix right away." (Tümen Akyıldız, 2026) 3. "Programs may include simple evaluation rubrics for AI-generated materials so that teacher candidates learn to judge accuracy, appropriateness, and creative value rather than accepting suggestions automatically." (Tümen Akyıldız, 2026)
31. I can participate in trainings to improve my prompt engineering skills.	1. "Continuous and comprehensive training for teachers is crucial to ensure that they can effectively integrate AI tools into their pedagogical practices." (Walter, 2024) 2. "The study promotes additional training programs for teachers in AI literacy, particularly in terms of prompt engineering. Teachers with proper training in developing effective prompts experienced easier incorporation of AI tools in lesson planning." (ElSayary et al., 2025) 3. "The main insight is that a single, targeted workshop can significantly improve undergraduates' AI self-efficacy, conceptual understanding and practical ability to write effective prompts, demonstrating the high value of integrating prompt engineering into discipline-specific courses." (Woo et al., 2025) 4. "Embedding early prompt-engineering clinics may accelerate the adoption of generative AI in teacher-education practice and reduce technology anxiety." (Davila-Moran et al., 2025)
32. I make an effort to continuously improve my teaching practices by benefiting from prompt engineering.	1. "The evolution of their prompting practice became a visible process in which creativity and pedagogical skill developed side by side through experience." (Tümen Akyıldız, 2026) 2. "Prompt engineering holds potential not only as a technical skill but also as a core element in teacher education programs that aim to promote 21st-century thinking skills and creativity in future educators." (Tümen Akyıldız, 2026) 3. "Examining such emerging strategies would offer valuable insights into how prompt engineering becomes a part of teachers' professional growth and everyday decision-making." (Tümen Akyıldız, 2026) 4. "This study provides insights for Information Systems educators on innovative professional growth, course design, and Scholarship for Teaching and Learning (SoTL). In addition, educators are afforded opportunities to leverage AI effectively in teaching practices, enhancing content creation and delivery." (Carl et al., 2024)

Note. Ed-PESS Administration, Scoring, and Interpretation Guidelines. Items are rated on a 5-point Likert scale (1 = "I cannot do this at all," 5 = "I can definitely do this"). Subscale scores are computed as the arithmetic mean of the items within each subscale, and the total score as the mean across all 26 items, yielding possible scores from 1.00 to 5.00. For formative interpretation, subscale means below 3.00 may indicate low self-efficacy requiring targeted support, means between 3.00 and 3.99 as developing, and means of 4.00 or above as consolidated; these bands are sample-based reference values and should be cross-validated before being treated as norms. Teacher educators may use Ed-PESS subscale profiles to identify which dimension (Basic, Pedagogical, Ethical, or CPD) requires the strongest curricular emphasis.

Appendix 2. The Results of the Comparisons Between the Upper 27% and Lower 27% Participant Groups

item	Group	M	SD	t
item_1	the lower 27%	2.25	.830	−13.655*
	the upper 27%	3.37	1.134	
item_2	the lower 27%	3.40	.944	−13.657*
	the upper 27%	4.58	.687	
item_3	the lower 27%	3.20	.872	−13.986*
	the upper 27%	4.70	.558	
item_4	the lower 27%	3.12	.797	−11.185*
	the upper 27%	4.67	.632	
item_5	the lower 27%	3.17	.933	−10.716*
	the upper 27%	4.77	.481	
item_6	the lower 27%	3.21	.890	−8.644*
	the upper 27%	4.78	.474	
item_7	the lower 27%	2.95	.835	−13.933*
	the upper 27%	4.43	.851	
item_8	the lower 27%	2.94	.747	−13.514*
	the upper 27%	4.31	.875	
item_9	the lower 27%	2.94	.747	−17.836*
	the upper 27%	4.20	1.077	
item_10	the lower 27%	2.78	.775	−15.374*
	the upper 27%	4.46	.759	
item_11	the lower 27%	2.70	.813	−15.652*
	the upper 27%	4.38	.768	
item_12	the lower 27%	2.67	.742	−15.161*
	the upper 27%	4.59	.628	
item_13	the lower 27%	2.91	.616	−15.639*
	the upper 27%	4.52	.709	
item_14	the lower 27%	3.02	.724	−17.427*
	the upper 27%	4.62	.561	
item_15	the lower 27%	2.72	.693	−16.138*
	the upper 27%	4.47	.776	
item_16	the lower 27%	2.93	.703	−17.705*
	the upper 27%	4.57	.631	
item_17	the lower 27%	2.94	.812	−19.891*
	the upper 27%	4.77	.481	
item_18	the lower 27%	3.07	.771	−14.524*
	the upper 27%	4.74	.519	
item_19	the lower 27%	3.15	.743	−13.192*
	the upper 27%	4.84	.432	
item_20	the lower 27%	2.91	.762	−16.849*
	the upper 27%	4.83	.412	
item_21	the lower 27%	3.01	.873	−16.534*
	the upper 27%	4.70	.580	
item_22	the lower 27%	2.96	.715	−16.888*
	the upper 27%	4.54	.807	
item_23	the lower 27%	2.88	.797	−12.125*
	the upper 27%	4.72	.575	
item_24	the lower 27%	2.90	.700	−15.156*
	the upper 27%	4.64	.639	
item_25	the lower 27%	2.93	.648	−13.333*
	the upper 27%	4.59	.608	
item_26	the lower 27%	2.52	.776	−16.380*
	the upper 27%	4.19	.963	
item_27	the lower 27%	2.58	.820	−11.688*
	the upper 27%	4.42	.722	
item_28	the lower 27%	2.63	.828	−9.765*
	the upper 27%	4.38	.845	
item_29	the lower 27%	2.90	.682	−11.888*
	the upper 27%	4.57	.611	
item_30	the lower 27%	2.70	.843	−13.655*
	the upper 27%	4.31	.903	
item_31	the lower 27%	2.53	.838	−13.657*
	the upper 27%	4.01	1.078	
item_32	the lower 27%	2.68	.722	−13.986*
	the upper 27%	4.25	.942	

*: $p < .05$.

Data availability

Research Link Provided
Ed-Pess _Data Source, https://drive.google.com/drive/folders/1j7iiqbvehmu10ru1rrekxr55szfvgwtx?usp=drive_link

References

Agirdag, O. (2025). Beyond prompt engineering: Prompting (l)iteracy, linguistic capital, and educational inequality. *Educational Theory*. https://doi.org/10.1111/edth.70057

Andewi, W., Waziana, W., Wibisono, D., Putra, K. A., Hastomo, T., & Oktarin, I. B. (2025). From prompting to proficiency: A mixed-methods analysis of prompting with ChatGPT versus lecturer interaction in an EFL classroom. *Journal of Studies in the English Language, 20*(2), 210–238. https://doi.org/10.64731/jsel.v20i2.282318

Arocha, J. G. (2025). Critical phenomenology of prompting in artificial intelligence. *Sophia-Colección de Filosofía de la Educación*, (39), 141–165. https://doi.org/10.17163/soph.n39.2025.04

Artino, A. R., Jr. (2012). Academic self-efficacy: From educational theory to instructional practice. *Perspectives on Medical Education, 1*(2), 76–85. https://doi.org/10.1007/s40037-012-0012-5

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Information Age Publishing.

Biberman-Shalev, L. (2025). Prompting theory into practice: Utilizing ChatGPT-4 in a curriculum planning course. *Education Sciences, 15*(2), Article 196. https://doi.org/10.3390/educsci15020196

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*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Boonstra, L. (2025). *Prompt Engineering*. Google [Whitepaper] https://www.kaggle.com/whitepaper-prompt-engineering.

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

Carl, K., Dignam, C., Kochan, M., Alston, C., & Green, D. (2024). Discovering prompt engineering: A qualitative study of nonexpert teachers' interactions with ChatGPT. *Issues in Information Systems, 25*(4), 205–220. https://doi.org/10.48009/4_iis_2024_117

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

Celik, I., Kontkanen, S., Laru, J., & Dalyanci, A. A. (2026). Co-constructing adaptive lesson plans with GenAI: Pre-service teachers' Intelligent-TPACK and prompt engineering strategies. *Computers & Education, 241*, Article 105485. https://doi.org/10.1016/j.compedu.2025.105485

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chiu, T. K. F., Ahmad, Z., & Çoban, M. (2025). Development and validation of teacher artificial intelligence (AI) competence self-efficacy (TAICS) scale. *Education and Information Technologies, 30*, 6667–6685. https://doi.org/10.1007/s10639-024-13094-z

Correia, A.-P., Hickey, S., & Xu, F. (2025). Realizing the possibilities of the large language models: Strategies for prompt engineering in educational inquiries. *Theory Into Practice, 64*(4), 434–447. https://doi.org/10.1080/00405841.2025.2528545

Crosthwaite, P. R., Smala, S., & Spinelli, F. (2025). Prompting for pedagogy? Australian F-10 teachers' generative AI prompting use cases. *Australian Educational Researcher, 52*(3), 1795–1825. https://doi.org/10.1007/s13384-024-00787-0

Davila-Moran, R. C., Sanchez Soto, J. M., Lopez Gomez, H. E., Silva Infantes, M., Arias Lizares, A., Huanca Rojas, L. M., & Cama Flores, S. J. (2025). Brief prompt-engineering clinic substantially improves AI literacy and reduces technology anxiety in first-year teacher-education students: A pre-post pilot study. *Education Sciences, 15*(8), Article 1010. https://doi.org/10.3390/educsci15081010

ElSayary, A., Kuhail, M. A., & Hojeij, Z. (2025). Examining the role of prompt engineering in utilizing generative AI tools for lesson planning: Insights from teachers' experiences and perceptions. *Human Behavior and Emerging Technologies*. https://doi.org/10.1155/hbe2/9986139

Federiakin, D., Molerov, D., Zlatkin-Troitschanskaia, O., & Maur, A. (2024). Prompt engineering as a new 21st century skill. *Frontiers in Education, 9*, Article 1366434. https://doi.org/10.3389/feduc.2024.1366434

Fornell, C., & Larcker, D. (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

Gibreel, O., & Arpaci, I. (2025). Development and validation of the prompt engineering competence scale (PECS). *Information Development*. https://doi.org/10.1177/02666669251336455. Advance online publication.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis*. Cengage Learning EMEA.

Hsu, H.-P. (2025). From programming to prompting: Developing computational thinking through large language model-based generative artificial intelligence. *TechTrends, 69*(3), 485–506. https://doi.org/10.1007/s11528-025-01052-6

Huang, K.-L., Liu, Y.-C., & Dong, M.-Q. (2024). Incorporating AIGC into design ideation: A study on self-efficacy and learning experience acceptance under higher-order thinking. *Thinking Skills and Creativity, 52*, Article 101508. https://doi.org/10.1016/j.tsc.2024.101508

Hwang, M., Jeens, R., & Lee, H.-K. (2025). Exploring learner prompting behavior and its effect on ChatGPT-assisted English writing revision. *The Asia-Pacific Education Researcher, 34*, 1157–1167. https://doi.org/10.1007/s40299-024-00930-6

Jacobsen, L. J., & Weber, K. E. (2025). The promises and pitfalls of large language models as feedback providers: A study of prompt engineering and the quality of AI-driven feedback. *AI, 6*(2). https://doi.org/10.3390/ai6020035

Jang, A., Oh, M., & Song, M. O. (2025). Question-centered prompt engineering for nursing simulation debriefing: Model development and validation. *Nurse Education in Practice, 88*, Article 104541. https://doi.org/10.1016/j.nepr.2025.104541

Jin, Z., Meng, G., Wang, X., Wang, J., Liu, C., & Zhang, J. (2025). Understanding user prompting behavior in generative AI: A component analysis. *Proceedings of the Association for Information Science and Technology, 62*(1), 941–945. https://doi.org/10.1002/pra2.1318

Kabeer, A., Bhat, R. A., Antony, S., & Tramboo, I. A. (2025). Enhancing creative writing skills in secondary school students through prompt engineering and artificial intelligence. *Forum for Linguistic Studies, 7*(3), 800–815. https://doi.org/10.30564/fls.v7i3.8511

Kang, L., Shi, X., & Zhu, K. (2025). Uncovering the mediation of disciplinary literacy in the effect of GAI prompt engineering on pre-service teachers' instructional design. *Education and Information Technologies, 30*(16), 22779–22802. https://doi.org/10.1007/s10639-025-13657-8

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. The Guilford Press.

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "Scale for the assessment of non-experts' AI literacy": An exploratory factor analysis. *Computers in Human Behavior Reports, 12*(4), Article 100338. https://doi.org/10.1016/j.chbr.2023.100338

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2024). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies, 29*(9), 11483–11515. https://doi.org/10.1007/s10639-023-12249-8

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education, 22*(1). https://doi.org/10.1186/s41239-025-00503-7

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship, 49*(4), Article 102720. https://doi.org/10.1016/j.acalib.2023.102720

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727.

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The hull method for selecting the number of common factors. *Multivariate Behavioral Research, 46*(2), 340–364. https://doi.org/10.1080/00273171.2011.564527

Mendes, A., Parizotto, A., Garcia, R., Garcia, M., Guedes, G., Vilela, R., Valle, P., Balancieri, R., & Silva, W. (2025). CoderBot 2.0: Integrating LLM and prompt engineering in the evolution of an educational chatbot. In *Simpósio Brasileiro de Informática na Educação (SBIE)* (pp. 1497–1506). https://doi.org/10.5753/sbie.2025.12321. SBC.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1177/016146810610800610

Moorhouse, B. L., Ho, T. Y., Wu, C., & Wan, Y. (2025). Pre-service language teachers' task-specific large language model prompting practices. *RELC Journal*. https://doi.org/10.1177/00336882251313701. Advance online publication.

Mzwri, K., & Turcsányi-Szabó, M. (2025). The impact of prompt engineering and a generative AI-driven tool on autonomous learning: A case study. *Education Sciences, 15*(2), Article 199. https://doi.org/10.3390/educsci15020199

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ognibene, D., Donabauer, G., Theophilou, E., Koyuturk, C., Yavari, M., Bursic, S., Telari, A., Testa, A., Boiano, R., Taibi, D., Hernandez-Leo, D., Kruschwitz, U., & Ruskov, M. (2025). Use me wisely: AI-driven assessment for LLM prompting skills development. *Educational Technology & Society, 28*(3), 184–201. https://doi.org/10.30191/ETS.202507_28(3).SP12

Oliveira, L., Tavares, C., Strzelecki, A., & Silva, M. (2025). Prompting minds: Evaluating how students perceive generative AI's critical thinking dispositions. *Electronic Journal of E-Learning, 23*(2), 1–18. https://doi.org/10.34190/ejel.23.2.3986

Park, J., & Choo, S. (2024). Generative AI prompt engineering for educators: Practical strategies. *Journal of Special Education Technology, 40*(3), 411–417. https://doi.org/10.1177/01626434241298954

Pratschke, B. M. (2024). *Generative AI and education: Digital pedagogies, teaching innovation and learning design*. Springer. https://doi.org/10.1007/978-3-031-67991-9

Preacher, K. J., Zhang, Z., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A comparison of objective methods. *Multivariate Behavioral Research, 48*(1), 28–56. https://doi.org/10.1080/00273171.2012.710386

Qian, Y. (2025). Prompt engineering in education: A systematic review of approaches and educational applications. *Journal of Educational Computing Research*. https://doi.org/10.1177/07356331251365189. Advance online publication.

Rabardel, P. (2002). *People and technology in the workplace: A cognitive ergonomic approach* (H. Wood, Trans.). Université Paris 8. HAL Open Science https://hal.science/hal-01020705.

Selwyn, N., Ljungqvist, M., & Sonesson, A. (2025). When the prompting stops: Exploring teachers' work around the educational frailties of generative AI tools. *Learning, Media and Technology, 50*(3), 310–323. https://doi.org/10.1080/17439884.2025.2537959

Serra, P., & Oliveira, A. (2025). AI-powered prompt engineering for education 4.0: Transforming digital resources into engaging learning experiences. *Education Sciences, 15*(12), Article 1640. https://doi.org/10.3390/educsci15121640

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Sigot, M., & Tassoti, S. (2025). An investigation of change in prompting strategies in a semester-long course on the use of GenAI. *Journal of Chemical Education, 102*(6), 2507–2513. https://doi.org/10.1021/acs.jchemed.4c01287

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.

Tümen Akyıldız, S. (2026). Prompt engineering and creative pedagogical thinking: Insights from a qualitative study with teacher candidates. *Thinking Skills and Creativity, 60*, Article 102105. https://doi.org/10.1016/j.tsc.2025.102105

UNESCO. (2024). *AI competency framework for teachers*. UNESCO Publishing.

Van Wyk, M. M. (2025). Student teachers' leveraging GenAI tools for academic writing, design, and prompting in an ODeL course. *Open Praxis, 17*(1), 95–107. https://doi.org/10.55982/openpraxis.17.1.711

Velásquez-Henao, J. D., Franco-Cardona, C. J., & Cadavid-Higuita, L. (2023). Prompt engineering: A methodology for optimizing interactions with AI-language models in the field of engineering. *Dyna, 90*(230), 9–17. https://doi.org/10.15446/dyna.v90n230.111700

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*(1). https://doi.org/10.1186/s41239-024-00448-3

Wang, B., Rau, P.-L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology, 42*(9), 1324–1337. https://doi.org/10.1080/0144929X.2022.2072768

Wang, Y. Y., & Chuang, Y. W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies, 29*(4), 4785–4808. https://doi.org/10.1007/s10639-023-12015-w

Woo, D. J., Wang, D., Yung, T., & Guo, K. (2025). Effects of a prompt engineering intervention on undergraduate students' AI self-efficacy, AI knowledge and prompt engineering ability: A mixed methods study. *British Educational Research Journal*. https://doi.org/10.1002/berj.70087. Advance online publication.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2