OPEN ACCESS

GLODAL Emerging Educational Technologies

Emerging Educational Technologies, 2026, 1,
260104
https://doi.org/10.65553/EET.260104

Validating and Profiling Emotional Engagement in Human - GenAI Collaborative Language Learning

Songming Liu[a] Yongliang Wang[b]

[a]Macao Polytechnic University, China

[b]North China University of Water Resources and Electric Power, China

Submission date: May 1, 2026; Revision date: October 25, 2026; Online: November 20, 2026

Abstract

Despite growing interest in generative artificial intelligence (GenAI) in language education, emotional engagement in human - GenAI collaborative language learning remains insufficiently conceptualized and measured. To address this gap, this study examined the psychometric properties of a 15-item measure of emotional engagement and identified learner profiles. Data came from 504 Chinese university students. Exploratory factor analysis in one random half of the sample (n = 252) yielded three factors: Positive Affective Engagement, Affective Connectedness in Collaboration, and Emotional Disaffection. Confirmatory factor analysis in the other half supported this structure (CFI = .926, TLI = .911, RMSEA = .073, SRMR = .048) and indicated better fit than one- and two-factor alternatives. The subscales showed good internal consistency (ω = .84-.87), convergent validity (AVE = .51-.58), and discriminant validity (HTMT ≤ .52). Configural, metric, scalar, and strict measurement invariance were supported across gender and year of study. Latent profile analysis identified high (29.8%), moderate (50.4%), and low (19.8%) engagement profiles that differed in level rather than shape. These findings support a multidimensional conceptualization of emotional engagement in human-GenAI collaborative language learning and provide a brief, psychometrically supported instrument for examining learners' affective experiences.

Keywords: emotional engagement, generative artificial intelligence, human - GenAI collaboration, foreign language learning, scale validation, latent profile analysis

Copyright © 2026 The Authors
Yongliang Wang, corresponding author: wangyongliang@ncwu.edu.cn

Introduction

Since the public release of ChatGPT in late 2022, generative artificial intelligence (GenAI) has rapidly become an increasingly prominent resource in higher education (Chan & Hu, 2023; Kasneci et al., 2023) and primary education (Wong & Fong, 2025). Language learning is among the domains substantially influenced by these developments, as large language models can engage in conversations, provide explanations and translations, model target-language expressions, and offer immediate feedback on learners' writing (Barrot, 2023; Kohnke et al., 2023; Lee et al., 2026; Warschauer et al., 2023). Rather than merely consulting GenAI as a supplementary tool, language learners increasingly engage in collaborative processes involving idea generation, meaning negotiation, and iterative revision (Su et al., 2023; Watson Todd, 2026; Yan, 2023). Such interactions reflect an emerging form of human-AI collaboration in which learners and intelligent systems contribute to shared learning tasks (Järvelä et al., 2023; Molenaar, 2022). This evolving relationship highlights the importance of understanding not only how learners interact with GenAI but also how they experience these interactions emotionally.

Human–GenAI collaboration involves emotional as well as cognitive processes. While emotional engagement in human–GenAI collaboration can be systematically measured, the linguistic stability of LLM-generated texts across proficiency prompts remains high (Carlo, 2026). Previous research has consistently demonstrated that emotions influence learners' effort, persistence, and learning outcomes (Fredricks et al., 2004; Pekrun, 2006). In second language acquisition, research has expanded beyond its traditional emphasis on foreign language anxiety (Horwitz et al., 1986; Teimouri et al., 2019) to encompass positive emotional experiences, particularly enjoyment (Dewaele & MacIntyre, 2014; MacIntyre & Gregersen, 2012). Within this development, learner engagement has emerged as a central construct for explaining the quality of learners' participation in language learning activities (Hiver et al., 2024; Philp & Duchesne, 2016). Emotional engagement, referring to learners' affective reactions to and involvement in learning activities, represents an important component of this broader construct (Skinner et al., 2009). In human–GenAI collaborative language learning, such engagement may involve not only enjoyment and interest in collaborative tasks but also feelings of connection with GenAI and negative emotional responses when collaboration proves unsatisfactory. These potentially distinct experiences highlight the need for a multidimensional understanding of emotional engagement in this emerging learning context.

Despite growing research interest in GenAI-assisted language learning, several gaps remain in understanding and measuring learners' emotional engagement. First, existing quantitative research has predominantly examined learners' technological acceptance, perceptions, and behavioral intentions (Liu & Ma, 2024; Zou & Huang, 2023), although increasing attention has been directed toward emotional experiences and academic engagement (Wang & Xue, 2024; Zhang & Liu, 2025). Available instruments, however, have largely originated from conventional classroom contexts and may not adequately capture the relational characteristics of human-GenAI collaboration, particularly learners' perceived emotional connection with an AI partner (Nass & Moon, 2000; Skjuve et al., 2021). Moreover, positive affect and emotional disaffection have often been examined separately, limiting an integrated understanding of their distinct contributions to emotional engagement (Dewaele & MacIntyre, 2014). Second, limited evidence is available regarding whether measures of emotional engagement function equivalently across different learner groups, despite measurement invariance being essential for meaningful comparisons (Putnick & Bornstein, 2016). Third, existing variable-centered approaches provide valuable insights into general patterns of emotional engagement but may overlook heterogeneity in how positive affect, affective connectedness, and emotional disaffection coexist within individual learners (see a systematic review, Xu, 2026). Person-centered methods such as latent profile analysis can help identify such heterogeneity (Spurk et al., 2020). Although person-centered approaches have increasingly been applied to

engagement in GenAI-mediated language learning (Teng & Huang, 2025, 2026), research specifically examining the multidimensional profiles of emotional engagement in human-GenAI collaboration remains limited (Teng, 2026).

To address these gaps, the present study examines emotional engagement in human-GenAI collaborative language learning using data from 504 Chinese university students. Specifically, it evaluates the factor structure, reliability, validity, and measurement invariance of a 15-item instrument comprising Positive Affective Engagement, Affective Connectedness in Collaboration, and Emotional Disaffection. It further employs latent profile analysis to identify distinct patterns of emotional engagement and examines whether profile membership varies across learner background characteristics. By integrating psychometric evaluation with person-centered analysis, this study seeks to advance understanding of both the multidimensional structure of emotional engagement and the heterogeneity of learners' affective experiences in human-GenAI collaborative language learning.

Literature review

Emotional Engagement in Language Learning

Engagement refers to the quality of learners' active involvement in learning activities. Fredricks et al. (2004) conceptualized engagement as comprising behavioral, cognitive, and emotional dimensions, while subsequent scholarship further recognized its agentic and social components (Philp & Duchesne, 2016; Reeve & Tseng, 2011). In language education, engagement represents an important process through which motivation is translated into meaningful learning participation (Hiver et al., 2024; Oga-Baldwin, 2019; Svalberg, 2009). However, emotional engagement remains less consistently conceptualized and operationalized than other dimensions of engagement. A systematic review of two decades of research highlighted the frequent reliance on general indicators of enjoyment, interest, or liking to assess emotional engagement (Hiver et al., 2024). Similarly, Sinatra et al. (2015) emphasized the importance of aligning the specificity of engagement measures with the learning activities under investigation. These observations suggest that emotional engagement should be conceptualized and measured in relation to the particular characteristics of the learning context rather than treated as a uniform emotional response across different educational settings.

Two conceptual considerations are particularly relevant to understanding emotional engagement in language learning. The first concerns emotional valence. Skinner et al. (2009) distinguished emotional engagement, characterized by enthusiasm, interest, and enjoyment, from emotional disaffection, reflected in discouragement, anxiety, and frustration. Although negatively associated, these dimensions represent distinguishable emotional experiences, indicating that the absence of positive engagement does not necessarily imply the presence of negative emotions. Similar distinctions have been established in second language acquisition. Dewaele and MacIntyre (2014) demonstrated that foreign language enjoyment and anxiety were moderately negatively correlated rather than opposite ends of a single continuum, a distinction subsequently supported in the Chinese language learning context (Li et al., 2018). Control-value theory further suggests that positive and negative achievement emotions arise from learners' evaluations of control and task value and may therefore have different antecedents and consequences (Pekrun, 2006). Positive emotions may broaden attention and facilitate the development of learning resources (Fredrickson, 2001; MacIntyre & Gregersen, 2012), whereas anxiety has been consistently associated with poorer language achievement (Teimouri et al., 2019). Taken together, these findings underscore the importance of distinguishing positive emotional engagement from emotional disaffection when examining learners' affective experiences.

The second consideration concerns the interpersonal and relational orientation of emotional engagement. Emotional experiences in language learning are shaped not only by learning tasks but also by interactions with others involved in the learning process. Philp and Duchesne

(2016) emphasized that engagement in language classrooms encompasses learners' emotional connections with interlocutors, reflecting the inherently social nature of language learning. From the perspective of self-determination theory, relatedness refers to the psychological need to experience connection and support in interpersonal relationships, which contributes to the maintenance of learning motivation (Ryan & Deci, 2000). Empirical research has similarly demonstrated associations between teacher-created emotional climates and learners' enjoyment and engagement (Dewaele & Li, 2021). These perspectives suggest that emotional engagement may encompass both affective responses to learning activities and perceived connections with interaction partners. Although such relational experiences have traditionally been examined in human interactions, their relevance to technology-mediated learning warrants further consideration, particularly when learners interact with conversational systems capable of providing responsive and personalized support.

GenAI as a Collaborative Partner in Language Learning

Before the emergence of GenAI, language learning chatbots generally relied on scripted interactions and demonstrated limitations in sustaining learners' interest over time. Fryer et al. (2017) found that learners' interest in speaking activities declined following initial interactions with chatbot partners, whereas comparable declines were not observed with human partners. This potential novelty effect was also highlighted in a subsequent systematic review (Huang et al., 2022). Compared with earlier chatbot systems, GenAI offers greater conversational flexibility, responsiveness, and adaptability, enabling learners to engage in open-ended exchanges, obtain personalized explanations, and receive immediate feedback (Kohnke et al., 2023). Research on second language writing has documented the integration of ChatGPT into different stages of composing, including idea generation, drafting, and revision (Barrot, 2023; Su et al., 2023; Yan, 2023). Nevertheless, analyses of learner interaction logs suggest considerable variation in how individuals collaborate with AI systems (Wang et al., 2023). These developments indicate that GenAI can support increasingly interactive forms of language learning, although learners' experiences may differ depending on how collaboration unfolds.

The expanding capabilities of GenAI have also encouraged a reconsideration of the relationship between learners and educational technologies. Molenaar (2022) conceptualized hybrid human–AI learning as a process in which learners and intelligent systems share aspects of control over learning activities, while Järvelä et al. (2023) highlighted the potential contribution of AI systems to collaborative learning regulation. Similarly, Jeon and Lee (2023) described the relationship between teachers and ChatGPT as complementary rather than substitutive. Within this perspective, GenAI is not merely a passive information source but can function as an interactive partner supporting the completion of shared learning tasks. Importantly, this collaborative orientation introduces an affective dimension into learner-technology interactions. Learners may evaluate collaboration not only according to the usefulness of AI-generated responses but also in terms of whether the interaction feels supportive, responsive, and emotionally comfortable. Consequently, understanding human-GenAI collaboration requires attention to learners' subjective experiences of interacting with an intelligent conversational partner.

The possibility of experiencing relational responses toward non-human systems is supported by research on human–computer interaction. Nass and Moon (2000) demonstrated that individuals may apply social expectations and interaction norms to computers despite recognizing their non-human nature. Research on companion chatbots further suggests that users can develop perceived feelings of being understood and supported through repeated interactions (Skjuve et al., 2021). In language learning, such perceived support may be particularly relevant because fear of negative evaluation constitutes an important component

of foreign language anxiety (Horwitz et al., 1986). Interacting with GenAI may provide opportunities to ask questions, explore linguistic expressions, and practice language use without the immediate interpersonal evaluation associated with some classroom situations. However, these perceived relational benefits should not be equated with genuine interpersonal understanding or emotional reciprocity. Moreover, GenAI-generated responses may be inaccurate, ambiguous, or insufficiently responsive to learners' needs, potentially generating disappointment, confusion, and frustration (Barrot, 2023; Warschauer et al., 2023). Human-GenAI collaboration may therefore involve positive emotional experiences, perceived affective connectedness, and emotional disaffection, highlighting the importance of examining these experiences as distinguishable dimensions.

Emotional Engagement in Human–GenAI Collaborative Language Learning

Building on these theoretical and empirical perspectives, the present study conceptualizes emotional engagement in human-GenAI collaborative language learning as comprising three related but distinguishable dimensions. Positive Affective Engagement encompasses enjoyment, interest, excitement, accomplishment, and anticipation associated with collaborative language learning activities. This dimension reflects positive emotional involvement in learning and corresponds to the activating positive emotions emphasized in control-value theory (Pekrun, 2006) and the engagement dimension proposed by Skinner et al. (2009). Affective Connectedness in Collaboration refers to learners' perceived emotional connection with GenAI, including feelings of support, understanding, reduced isolation, and comfort during interaction. This dimension draws on the relational characteristics of language learning engagement (Philp & Duchesne, 2016; Ryan & Deci, 2000) and extends consideration of perceived social responses to non-human conversational partners (Nass & Moon, 2000; Skjuve et al., 2021). Emotional Disaffection encompasses discouragement, disappointment, anxiety, stress, and frustration arising from unsatisfactory collaborative learning experiences. It is consistent with the distinction between engagement and disaffection proposed by Skinner et al. (2009) and with research demonstrating the relevance of negative emotions in foreign language learning (Dewaele & MacIntyre, 2014; Horwitz et al., 1986). Although these dimensions may be interrelated, they capture different aspects of learners' emotional experiences. Positive Affective Engagement concerns enjoyment and interest in collaborative learning, Affective Connectedness in Collaboration concerns perceived relational support during interaction, and Emotional Disaffection concerns negative emotional responses to difficulties encountered in collaboration.

Existing instruments provide important foundations for examining these experiences but do not fully address their combined characteristics in human–GenAI collaborative language learning. Measures of foreign language enjoyment and anxiety primarily assess positive and negative emotional experiences within conventional language learning contexts (Dewaele & MacIntyre, 2014; Li et al., 2018). Multidimensional engagement instruments generally position emotional engagement alongside behavioral and cognitive engagement, often operationalizing it through a limited number of positively oriented items (Hiver et al., 2024; Reeve & Tseng, 2011). Similarly, technology acceptance measures typically examine enjoyment or attitudes as predictors of technology use intentions rather than as components of multidimensional emotional engagement (Liu & Ma, 2024; Zou & Huang, 2023). Although these approaches capture relevant aspects of learners' affective experiences, they do not explicitly integrate positive emotional involvement, perceived affective connectedness with GenAI, and emotional disaffection within a unified measurement framework. To address this limitation, the present study examines items adapted from previous research on learner-GenAI relationships and cognitive, social, and emotional experiences in foreign language learning (Guo & Li, 2022; Jin, 2025). This approach allows the three proposed dimensions to be evaluated within the specific context of human-GenAI collaborative language learning.

An additional consideration concerns whether emotional engagement can be measured consistently across different groups of learners. Comparisons of latent constructs require evidence that measurement instruments operate equivalently across relevant populations (Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). Without establishing measurement invariance, differences in observed scores may reflect variations in item interpretation or measurement properties rather than substantive differences in emotional engagement. This issue is particularly relevant to human-GenAI collaborative language learning, where learners with different demographic and educational backgrounds may encounter different forms of technology-mediated interaction. Nevertheless, measurement invariance has received limited attention in existing instruments examining GenAI-related language learning experiences. Evaluating invariance across gender and year of study can therefore provide important evidence concerning the comparability of emotional engagement scores and the suitability of the instrument for subsequent group-based investigations.

Finally, understanding emotional engagement requires consideration of both general relationships among constructs and differences in how emotional experiences are combined within individual learners. Previous research has predominantly adopted variable-centered approaches, which describe associations and average tendencies across samples. In contrast, person-centered methods, such as latent profile analysis, identify subgroups of learners characterized by distinct configurations of emotional experiences (Spurk et al., 2020). Importantly, such configurations may differ in level or shape. Level differences occur when profiles display broadly similar patterns across dimensions but differ in their overall intensity, whereas shape differences emerge when learners exhibit qualitatively different combinations of emotional experiences, such as high enjoyment alongside substantial emotional disaffection (Marsh et al., 2009; Morin & Marsh, 2015). Distinguishing between these possibilities has implications for understanding learner heterogeneity and informing instructional support. Profiles differing primarily in level may indicate varying degrees of emotional involvement, whereas profiles differing in shape may suggest more complex affective configurations requiring differentiated forms of support. Accordingly, examining latent profiles can complement psychometric validation by clarifying how the proposed dimensions of emotional engagement coexist across learners in human-GenAI collaborative language learning.

Rationale for the Present Study

The present study had two aims: to evaluate the psychometric properties of a three-dimensional measure of emotional engagement in human-GenAI collaborative language learning, and to describe heterogeneity in emotional engagement among learners. Four research questions were addressed.

RQ1. What is the factor structure of the emotional engagement measure, and does a three-factor model fit better than alternative models?

RQ2. Do the subscales show adequate reliability, convergent validity, and discriminant validity?

RQ3. Is the measure invariant across gender and year of study?

RQ4. What latent profiles of emotional engagement can be identified, and is profile membership related to gender, year of study, English proficiency, and the GenAI tool used?

Methodology

Participants

Participants were 504 students from three universities in China who used GenAI tools in their foreign language learning. The sample comprised 230 men (45.6%) and 274 women (54.4%).

By year of study, 115 (22.8%) were first-year undergraduates, 145 (28.8%) second-year, 108 (21.4%) third-year, and 90 (17.9%) fourth-year, and 46 (9.1%) were postgraduates. Participants named the GenAI tool they currently used. The most frequent were ChatGPT (n = 132, 26.2%), ERNIE Bot (n = 77, 15.3%), Doubao (n = 66, 13.1%), Kimi (n = 59, 11.7%), DeepSeek (n = 48, 9.5%), and Zhipu Qingyan (n = 43, 8.5%); the remaining 79 participants (15.7%) used Tongyi Qianwen, Claude, Tencent Yuanbao, or iFlytek Spark. College English Test Band 4 (CET-4) scores were reported by 371 participants (M = 462.41, SD = 62.38, range = 277–639).

Procedure

Data were collected over a one-week period through Wenjuanxing, an online survey platform widely used in China. Faculty members at the three participating universities distributed the survey link to students who met the eligibility criteria. The first page of the questionnaire presented an electronic informed-consent statement explaining the purpose of the study, the voluntary nature of participation, the confidentiality of responses, and participants' right to discontinue the survey. Participants reviewed this information before proceeding to the questionnaire.

Participants first provided demographic information and then completed the study measures. The emotional engagement items analyzed here were administered as part of a larger questionnaire. The questionnaire presented the English and Chinese versions of the items in parallel.

After data collection, responses were screened for data quality. Questionnaires were classified as invalid when they showed one or more of the following characteristics: an unreasonably short completion time, invariant response patterns across a series of items, incomplete responses, or evident inconsistencies among responses. On the basis of these criteria, 37 questionnaires were excluded, leaving 504 valid cases for analysis. The 15 emotional engagement items had no missing values.

Measures

Emotional engagement in human–GenAI collaborative language learning was assessed using a 15-item instrument comprising three dimensions: Positive Affective Engagement, Affective Connectedness in Collaboration, and Emotional Disaffection, with five items representing each dimension (see the Appendix). The items were adapted from previous research examining learner–GenAI relationships and cognitive, social, and emotional experiences in foreign language learning (Guo & Li, 2022; Jin, 2025). Participants responded using a 5-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree). All responses were reverse-coded before analysis to ensure that higher scores indicated stronger endorsement of the construct represented by each dimension. Accordingly, higher scores on Positive Affective Engagement and Affective Connectedness in Collaboration reflected greater positive emotional involvement and perceived connectedness, respectively, whereas higher scores on Emotional Disaffection indicated greater negative emotional experiences.

Positive Affective Engagement assessed learners' enjoyment, interest, excitement, sense of accomplishment, and anticipation associated with human–GenAI collaborative language learning (sample item: "I enjoy learning through Human–GenAI collaboration"). Affective Connectedness in Collaboration captured learners' perceived emotional connection with GenAI, including feelings of support, understanding, reduced isolation, and comfort during interaction (sample item: "During Human–GenAI collaborative learning, I feel a sense of connection with GenAI"). Emotional Disaffection measured learners' negative emotional responses to collaborative learning experiences, including discouragement, disappointment, anxiety, stress, and frustration (sample item: "During Human–GenAI collaborative foreign language learning, I feel discouraged when my progress is unsatisfactory"). Together, these

dimensions were intended to capture distinguishable aspects of learners' affective experiences during human-GenAI collaboration.

Background variables included gender, year of study, self-reported College English Test Band 4 (CET-4) scores, and the GenAI tool currently used by participants.

Data Analysis

The analysis proceeded in five steps. First, item quality was examined through descriptive statistics, corrected item-total correlations, Cronbach's α if an item was deleted, and an extreme-groups comparison of the top and bottom 27% of scorers on each subscale.

Second, the sample was randomly divided into two halves of 252 cases, a practice recommended so that a structure is not explored and confirmed on the same data (Fabrigar et al., 1999; Worthington & Whittaker, 2006). Exploratory factor analysis (EFA) was conducted on the first half using principal axis factoring with promax rotation, because the factors were expected to correlate. The number of factors was determined by parallel analysis (Horn, 1965) with 1,000 simulated data sets and the 95th-percentile criterion.

Third, confirmatory factor analysis (CFA) was conducted on the second half. Four models were compared: a one-factor model (M1); a two-factor model that combined Positive Affective Engagement and Affective Connectedness into one positive factor (M2); the hypothesized correlated three-factor model (M3); and a bifactor model with a general factor and three specific factors (M4). A second-order model was not tested separately because, with three first-order factors, it is statistically equivalent to M3 (Brown, 2015). Models were estimated by maximum likelihood. Skewness and kurtosis were below |1| for every item, and 5-point items can be treated as continuous under these conditions with little bias (Rhemtulla et al., 2012). Fit was judged by the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA) with its 90% confidence interval, and the standardized root mean square residual (SRMR). Values of CFI and TLI at or above .90 and of RMSEA and SRMR at or below .08 were taken as acceptable, and CFI and TLI at or above .95 with RMSEA at or below .06 as good (Hu & Bentler, 1999; Marsh et al., 2004). The Akaike and Bayesian information criteria (AIC and BIC) were used to compare models.

Fourth, reliability and validity were evaluated in the full sample. Internal consistency was estimated with Cronbach's α and McDonald's ω (Dunn et al., 2014; Hayes & Coutts, 2020), together with composite reliability (CR). Convergent validity was assessed by the average variance extracted (AVE), with .50 as the threshold, and discriminant validity by the Fornell-Larcker criterion (Fornell & Larcker, 1981) and the heterotrait-monotrait ratio of correlations (HTMT), with .85 as the threshold (Henseler et al., 2015). Measurement invariance across gender and year of study was tested with multigroup CFA in a sequence of configural, metric, scalar, and strict models (Vandenberg & Lance, 2000). Invariance was retained when CFI decreased by no more than .010 and RMSEA increased by no more than .015 (Chen, 2007; Cheung & Rensvold, 2002).

Fifth, latent profile analysis (LPA) was conducted on the three subscale mean scores, with variances constrained to be equal across profiles and indicators assumed to be locally independent. Models with one to six profiles were compared on AIC, BIC, the sample-size-adjusted BIC (SABIC), entropy (Celeux & Soromenho, 1996), the bootstrap likelihood ratio test (BLRT) with 100 bootstrap samples, and the size and interpretability of the smallest profile (Nylund et al., 2007; Spurk et al., 2020). Each model was estimated from multiple random starts. Participants were assigned to their most likely profile, and profiles were compared on background variables using chi-square tests and one-way analysis of variance.

Because all data were self-reported at one time point, common method variance was examined. A single factor accounted for 35.8% of the variance in the 15 items, below the conventional 50%

threshold, and the one-factor CFA model fit the data poorly (see Section 4.3), which suggests that a single method factor does not account for the covariation among items (Podsakoff et al., 2003).

Results

Descriptive Statistics and Item Analysis

Table 1 presents the item statistics. Item means ranged from 3.89 to 3.95 for Positive Affective Engagement, from 3.58 to 3.62 for Affective Connectedness in Collaboration, and from 2.30 to 2.33 for Emotional Disaffection. Subscale means were 3.93 (SD = 0.63), 3.60 (SD = 0.68), and 2.31 (SD = 0.64), respectively. One-sample t tests showed that all three differed from the scale midpoint of 3, t(503) = 33.25, 19.69, and −23.92, all ps < .001. Participants thus tended to endorse positive affect and connectedness and to disagree with statements of disaffection. Skewness ranged from −0.40 to 0.34 and kurtosis from −0.66 to 0.15, and no item showed a floor or ceiling effect. Corrected item-total correlations ranged from .58 to .76. Every item discriminated between high and low scorers, ts = 17.85-27.76, and deleting any item would not have increased the α of its subscale. All 15 items were retained.

Table 1. Descriptive Statistics and Item Analysis for the 15 Items (N = 504)

Item	Wording	M	SD	Skewness	Kurtosis	CITC	α if deleted	t
Positive Affective Engagement ***(α = .835, subscale M = 3.93, SD = 0.63)***								
1	I enjoy learning through Human–GenAI collaboration.	3.93	0.80	−0.29	−0.51	.603	.811	23.20
2	I feel interested when exploring new foreign language expressions, linguistic knowledge, or related information together with AI.	3.89	0.80	−0.28	−0.50	.675	.791	23.39
3	I feel excited when human–AI collaboration helps me make progress on challenging foreign language learning tasks.	3.95	0.80	−0.25	−0.66	.637	.802	24.28
4	I feel a sense of accomplishment after completing foreign language learning tasks in collaboration with AI.	3.92	0.84	−0.40	−0.37	.684	.788	23.69
5	I look forward to carrying out foreign language learning tasks through human–AI collaboration.	3.95	0.80	−0.34	−0.47	.581	.817	19.72
Affective Connectedness in Collaboration ***(α = .871, subscale M = 3.60, SD = 0.68)***								
6	During Human–GenAI collaborative learning, I feel a sense of connection with GenAI.	3.62	0.86	−0.19	−0.31	.669	.851	24.19
7	I feel that the AI provides me with support.	3.58	0.83	−0.05	−0.35	.675	.849	22.26
8	When the AI responds to my questions, I feel that my learning needs are understood.	3.60	0.86	−0.10	−0.47	.758	.828	27.76
9	When dealing with challenging tasks, collaborating with AI makes me feel less isolated and helpless.	3.60	0.81	0.01	−0.43	.705	.842	24.57

Item	Wording	M	SD	Skewness	Kurtosis	CITC	α if deleted	t
10	I feel comfortable asking the AI questions.	3.58	0.82	−0.22	−0.14	.677	.848	21.51
Emotional Disaffection *(α = .851, subscale M = 2.31, SD = 0.64)*								
11	During Human–GenAI collaborative foreign language learning, I feel discouraged when my progress is unsatisfactory.	2.30	0.79	0.04	−0.43	.682	.816	21.70
12	I feel disappointed when the collaboration does not bring the expected improvement.	2.32	0.84	0.11	−0.40	.679	.816	23.81
13	I feel anxious when I obtain unsatisfactory results.	2.32	0.80	0.20	−0.26	.675	.818	20.91
14	I feel stressed when I am confused by AI-generated feedback.	2.33	0.83	0.13	−0.46	.672	.818	22.18
15	I feel frustrated when the AI cannot adequately resolve my difficulties.	2.30	0.81	0.34	0.15	.604	.836	17.85

Note. CITC = corrected item–total correlation with the remaining four items of the same subscale. The *t* values compare the top and bottom 27% of scorers on the subscale total (Welch's test); all *p*s < .001.

Exploratory Factor Analysis

The data in the first half-sample were suitable for factor analysis, KMO = .858, Bartlett's χ 2(105) = 1868.71, p < .001. The first three eigenvalues (5.51, 2.60, and 1.55) exceeded the corresponding 95th-percentile eigenvalues from the parallel analysis (1.53, 1.40, and 1.32), whereas the fourth (0.83) did not (1.25; Figure 1). Three factors were extracted, accounting for 55.7% of the variance. As Table 2 shows, every item loaded on its intended factor, with primary loadings between .62 and .84 and no cross-loading above .13. Communalities ranged from .41 to .70. Positive Affective Engagement correlated .45 with Affective Connectedness in Collaboration and −.52 with Emotional Disaffection, and the latter two factors correlated −.26.

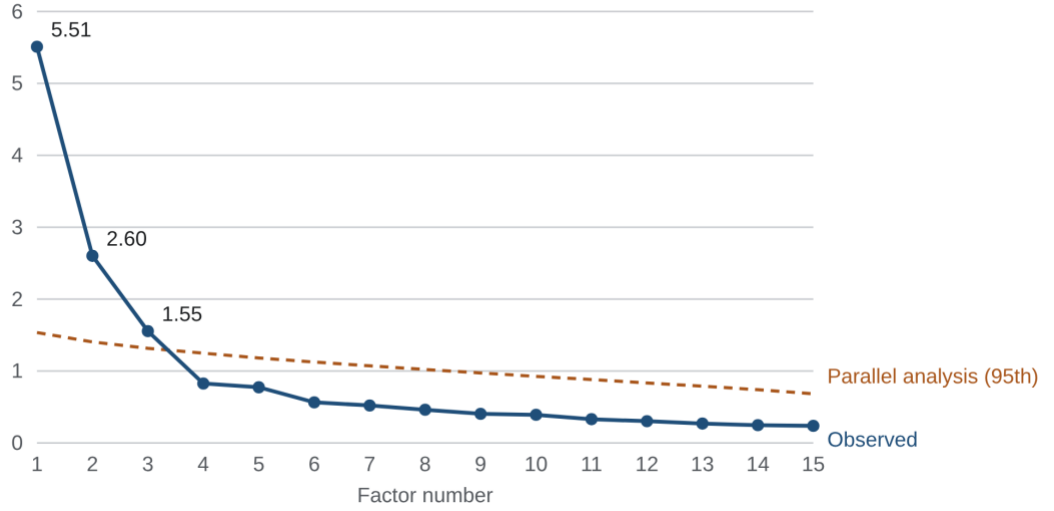


Figure 1 Scree Plot With Parallel Analysis (n = 252)

Table 2. Promax-Rotated Pattern Matrix from the Exploratory Factor Analysis (n = 252)

Item	Wording	Factor 1	Factor 2	Factor 3	h^2
1	I enjoy learning through Human–GenAI collaboration.	**.648**	.088	.021	.465
2	I feel interested when exploring new foreign language expressions, linguistic knowledge, or related information together with AI.	**.711**	.023	−.049	.560
3	I feel excited when human–AI collaboration helps me make progress on challenging foreign language learning tasks.	**.768**	−.033	.023	.552
4	I feel a sense of accomplishment after completing foreign language learning tasks in collaboration with AI.	**.774**	−.003	−.006	.602
5	I look forward to carrying out foreign language learning tasks through human–AI collaboration.	**.623**	−.029	−.057	.412
6	During Human–GenAI collaborative learning, I feel a sense of connection with GenAI.	−.008	**.739**	.005	.539
7	I feel that the AI provides me with support.	.063	**.736**	.028	.576
8	When the AI responds to my questions, I feel that my learning needs are understood.	.014	**.840**	.043	.699
9	When dealing with challenging tasks, collaborating with AI makes me feel less isolated and helpless.	−.026	**.774**	−.011	.587
10	I feel comfortable asking the AI questions.	−.017	**.783**	−.030	.614
11	During Human–GenAI collaborative foreign language learning, I feel discouraged when my progress is unsatisfactory.	−.038	.054	**.738**	.556
12	I feel disappointed when the collaboration does not bring the expected improvement.	−.002	−.004	**.772**	.599
13	I feel anxious when I obtain unsatisfactory results.	−.045	.062	**.770**	.607
14	I feel stressed when I am confused by AI-generated feedback.	−.028	.013	**.710**	.521
15	I feel frustrated when the AI cannot adequately resolve my difficulties.	.098	−.132	**.685**	.462

Note. Factor 1 = Positive Affective Engagement; Factor 2 = Affective Connectedness in Collaboration; Factor 3 = Emotional Disaffection. Loadings of |.40| or greater are in bold. h2 = communality.

Confirmatory Factor Analysis

Table 3 reports the fit of the competing models. In the holdout sample, the one-factor model (M1) and the two-factor model (M2) fit poorly. The hypothesized three-factor model (M3) fit acceptably, $\chi 2(87) = 204.91$, CFI = .926, TLI = .911, RMSEA = .073, 90% CI [.060, .086], SRMR = .048, and had much lower AIC and BIC than M1 and M2. Results in the full sample were similar, $\chi 2(87) = 360.22$, CFI = .921, TLI = .904, RMSEA = .079, 90% CI [.071, .087], SRMR = .040. Standardized loadings in the holdout sample ranged from .61 to .83 (Table 4).

The bifactor model (M4) had somewhat better fit indices than M3 but was not retained, for three reasons. Its BIC in the holdout sample (7899) was higher than that of M3 (7892). Its general factor was defined mainly by Items 1-3, whose loadings on their specific factor were near zero or negative (Table 4), so the general factor largely duplicated Positive Affective Engagement instead of representing variance common to all 15 items. And bifactor models tend to fit better than correlated-factor models even when they are not the data-generating structure (Bonifay et al., 2017; Reise, 2012). A higher-order interpretation was not supported either. When the correlations among the three factors were expressed as loadings on a second-order factor, the standardized loading of Positive Affective Engagement was .988 in the holdout sample and .999 in the full sample.

The RMSEA of M3 was at the upper end of the acceptable range. Inspection of residual correlations showed that the misfit was localized in a few pairs of items within the same factor, most clearly Items 4 and 5, Items 9 and 10, Items 14 and 15, and Items 11 and 12. Allowing these residuals to covary produced good fit in the holdout sample, $\chi 2(83) = 90.28$, CFI = .995, TLI = .994, RMSEA = .019, SRMR = .037 (M5 in Table 3), without changing the pattern of loadings. Because these modifications were made post hoc, M5 is reported for information only, and M3 is retained as the measurement model.

Table 3. Fit Indices for Competing Measurement Models

Model	χ^2	df	χ^2/df	CFI	TLI	RMSEA [90% CI]	SRMR	AIC	BIC
Holdout sample (n = 252)									
M1: One-factor	840.31	90	9.34	.532	.454	.182 [.171, .193]	.145	8352	8511
M2: Two-factor	511.10	89	5.74	.737	.689	.137 [.126, .149]	.114	8025	8188
M3: Correlated three-factor	204.91	87	2.36	.926	.911	.073 [.060, .086]	.048	7723	7892
M4: Bifactor	145.24	75	1.94	.956	.939	.061 [.046, .076]	.039	7687	7899
M5: M3 + 4 residual covariances	90.28	83	1.09	.995	.994	.019 [.000, .041]	.037	7616	7800
Full sample (N = 504)									
M1: One-factor	1727.72	90	19.20	.525	.445	.190 [.182, .198]	.147	16686	16876
M2: Two-factor	1028.00	89	11.55	.727	.678	.145 [.137, .153]	.121	15988	16182
M3: Correlated three-factor	360.22	87	4.14	.921	.904	.079 [.071, .087]	.040	15324	15527
M4: Bifactor	263.51	75	3.51	.945	.923	.071 [.061, .080]	.034	15251	15505
M5: M3 + 5 residual covariances	102.79	82	1.25	.994	.992	.022 [.000, .035]	.027	15077	15300

Note. M2 combines Positive Affective Engagement and Affective Connectedness in Collaboration into one factor. M5 adds residual covariances for item pairs 4–5, 9–10, 14–15, 11–12 in the holdout sample and 14–15, 4–5, 9–10, 11–12, 2–3 in the full sample. CI = confidence interval.

Table 4. Standardized Factor Loadings for the Three-Factor and Bifactor Models

Item	Factor	M3 λ (n = 252)	M3 λ (N = 504)	M4 general λ (N = 504)	M4 specific λ (N = 504)
1	PAE	.661	.671	.652	.044
2	PAE	.796	.779	.807	−.091
3	PAE	.719	.732	.775	−.180
4	PAE	.713	.738	.735	.548
5	PAE	.615	.626	.583	.359
6	ACC	.733	.737	.314	.673
7	ACC	.710	.736	.370	.633
8	ACC	.830	.827	.365	.748
9	ACC	.754	.761	.356	.668
10	ACC	.682	.731	.343	.642
11	ED	.762	.775	−.357	.714
12	ED	.746	.776	−.395	.683
13	ED	.710	.745	−.400	.621
14	ED	.739	.711	−.399	.568
15	ED	.655	.643	−.362	.512

Note. PAE = Positive Affective Engagement; ACC = Affective Connectedness in Collaboration; ED = Emotional Disaffection; M3 = correlated three-factor model; M4 = bifactor model.

Reliability and Validity

Table 5 summarizes reliability and convergent validity. Cronbach's α ranged from .835 to .871 and McDonald's ω from .836 to .872. The AVE of each subscale exceeded .50. As Table 6 shows, the square root of each AVE (.71-.76) was larger than the absolute correlations of that factor with the other two (.24-.52), and the HTMT ratios (.25-.52) were well below .85. The three subscales were therefore reliable and empirically distinct.

Table 5. Reliability and Convergent Validity (N = 504)

Subscale	α	ω	CR	AVE	Standardized λ range
Positive Affective Engagement	.835	.836	.836	.506	.626–.779
Affective Connectedness in Collaboration	.871	.872	.872	.577	.731–.827
Emotional Disaffection	.851	.851	.852	.536	.643–.776

Note. ω and composite reliability (CR) were computed from the M3 solution. AVE = average variance extracted; λ = standardized factor loading.

Table 6. Discriminant Validity (N = 504)

Subscale	1	2	3
1. Positive Affective Engagement	.711	*.465*	*.524*
2. Affective Connectedness in Collaboration	.464	.759	*.250*
3. Emotional Disaffection	−.524	−.244	.732

Note. Square roots of AVE are on the diagonal in bold. Latent correlations from M3 are below the diagonal, and HTMT ratios are above the diagonal in italics.

Measurement Invariance

Table 7 presents the invariance tests. For gender, constraining loadings, intercepts, and residual variances in turn produced nonsignificant changes in $\chi 2$, changes in CFI between .000 and +.003, and decreases in RMSEA. For year of study, the largest decrease in CFI was .004, at the scalar step, and RMSEA again decreased at every step. Configural, metric, scalar, and strict invariance were therefore supported for both grouping variables. The absolute fit of the configural models was similar to that of M3 in the full sample. The postgraduate group was small (n = 46), and the results for year of study should be read with this in mind.

Table 7. Tests of Measurement Invariance Across Gender and Year of Study

Model	χ^2	df	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$ (Δdf)	p	ΔCFI	ΔRMSEA
Gender (male n = 230, female n = 274)										
Configural	461.47	174	.917	.899	.081	.046	—	—	—	—
Metric	467.00	186	.919	.908	.077	.048	5.54 (12)	.938	+0.002	−0.004
Scalar	480.63	198	.918	.913	.075	.049	13.62 (12)	.325	0.000	−0.002
Strict	485.63	213	.921	.922	.071	.049	5.01 (15)	.992	+0.003	−0.004
Year of study (five groups)										
Configural	731.43	435	.915	.897	.082	.064	—	—	—	—
Metric	774.84	483	.916	.909	.077	.074	43.42 (48)	.661	+0.001	−0.005
Scalar	835.40	531	.912	.913	.075	.077	60.56 (48)	.105	−0.004	−0.002
Strict	887.45	591	.915	.924	.071	.080	52.04 (60)	.758	+0.002	−0.005

Note. Group sizes for year of study were 115 (first year), 145 (second year), 108 (third year), 90 (fourth year), and 46 (postgraduate). Each model is compared with the preceding one.

Latent Profiles of Emotional Engagement

Table 8 shows the fit of the latent profile models. BIC was lowest for the three-profile model. SABIC was slightly lower for the four-profile model, but the additional profile comprised only 3.5% of the sample (about 18 participants), too few to be estimated stably; profile size and interpretability should be weighed alongside statistical indices (Spurk et al., 2020). The BLRT was significant up to four profiles and nonsignificant thereafter. The three-profile model was retained on the grounds of BIC, profile size, and parsimony. Its entropy was .695 and the mean posterior probabilities of the assigned profiles ranged from .85 to .88.

Table 8. Fit Statistics for Latent Profile Models with One to Six Profiles (N = 504)

Profiles	LL	AIC	BIC	SABIC	Entropy	BLRT p	Profile proportions
1	−1492.61	2997.22	3022.55	3003.51	—	—	100.0%
2	−1408.29	2836.59	2878.81	2847.07	.595	.010	52.6% / 47.4%
3	−1387.96	2803.91	2863.03	2818.59	.695	.010	49.4% / 30.4% / 20.2%
4	−1380.15	2796.31	2872.31	2815.18	.784	.020	43.7% / 28.6% / 24.2% / 3.5%
5	−1377.31	2798.62	2891.51	2821.68	.677	.683	28.5% / 25.7% / 24.1% / 18.1% / 3.5%
6	−1370.73	2793.47	2903.25	2820.73	.669	.188	22.7% / 22.2% / 20.6% / 18.4% / 12.5% / 3.7%

Note. The retained solution is in bold. LL = log-likelihood; SABIC = sample-size-adjusted BIC; BLRT = bootstrap likelihood ratio test comparing *k* with *k* − 1 profiles.

The profiles are shown in Figure 2 and Table 9. The first, labeled high engagement (n = 150, 29.8%), combined very high Positive Affective Engagement (M = 4.67) with high Affective Connectedness (M = 3.95) and low Emotional Disaffection (M = 1.89). The second, moderate engagement (n = 254, 50.4%), was close to the sample means on all three subscales. The third, low engagement (n = 100, 19.8%), scored near the scale midpoint on all three subscales (Ms = 3.05, 3.06, and 2.83). The profiles differed significantly on each subscale, F(2, 501) = 1101.99 for Positive Affective Engagement, 64.92 for Affective Connectedness, and 88.49 for Emotional Disaffection, all ps < .001. The profiles were ordered in the same way on every indicator: the higher a profile's positive affect, the higher its connectedness and the lower its disaffection. No profile combined high positive affect with high disaffection.

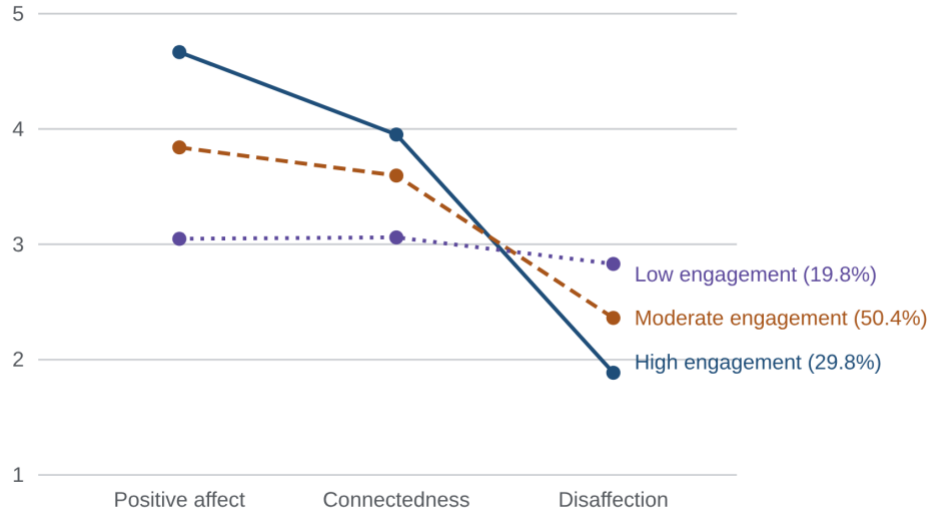


Figure 2 Subscale Means of the Three Latent Profiles

Note. Positive affect = Positive Affective Engagement; Connectedness = Affective Connectedness in Collaboration; Disaffection = Emotional Disaffection.

Table 9. Subscale Means and Standard Deviations by Latent Profile

Profile	n (%)	Positive Affective Engagement	Affective Connectedness	Emotional Disaffection	Mean posterior probability
High engagement	150 (29.8%)	4.67 (0.24)	3.95 (0.56)	1.89 (0.57)	.884
Moderate engagement	254 (50.4%)	3.84 (0.28)	3.60 (0.63)	2.36 (0.53)	.853
Low engagement	100 (19.8%)	3.05 (0.30)	3.06 (0.60)	2.83 (0.59)	.851
Total sample	504	3.93 (0.63)	3.60 (0.68)	2.31 (0.64)	—

Note. Subscale entries are *M* (*SD*) on the 1–5 scale. Profile membership is based on the highest posterior probability.

Table 10 compares the profiles on background variables. Profile membership was not significantly associated with gender, χ 2(2) = 5.36, p = .068, Cramér's V = .10; year of study, χ 2(8) = 11.06, p = .199, V = .11; or GenAI tool, χ 2(12) = 11.39, p = .496, V = .11. CET-4 scores did not differ among profiles, F(2, 368) = 0.78, p = .458.

Table 10. *Background Characteristics by Latent Profile*

Variable	High engagement (n = 150)	Moderate engagement (n = 254)	Low engagement (n = 100)
Gender: $\chi^2(2) = 5.36$, $p = .068$, Cramér's $V = .103$			
Male	58 (38.7%)	128 (50.4%)	44 (44.0%)
Female	92 (61.3%)	126 (49.6%)	56 (56.0%)
Year of study: $\chi^2(8) = 11.06$, $p = .199$, Cramér's $V = .105$			
First year	41 (27.3%)	57 (22.4%)	17 (17.0%)
Second year	40 (26.7%)	67 (26.4%)	38 (38.0%)
Third year	37 (24.7%)	52 (20.5%)	19 (19.0%)
Fourth year	20 (13.3%)	53 (20.9%)	17 (17.0%)
Postgraduate	12 (8.0%)	25 (9.8%)	9 (9.0%)
Primary AI tool: $\chi^2(12) = 11.39$, $p = .496$, Cramér's $V = .106$			
ChatGPT	43 (28.7%)	61 (24.0%)	28 (28.0%)
Doubao	19 (12.7%)	38 (15.0%)	9 (9.0%)
ERNIE Bot	21 (14.0%)	37 (14.6%)	19 (19.0%)
Kimi	23 (15.3%)	29 (11.4%)	7 (7.0%)
Zhipu Qingyan	11 (7.3%)	24 (9.4%)	8 (8.0%)
DeepSeek	16 (10.7%)	22 (8.7%)	10 (10.0%)

Variable	High engagement (n = 150)	Moderate engagement (n = 254)	Low engagement (n = 100)
Other	17 (11.3%)	43 (16.9%)	19 (19.0%)
CET-4 score (n = 371): $F(2, 368) = 0.78$, $p = .458$			
***M* (*SD*)**	468.5 (69.1)	459.3 (59.2)	461.1 (59.5)

Note. Entries are *n* (% within profile) unless otherwise indicated. Other = Tongyi Qianwen, Claude, Tencent Yuanbao, and iFlytek Spark. CET-4 = College English Test Band 4.

Discussion

A Three-Dimensional Structure

The first research question concerned factor structure. Exploratory and confirmatory analyses in independent half-samples converged on three factors, and models that merged the dimensions fit far worse. Two aspects of this result are of theoretical interest.

The first is that Affective Connectedness in Collaboration was distinct from Positive Affective Engagement. The two correlated moderately (latent r = .46), and a model that merged them was clearly rejected. Enjoying work with GenAI and feeling connected to and understood by it are therefore related but different experiences. This extends the argument that relational feelings are part of engagement in language learning (Philp & Duchesne, 2016) to a non-human partner, and it is consistent with evidence that people respond socially to conversational systems (Nass & Moon, 2000; Skjuve et al., 2021). Connectedness was endorsed less strongly than positive affect (M = 3.60 vs. 3.93). A plausible reading is that most learners find GenAI collaboration pleasant and rewarding, while a smaller share experience it as relational. Instruments that assess only enjoyment or interest would miss this difference.

The second is that Emotional Disaffection was a separate factor and not the negative pole of a single continuum. Its correlation with Positive Affective Engagement was moderate and negative (latent r = −.52), close to the relation between enjoyment and anxiety reported for language classrooms (Dewaele & MacIntyre, 2014) and in line with the distinction between engagement and disaffection drawn by Skinner et al. (2009). Its correlation with Affective Connectedness was weak (latent r = −.24). Feeling supported by GenAI thus offers only limited protection against discouragement or frustration when collaboration fails to help. This is understandable from the standpoint of control-value theory (Pekrun, 2006): negative emotions in this setting are likely to arise from appraisals of low control, for instance when feedback is confusing or a difficulty remains unresolved, and such appraisals can occur whether or not the learner feels at ease with the system.

Neither a general factor nor a higher-order factor was supported. The bifactor solution did not yield an interpretable general factor, and the second-order loading of Positive Affective Engagement approached unity. A single total score would therefore be dominated by positive affect and would obscure the other two dimensions. Researchers using the scale should report the three subscale scores separately.

Fit of the three-factor model was acceptable, not excellent. The misfit was traced to four within-factor item pairs with related content: accomplishment and anticipation (Items 4 and 5), feeling less isolated and feeling comfortable asking questions (Items 9 and 10), stress and frustration (Items 14 and 15), and discouragement and disappointment (Items 11 and 12). Such local dependence is common when items within a subscale describe neighboring emotions, and strict application of cutoff values to multi-item scales has been criticized as unrealistic (Marsh et al., 2004). The pattern nevertheless identifies where the scale could be refined. Rewording one item in each pair, or replacing it with an item that samples a different emotion, may

reduce redundancy in future versions.

Reliability, Validity, and Invariance

The second and third research questions concerned the quality and comparability of scores. All three subscales showed good internal consistency with only five items each, exceeded the AVE threshold for convergent validity, and met both the Fornell-Larcker and HTMT criteria for discriminant validity. The scale is short enough to be included in multi-construct surveys or administered repeatedly.

Full invariance across gender and across five year-of-study groups means that men and women, and students at different stages of study, interpret the items in the same way and use the response scale similarly. Observed and latent mean differences between these groups can therefore be attributed to the construct and not to the instrument (Putnick & Bornstein, 2016). Evidence of this kind is seldom reported for instruments in this field, and it provides a foundation for group comparisons in later research.

Profiles That Differ in Level

The fourth research question concerned heterogeneity. Three profiles were identified. Their defining feature is that they were ordered: the high, moderate, and low profiles kept the same rank on positive affect and connectedness and the reverse rank on disaffection. In the terms of Morin and Marsh (2015), the profiles differed in level and not in shape. Learners who combined high enjoyment with high distress, or strong connectedness with little enjoyment, did not form a distinguishable group. One explanation is that behavioral, cognitive, and emotional engagement can diverge from one another, whereas the facets within emotional engagement tend to move together. When profiles differ mainly in level, a dimensional description of individual differences is more parsimonious than a typology (Marsh et al., 2009), and for most purposes the three subscale scores will be more informative than profile membership.

The content of the profiles is still informative. Half of the sample was moderately engaged, and nearly a third was highly engaged, which is consistent with the generally positive emotional responses reported in the literature (Zhang & Liu, 2025). The low profile, one fifth of the sample, was not characterized by strong negative emotion. Its members scored near the midpoint on all three subscales, including disaffection. These learners appear indifferent or undecided, not distressed. The profiles were separated chiefly by positive affect, and much less by connectedness or disaffection. The main difference among learners, in other words, was how much they enjoyed and valued the collaboration.

Profile membership was unrelated to gender, year of study, English proficiency, or the GenAI tool used. Emotional engagement with GenAI was thus spread evenly across the demographic and proficiency groups examined here, and no tool stood out as more or less emotionally engaging. This pattern agrees with evidence that engagement profiles in GenAI-mediated learning are associated with malleable features of the learning environment (Xu, 2026), and it suggests that explanations should be sought in how learners use GenAI and in the instructional context more than in who the learners are. The null findings should be interpreted cautiously, because the tool categories were broad and CET-4 scores were available for only 74% of the sample.

Conclusion

This study provides empirical evidence for a three-dimensional conceptualization of emotional engagement in human-GenAI collaborative language learning, comprising Positive Affective Engagement, Affective Connectedness in Collaboration, and Emotional Disaffection. The 15-item instrument demonstrated satisfactory reliability, convergent validity, and discriminant

validity, alongside configural, metric, scalar, and strict measurement invariance across gender and year of study. Notably, the identification of affective connectedness as a distinct dimension highlights the importance of perceived relational experiences in human-GenAI collaboration, extending conventional understandings of emotional engagement beyond enjoyment and negative emotional responses. Latent profile analysis further identified three groups characterized by high, moderate, and low emotional engagement, with differences primarily reflecting levels rather than configurations of affective experiences. These findings contribute to a multidimensional understanding of emotional engagement in technology-mediated language learning and provide a brief instrument for examining learners' affective responses to GenAI collaboration. From a pedagogical perspective, the findings highlight the value of attending to learners' enjoyment, perceived connectedness, and emotional difficulties when designing and supporting human-GenAI collaborative language learning activities.

Several limitations should be acknowledged. First, the study relied on cross-sectional self-reported data from 504 students at three Chinese universities, which limits conclusions about temporal stability and generalizability to other educational and cultural contexts. Although the reported analyses did not indicate that a single factor accounted for the observed item relationships, potential common method variance cannot be entirely excluded. Second, the psychometric evaluation primarily focused on internal structure, reliability, and measurement invariance, without incorporating external criteria to establish criterion-related or predictive validity. Test-retest reliability was also not examined, and several item pairs exhibited local dependence, indicating opportunities for further refinement. Third, the latent profile solution showed moderate classification precision, suggesting that individual profile assignments should be interpreted cautiously. Furthermore, although demographic characteristics, English proficiency, and GenAI tool use were examined in relation to profile membership, the study did not assess the frequency or specific patterns of GenAI use. These limitations constrain the extent to which the identified profiles can explain the contextual and interactional processes underlying learners' emotional engagement.

Future research should extend the present findings in several directions. First, cross-cultural and longitudinal studies involving learners from different educational levels and institutional contexts are needed to evaluate the generalizability, temporal stability, and measurement equivalence of the proposed instrument. Second, further validation could incorporate established measures of foreign language enjoyment and anxiety, behavioral and cognitive engagement, actual GenAI use, and language learning outcomes to strengthen criterion-related and predictive validity evidence. Refining items exhibiting local dependence, cross-validating the revised instrument in independent samples, and examining its properties using estimation methods appropriate for ordinal responses would also be valuable. Third, future studies could investigate how the frequency, purposes, and interactional patterns of GenAI use, together with instructional conditions, relate to different emotional engagement profiles. Longitudinal person-centered approaches, particularly latent transition analysis, may further clarify whether learners move between profiles over time and which contextual factors are associated with these transitions. Such investigations would contribute to a more comprehensive understanding of the dynamic and context-dependent nature of emotional engagement in human-GenAI collaborative language learning.

Acknowledgement

We gratefully acknowledge the anonymous reviewers for their diligent efforts in reviewing this study.

Declaration of Interest Statement

There is no potential conflict of interest.

References

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57,* Article 100745. https://doi.org/10.1016/j.asw.2023.100745

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science, 5*(1), 184–186. https://doi.org/10.1177/2167702616657069

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Carlo, M. (2026). Same CEFR, different prompt: Instruction-paraphrase sensitivity in CEFR-conditioned language model writing. *Applied Language Sciences, 2,* 260113. https://doi.org/10.65553/ALS.260113

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*(2), 195–212. https://doi.org/10.1007/BF01246098

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

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

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

Dewaele, J.-M., & Li, C. (2021). Teacher enthusiasm and students' social-behavioral learning engagement: The mediating role of student enjoyment and boredom in Chinese EFL classes. *Language Teaching Research, 25*(6), 922–945. https://doi.org/10.1177/13621688211014538

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*(3), 218–226. https://doi.org/10.1037/0003-066X.56.3.218

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior, 75,* 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Guo, N., & Li, R. (2022). Measuring Chinese English-as-a-foreign-language learners' resilience: Development and validation of the foreign language learning resilience scale. *Frontiers in Psychology, 13,* 1046340. https://doi.org/10.3389/fpsyg.2022.1046340

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

https://doi.org/10.1080/19312458.2020.1718629
Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8
Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2024). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research, 28*(1), 201–230. https://doi.org/10.1177/13621688211001289
Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447
Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118
Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning–Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237–257. https://doi.org/10.1111/jcal.12610
Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology, 54*(5), 1057–1076. https://doi.org/10.1111/bjet.13325
Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies, 28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1
Jin, S.-H. (2025). Measures of learner-generative ai relationships. *Computers and Education Open, 8,* 100258. https://doi.org/10.1016/j.caeo.2025.100258
Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103,* Article 102274. https://doi.org/10.1016/j.lindif.2023.102274
Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868
Lee, Y.-J., Parrila, R., & Colenbrander, D. (2026). Adopt, adapt, or reject? Analysing student-AI interaction in university curriculum-aligned writing tasks. *Applied Language Sciences, 2,* 260102. https://doi.org/10.65553/ALS.260102
Li, C., Jiang, G., & Dewaele, J.-M. (2018). Understanding Chinese high school students' foreign language enjoyment: Validation of the Chinese version of the Foreign Language Enjoyment scale. *System, 76,* 183–196. https://doi.org/10.1016/j.system.2018.06.004
Liu, G., & Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching, 18*(2), 125–138. https://doi.org/10.1080/17501229.2023.2240316
MacIntyre, P., & Gregersen, T. (2012). Emotions that facilitate language learning: The positive-broadening power of the imagination. *Studies in Second Language Learning and Teaching, 2*(2), 193–213. https://doi.org/10.14746/ssllt.2012.2.2.4
Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling, 16*(2), 191–225. https://doi.org/10.1080/10705510902751010

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4), 632–645. https://doi.org/10.1111/ejed.12527

Morin, A. J. S., & Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling, 22*(1), 39–59. https://doi.org/10.1080/10705511.2014.919825

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Oga-Baldwin, W. L. Q. (2019). Acting, thinking, feeling, making, collaborating: The engagement process in foreign language learning. *System, 86,* Article 102128. https://doi.org/10.1016/j.system.2019.102128

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics, 36,* 50–72. https://doi.org/10.1017/S0267190515000094

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41,* 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology, 36*(4), 257–267. https://doi.org/10.1016/j.cedpsych.2011.05.002

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist, 50*(1), 1–13. https://doi.org/10.1080/00461520.2014.1002924

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493–525. https://doi.org/10.1177/0013164408323233

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion–A study of human-chatbot relationships. *International Journal of Human-Computer Studies, 149,* Article 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior, 120,* Article 103445. https://doi.org/10.1016/j.jvb.2020.103445

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing

classrooms. *Assessing Writing, 57,* Article 100752. https://doi.org/10.1016/j.asw.2023.100752

Svalberg, A. M.-L. (2009). Engagement with language: Interrogating a construct. *Language Awareness, 18*(3–4), 242–258. https://doi.org/10.1080/09658410903197264

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition, 41*(2), 363–387. https://doi.org/10.1017/S0272263118000311

Teng, M. F. (2026). The impact of GenAI feedback on Chinese EFL students' emotional engagement: A mixed effect modelling approach. *The Asia-Pacific Education Researcher,* 1–17. https://doi.org/10.1007/s40299-025-01067-w

Teng, M. F., & Huang, J. (2025). Incorporating ChatGPT for EFL writing and its effects on writing engagement. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 15*(1), 1–21. https://doi.org/10.4018/IJCALLT.367874

Teng, M. F., & Huang, J. (2026). Assessing ChatGPT feedback for EFL learners' engagement and writing performance. *International Journal of Applied Linguistics.* https://doi.org/10.1111/ijal.70155

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education, 194,* Article 104703. https://doi.org/10.1016/j.compedu.2022.104703

Wang, Y., & Xue, L. (2024). Using AI-driven chatbots to foster Chinese EFL students' academic engagement: An intervention study. *Computers in Human Behavior, 159,* 108353. https://doi.org/10.1016/j.chb.2024.108353

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing, 62,* Article 101071. https://doi.org/10.1016/j.jslw.2023.101071

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

Watson Todd, R. (2026). Using generative AI to revise articles: How much is too much? *Applied Language Sciences, 2,* 260108. https://doi.org/10.65553/ALS.260108

Wong, W. L., & Fong, S. (2025). A new era with artificial intelligence powered learner corpora for teaching and learning of English as a second language in primary school education. *Applied Language Sciences, 1,* 250105. https://doi.org/10.65553/ALS.250105

Xu, X. (2026). Emotions and engagement in generative AI-supported L2 writing: A systematic review. *Emerging Educational Technologies, 1,* 260103. https://doi.org/10.65553/EET.260103

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*(11), 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

Zhang, S., & Liu, X. (2025). Learner emotions in AI-assisted English as a second/foreign language learning: A systematic review of empirical studies. *Frontiers in Psychology, 16,* Article 1652806. https://doi.org/10.3389/fpsyg.2025.1652806

Zou, M., & Huang, L. (2023). To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model. *Frontiers in Psychology, 14,* Article 1259531. https://doi.org/10.3389/fpsyg.2023.1259531

Author Bios

Songming Liu received his M.A. degree from Inner Mongolia University, China. He is currently working toward the Ph.D. degree in Applied Science of Linguistics with the Faculty of Languages and Translation, Macao Polytechnic University, Macao, China. His research interests lie at the intersection of psycholinguistics, cognitive science, and educational technology. Specifically, his current research focuses on artificial intelligence-assisted language learning among trilingual learners, with a special emphasis on metacognition and positive psychology.

E-mail: liusongming314@gmail.com

ORCiD: https://orcid.org/0009-0007-0580-797X

Yongliang Wang is an Associate Professor of Applied Linguistics and a Ph.D. degree holder. He works at the School of Foreign Studies, North China University of Water Resources and Electric Power, China, and serves as a postgraduate supervisor in Applied Linguistics at the School of Foreign Languages and Cultures, Nanjing Normal University, China. He also acts as a blind external examiner for Ph.D. dissertations in language education from institutions in the United Kingdom, Poland, Australia, the USA, and other countries. Additionally, he serves as a peer reviewer for over 40 SSCI-indexed international journals in EFL education. He has been recognized as one of the world's top 2% scientists in Languages and Linguistics by Stanford University for three consecutive years (2024–2026), and he was named a Highly Cited Chinese Researcher by Elsevier for four consecutive years (2022–2025). His recent academic publications have appeared in top-tier journals such as *Computers in Human Behavior, System, Journal of Multilingual and Multicultural Development, and Applied Linguistics Review.*

E-mail: wangyongliang@ncwu.edu.cn

ORCiD: http://orcid.org/0000-0002-4672-8481

Appendix

Items of the Emotional Engagement Measure

Item	Subscale	Wording
1	Positive Affective Engagement	I enjoy learning through Human–GenAI collaboration.
2	Positive Affective Engagement	I feel interested when exploring new foreign language expressions, linguistic knowledge, or related information together with AI.
3	Positive Affective Engagement	I feel excited when human–AI collaboration helps me make progress on challenging foreign language learning tasks.
4	Positive Affective Engagement	I feel a sense of accomplishment after completing foreign language learning tasks in collaboration with AI.
5	Positive Affective Engagement	I look forward to carrying out foreign language learning tasks through human–AI collaboration.
6	Affective Connectedness in Collaboration	During Human–GenAI collaborative learning, I feel a sense of connection with GenAI.
7	Affective Connectedness in Collaboration	I feel that the AI provides me with support.
8	Affective Connectedness in Collaboration	When the AI responds to my questions, I feel that my learning needs are understood.
9	Affective Connectedness in Collaboration	When dealing with challenging tasks, collaborating with AI makes me feel less isolated and helpless.
10	Affective Connectedness in Collaboration	I feel comfortable asking the AI questions.
11	Emotional Disaffection	During Human–GenAI collaborative foreign language learning, I feel discouraged when my progress is unsatisfactory.
12	Emotional Disaffection	I feel disappointed when the collaboration does not bring the expected improvement.
13	Emotional Disaffection	I feel anxious when I obtain unsatisfactory results.
14	Emotional Disaffection	I feel stressed when I am confused by AI-generated feedback.
15	Emotional Disaffection	I feel frustrated when the AI cannot adequately resolve my difficulties.