Applied Language Sciences

RESEARCH Open Access

Measuring Autonomous and Efficacious GenAI Use and GenAI Reliance in EFL Learning: Initial Scale Development and Validation

Olga Li[1]* & Atsushi Mizumoto[2]

*Correspondence:
liolga@pu-toyama.ac.jp

[1] Toyama Prefectural University, Japan

[2] Kansai University, Japan

Check for updates

Abstract

This study developed and examined initial validity evidence for a scale assessing autonomous and reliant use of generative artificial intelligence (GenAI) in English as a foreign language (EFL) learning. Participants were 416 students enrolled in required English courses at one Japanese university. The sample was randomly divided into an exploratory sample (n = 205) and an independent confirmatory sample (n = 211). Exploratory factor analysis of 10 initial items supported two factors: Autonomous and Efficacious GenAI Use (six items) and GenAI Reliance (three items). One of the ten items (REL4), which concerned worry about reliance, did not load meaningfully on either factor and was treated as an exploratory single item. The initial confirmatory factor analysis showed mixed fit. A post hoc model allowing a residual covariance between two autonomy items (AU1 and AU2) showed improved fit (robust CFI = .959, robust TLI = .940, robust RMSEA = .073, and SRMR = .060). Internal-consistency coefficients were .86 for Autonomous and Efficacious GenAI Use and .70 for GenAI Reliance, and the latent factor correlation was small and nonsignificant (r = .11, 95% CI [−.07, .30]). The heterotrait-monotrait ratio was .175, 95% CI [.132, .343]. These findings provide preliminary evidence that Autonomous and Efficacious GenAI Use and GenAI Reliance may represent distinct dimensions. Further validation with new, more diverse samples and external criteria is required.

Keywords: generative artificial intelligence; learner autonomy; GenAI reliance; self-efficacy; EFL; scale development

GloDAL Global Digital Applied Linguistics Association

© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

1. Introduction

Generative AI (GenAI) has rapidly become part of everyday and educational life. Its influence is particularly evident in foreign language education, where AI has become a major focus of research in second-language learning and applied linguistics; ChatGPT-related studies, in particular, have expanded rapidly since its public release (Kartal & Yeşilyurt, 2024; Li et al., 2024; Yang & Li, 2024). In university EFL classrooms, teachers increasingly encounter students who already have experience using GenAI tools in both private and educational settings and who form their own views about what these tools can and cannot do for their learning (Liu et al., 2024). This rapid uptake raises a question about whether autonomous and reliant GenAI use represent opposite ends of a single continuum or distinguishable dimensions of learners' engagement. The present study addresses this question by developing and initially validating a scale that measures both Autonomous and Efficacious GenAI Use and GenAI Reliance among Japanese university EFL learners.

2. Literature Review

Learner autonomy has long been regarded as an important educational goal because autonomous learners are expected to take an active role in planning, monitoring, and evaluating their own learning. Holec (1981) defined autonomy as "the ability to take charge of one's own learning" (p. 3), and Little (1991) described it as a capacity involving "critical reflection, decision-making, and independent action" (p. 4). Teng (2019) further linked autonomous learning to self-planning, self-management, self-reflection, and self-evaluation, and argued that helping learners increase their autonomy is part of teachers' responsibility. In GenAI-mediated learning, these dimensions may become even more important because learners must decide not only what and how to learn, but also when, why, and how to use AI support (Teng, 2026).

Autonomy does not mean learning without support. Holec (1981) explained that self-directed learning does not necessarily exclude the presence of teachers or teaching aids. Little (1991) also argued that autonomy is not the absence of a teacher; rather, it involves learners' critical reflection and responsibility for decision-making in their learning process. This distinction is crucial in GenAI-mediated learning: if autonomy is understood as informed decision-making rather than as unaided learning or performance, the use of GenAI does not necessarily indicate reliance (Reinders & White, 2016). Whether GenAI supports or erodes autonomy therefore depends less on how often learners use it than on how they use it.

Recent studies suggest that GenAI tools can support autonomous learning. They can reduce language-learning anxiety, provide real-time, individualized feedback, support grammar and vocabulary development, and help learners brainstorm ideas or prepare presentations. These affordances largely stem from GenAI's capacity to provide immediate, individualized support, which teachers may struggle to offer continuously in large or time-limited classroom settings. Koç and Savaş (2025) found that AI chatbots can provide learners with opportunities to practice speaking and may foster learner autonomy. Similarly, Zou et al. (2026) indicated that GenAI use may support affective and motivational dimensions of autonomous learning, while Guan et al. (2025) showed that GenAI tools can provide flexible access to feedback, explanations, and practice opportunities. Consistent with these affordances, a recent meta-analysis found that chatbots have a medium-sized positive effect on second-language learning (Lyu et al., 2025), and EFL learners have generally shown favorable acceptance of GenAI tools for language study (Ma, 2025).

However, such affordances are likely to promote autonomy only when learners can plan, monitor, evaluate, and revise their AI use (Anders & Speltz, 2025). In other words, learners need to use AI with, rather than instead of, their own judgment.

At the same time, researchers have raised concerns about cognitive and affective risks of GenAI use. For example, Goh et al. (2025) suggested that AI reliance may be associated with weaker critical thinking, while Maral et al. (2026) reported associations with problematic internet and AI use. Jose et al. (2025) described this tension as the "cognitive paradox of AI in education", arguing that although AI can support learning through accessibility and efficiency, overreliance may negatively affect critical thinking, problem-solving, creativity, and memory retention. Gerlich (2025) provided further empirical support for this concern, showing that frequent AI tool use was negatively associated with critical thinking and that cognitive offloading mediated this relationship. GenAI Reliance, in this case, is not simply frequent use; it may also be affective and cognitive, for example, feeling anxious when AI support is unavailable, or turning to AI before attempting to think independently. It remains unclear whether such reliance bolsters or weakens learners' confidence and independent ability.

Recent research on GenAI in language education has examined learners' perceptions of its benefits and limitations for academic communication (Liu et al., 2024), its real-time use for idea generation, language support, and writing process (Khanh-Quynh & Cong-Lem, 2026), the processing of GenAI feedback (Yan, 2024), learners' motivation (Teng, 2026; Zare et al., 2025), and learning outcomes (Hong & Shin, 2025). These studies offer valuable insights into what GenAI can provide and how learners perceive it. However, less attention has focused on how learners regulate their reliance on GenAI. Existing research has not sufficiently measured whether learners' AI use reflects autonomous decision-making, critical evaluation, or reliance. Autonomous and reliant patterns of use have rarely been measured together within a single instrument, yet doing so is essential for understanding learners' engagement with AI as more than a matter of frequency.

Existing measurement studies provide important foundations. Murase's MILLA (2015) was developed to measure learner autonomy in a Japanese EFL context and includes technical autonomy subcomponents such as planning, reflection, management, self-evaluation, and record-keeping. However, because it was designed before the spread of GenAI, it cannot fully capture AI-mediated forms of autonomy. This contextual gap is important because Japanese university EFL learners may have experienced relatively teacher-led instruction and may feel anxious about making mistakes in English. Therefore, a scale developed for this context needs to capture not only general autonomy-related skills but also how learners make decisions and maintain confidence when using new forms of support such as GenAI.

Liu et al. (2025) introduced a scale designed to measure the strategic use of AI in self-regulated learning among Chinese students, including strategies for evaluating AI output and refining prompts. Their instrument provides a detailed account of how learners regulate both AI-assisted task processing and AI application. However, it was designed for comprehensive strategy coverage and includes no dimension capturing learners' reliance on AI. More recently, Ayaz et al. (2026) developed an instrument designed to measure AI-enhanced self-regulated learning and demonstrated the importance of measuring how learners plan, monitor, and evaluate their learning with AI support. However, this scale is grounded primarily in self-regulated learning and does not specifically address GenAI-mediated learner autonomy and reliance in EFL learning.

3. Research Gaps and Research Questions

GenAI reliance should not be understood simply as frequent or habitual use. It may involve emotional and cognitive dimensions, such as feeling anxious without AI support, relying on AI before attempting to think independently, or using AI as a source of affective safety. At the same time, it remains unclear whether such emotional reliance supports learners' confidence or undermines their development of independent language-learning abilities.

A bipolar account of autonomous use and reliance implies that greater autonomy necessarily corresponds to less reliance. A two-dimensional account allows these tendencies to vary separately. This distinction affects how the two patterns of use should be measured and interpreted. Assessing both within the same instrument makes it possible to examine whether a single continuum adequately represents learners' responses.

Accordingly, the present study develops and initially validates a questionnaire measuring autonomous and efficacious GenAI use and GenAI reliance among Japanese university EFL learners. It addresses two research questions. First, what factor structure and internal-consistency reliability are supported by the responses? Second, does the evidence support autonomous and efficacious use and reliance as opposite poles of a continuum, or as empirically distinguishable dimensions?

4. Method

4.1 *Research design*

This study employed a quantitative scale development and validation design. The purpose was to develop and evaluate a scale measuring Autonomous and Efficacious GenAI Use and AI Reliance among EFL learners. The study involved item development based on prior literature and newly introduced items; administering a questionnaire to Japanese university EFL students; conducting exploratory factor analysis; model-based evaluation; reliability analysis; and examining the relationship between Autonomous and Efficacious GenAI Use and GenAI Reliance.

4.2 *Participants and context*

This study included 416 participants: 244 first-year, 99 second-year, and 73 graduate engineering students from the same university in Japan. They ranged in age from 18 to 23 years. According to the Common European Framework of Reference for Languages (CEFR), undergraduates were at the A1-A2 level and graduate students at the A2-B2 level as determined by a placement test. All students took required English courses during the spring semester of 2026. All participants have been studying English as a foreign language since the compulsory education stage, giving them more than six years of English-learning experience. Informal classroom inquiries suggested that many students had prior experience with GenAI. However, the study did not systematically measure the extent and frequency of their GenAI use. After receiving an explanation of the study, all participants provided written (paper-based) informed consent, and participation was voluntary and anonymous.

Graduate students were included because the study aimed to develop a scale applicable to university EFL learners rather than exclusively to undergraduates. Including both groups allowed the initial examination of the scale across different stages of university education. However, potential differences in learning purposes and GenAI experience between undergraduate and graduate students should be considered when interpreting the findings.

The sample of 416 participants was randomly split into two groups to perform EFA

and CFA analyses. The first sample group included 118 first-year, 48 second-year undergraduate, and 39 graduate students for a total of 205. This sample size is sufficiently large to conduct exploratory factor analysis (EFA) (Takeuchi & Mizumoto, 2023). The second sample group included 126 first-year students, 51 second-year students, and 34 graduate students, for a total of 211. The participants provide a relevant context for examining GenAI-mediated autonomy because many learners study English as a required subject and increasingly have access to AI-based language support.

4.3 *Scale development*

Scale development drew on relevant literature and aimed to capture three components: learner autonomy (AU), efficacious GenAI use (EF), and GenAI reliance (REL) in GenAI-mediated EFL learning. Learner autonomy (AU1-AU4) items were adapted from established autonomy instruments and validation procedures (Alamer et al., 2025; Murase, 2015; Nguyen & Habók, 2021), reframed for GenAI use, such as deciding when and why to use GenAI, attending to and acting on AI feedback, and being aware of one's strengths and weaknesses. AI self-efficacy items (EF1-EF2) were adapted from technology-acceptance work on perceived ease of use and effective tool use with ChatGPT (Alsumairi & Madini, 2025) and capture confidence in using GenAI and refining prompts. GenAI Reliance (REL1-REL4) items were informed by work on AI dependence and affective reliance/affective safety (Alsumairi & Madini, 2025; Goh et al., 2025; Liu et al., 2025), capturing anxiety without AI and reliance on AI before independent thinking.

The initial item pool consisted of 10 items: four addressing learner autonomy (AU1–AU4), two addressing efficacious GenAI use (EF1–EF2), and four addressing GenAI reliance (REL1–REL4). The items were developed based on the theoretical framework and relevant literature, with the aim of capturing key aspects of each construct using a concise set of items.

All items were originally written in Japanese and designed to address a single idea, with examples of GenAI tools (e.g., ChatGPT) included where appropriate. No reverse-keyed items were used. Two experienced English educators reviewed the items: a Japanese English instructor with a PhD who regularly uses English professionally, and an Australian native English-speaking instructor with an MA in Applied Linguistics and 30 years of English teaching experience in Japan. The reviewers examined the naturalness of the Japanese wording, item clarity, and relevance to the intended constructs.

The initial 10 items were retained for administration. The Japanese items were administered to participants and subsequently translated into English for reporting in this article.

4.4 *Measures*

The final scale consisted of nine items to measure two components: Autonomous and Efficacious GenAI Use (AU1-AU4 and EF1-EF2) and GenAI Reliance (REL1-REL3). REL4 was retained as a supplementary descriptive item reflecting Concern about Reliance and was not included in the final scale scores. The full item list can be found in the Appendix. A 6-point Likert scale was used (1 = does not apply to me at all… 6 = applies to me very much); an even-numbered scale was adopted to avoid a neutral midpoint and elicit a directional response. Scale scores were computed as item means without any reverse coding.

4.5 *Data collection*

The survey was administered during the spring semester of 2026 via online Google Forms. A total of 416 students participated in this study, including 244 first-year, 99 second-year, and 73 graduate engineering students from the same university in Japan. Of the 455 students invited during required English classes, 416 completed the survey.

The survey was approved by the research ethics committee of Toyama Prefectural University (approval No. R6-29).

4.6 *Data analysis*

All analyses were conducted in R (version 4.5.1). The 416 respondents were randomly divided into an exploratory sample (n = 205) and an independent confirmatory sample (n = 211) to explore the factor structure and then cross-validate it on separate data.

For the exploratory factor analysis (EFA), the suitability of the data for factoring was first assessed using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. Factors were then extracted using principal-axis factoring with promax (oblique) rotation, and the number of factors to retain was determined by considering parallel analysis, eigenvalues, and the interpretability of the resulting solutions.

The confirmatory factor analysis (CFA) was conducted on an independent validation sample using the robust maximum likelihood (MLR) estimator. Model fit was evaluated with the robust comparative fit index (CFI) and Tucker–Lewis index (TLI), for which values of .90 or above were considered acceptable and values of .95 or above good, together with the root mean square error of approximation (RMSEA ≤ .08) and the standardized root mean square residual (SRMR ≤ .08). One residual covariance was added on the basis of modification indices, as it was both statistically indicated and theoretically justifiable.

Dimensional distinctness was assessed by comparing the modified two-factor model with a one-factor model using a Satorra–Bentler scaled chi-square difference test. Both models included the same nine items and the AU1–AU2 residual covariance. The one-factor model allowed either sign for the freely estimated loadings. The 95% confidence interval for the standardized latent factor correlation was also examined to assess the magnitude and uncertainty of the association.

The heterotrait-monotrait ratio of correlations (HTMT) provided a second check of whether the two item sets measured different dimensions (Henseler et al., 2015). HTMT compares correlations between items assigned to different scales with correlations among items assigned to the same scale. Values closer to 1 indicate that the two item sets behave more alike; values below .85 support treating them as distinct. We calculated HTMT from absolute Pearson correlations, excluding REL4, and estimated percentile 95% confidence intervals from 10,000 respondent-level bootstrap samples. The CFA sample provided the primary test, and the EFA sample served as a sensitivity check.

For example, if an item about deciding when to use GenAI correlated as strongly with the reliance items as it did with the other autonomous-use items, the two scales would be difficult to distinguish. Lower correlations across the two item sets would support measuring the dimensions separately.

The internal-consistency reliability of the resulting scales was examined using Cronbach's alpha. Finally, descriptive statistics were computed for the scale scores, and Pearson correlations among them were estimated, with p-values adjusted for multiple comparisons using the Holm procedure.

Because a small overall correlation does not show whether individual learners can score high on both dimensions, we plotted each respondent's two scale means. A point in the upper-right part of Figure 2 represents a learner with elevated scores on both scales. We

counted respondents whose mean scores were at least 4 or 5 on both scales. On the six-point response scale, 4 corresponded to "Slightly applies to me," and 5 to "Applies to me." Exact binomial 95% confidence intervals described the uncertainty around these proportions. Jitter separated overlapping points in the figure but did not change the scores or counts.

These cutoffs describe how often the score combinations occurred in each sample; they do not define latent groups or show causality or stability over time. We repeated the analysis in the EFA sample as a sensitivity check.

To ensure transparency and reproducibility, the data and the R analysis code are openly available on OSF (https://osf.io/5n76p/overview).

5. Results

5.1 *Validity: Exploratory Factor Analysis*

An exploratory factor analysis (EFA) was conducted using the EFA sample (n = 205) to examine the underlying structure of the 10 initial items. The data were suitable for factor analysis, with an overall KMO/MSA value of .77. Bartlett's test of sphericity was significant, χ^2 (45) = 643.91, $p < .001$, indicating that the correlation matrix was factorable (Table 1).

Table 1. Factorability Checks for the Exploratory Factor Analysis

Index	Value
Kaiser–Meyer–Olkin (overall MSA)	.77
Bartlett's test of sphericity, χ^2(45)	643.91
p	< .001

Note. The EFA was conducted on the exploratory sample (n = 205) using the initial 10 items, including REL4. MSA = measure of sampling adequacy.

Principal-axis factoring with promax (oblique) rotation was used, and two- and three-factor solutions were compared. Parallel analysis supported two factors. The three-factor solution was difficult to interpret because AU2 largely defined the third factor, AU1 cross-loaded on the first and third factors, and the two autonomy-related factors correlated strongly (.68). The two-factor solution was therefore retained. AU1–AU4 and EF1–EF2 formed the first factor, Autonomous and Efficacious GenAI Use, whereas REL1–REL3 formed the second, GenAI Reliance (Table 2). The correlation between the two retained factors was .29. REL4 was initially included to capture learners' concern about their reliance on GenAI. However, it did not load meaningfully on either factor and had low communality (.08). Consequently, we excluded it from the final nine-item scale and retained only as a supplementary descriptive item. The section on CFA and dimensional distinctness reports the independent CFA results.

Table 2. Exploratory Factor Analysis: Two-Factor Promax Solution (n = 205)

Item	F1: Autonomous and Efficacious GenAI Use	F2: GenAI Reliance	h^2
AU1	.76	-.06	.56
AU2	.78	-.04	.59
AU3	.72	-.15	.47

Item	F1: Autonomous and Efficacious GenAI Use	F2: GenAI Reliance	h^2
AU4	.54	-.04	.28
EF1	.71	.04	.52
EF2	.58	.06	.35
REL1	.10	.55	.35
REL2	-.02	.88	.77
REL3	-.15	.71	.47
REL4	.23	.11	.08

Note. Extraction method: principal-axis factoring; rotation method: promax. All pattern coefficients are shown. h^2 = communality. The factor correlation was $r = .29$. REL4 was included in the initial EFA but did not load meaningfully on either factor and showed low communality; it was therefore excluded from the final scale and retained only for supplementary descriptive analysis.

5.2 *CFA and Dimensional Distinctness*

A confirmatory factor analysis (CFA) was conducted on an independent validation sample ($n = 211$) using the robust maximum likelihood (MLR) estimator. Based on the EFA, the model specified two correlated latent factors: Autonomous and Efficacious GenAI Use (AU1–AU4, EF1–EF2) and GenAI Reliance (REL1–REL3). REL4 was excluded from the CFA model and retained only as a supplementary descriptive item. The initial model showed mixed fit, robust CFI = .920, robust TLI = .889, robust RMSEA = .100, and SRMR = .067 (Table 3).

Table 3. Model Fit Indices for the One-Factor and Two-Factor CFA Models ($n = 211$)

Model	χ^2 (df)	p	CFI	TLI	RMSEA [90% CI]	SRMR
One-factor	150.50 (26)	< .001	.768	.678	.170 [.144, .197]	.126
Initial (two-factor)	70.10 (26)	< .001	.920	.889	.100 [.072, .128]	.067
Modified (two-factor)	47.61 (25)	.004	.959	.940	.073 [.040, .104]	.060

Note. CFA = confirmatory factor analysis; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Estimator = MLR. CFI, TLI, and RMSEA are robust indices, and χ^2 is the scaled test statistic. Both the one-factor and modified two-factor models included the AU1–AU2 residual covariance. The initial two-factor model included no residual covariance.

Modification indices indicated a residual covariance between AU1 and AU2. This respecification was theoretically justified because both items refer to intentional, self-regulated decisions about when and why to use generative AI in English learning. After adding this residual covariance, the modified model showed an improved fit, although some indices remained below conventional criteria for good fit: $\chi^2(25) = 47.61$, p = .004; robust CFI = .959, robust TLI = .940, robust RMSEA = .073, 90% CI [.040, .104], and SRMR = .060 (Table 3). Compared with the initial two-factor model, the modified model showed improvements in CFI (ΔCFI = .039), TLI (ΔTLI = .051), RMSEA (ΔRMSEA = −.027), and SRMR (ΔSRMR = −.007), as reported in Table 3.

Standardized factor loadings for the modified model are reported in Table 4 and depicted in Figure 1. All loadings were statistically significant ($p < .001$). Loadings for Autonomous and Efficacious GenAI Use ranged from .635 to .769, and loadings for GenAI

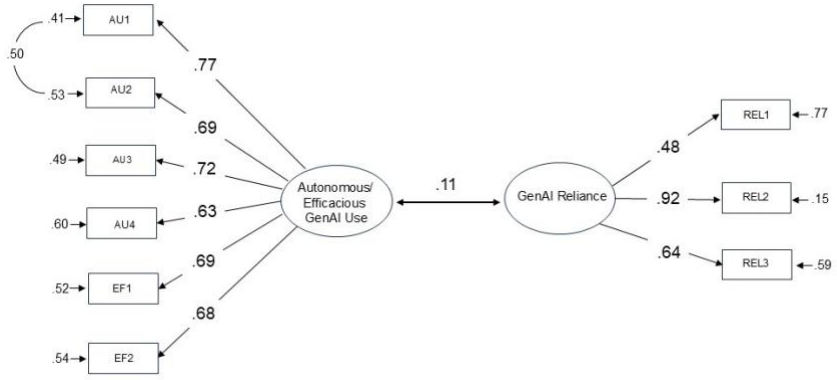
Reliance ranged from .483 to .922. The residual covariance between AU1 and AU2 was significant (standardized estimate = .496, $p < .001$). The one-factor model fit poorly (robust CFI = .768, robust RMSEA = .170) and showed worse fit than the modified two-factor model, scaled $\Delta\chi^2(1) = 70.01$, $p < .001$. The latent factor correlation was small, r = .11, 95% CI [−.07, .30].

Table 4. Standardized Factor Loadings for the Modified CFA Model (*n* = 211)

Factor / Item	Standardized loading
Autonomous and Efficacious GenAI Use	
AU1	.769
AU2	.688
AU3	.716
AU4	.635
EF1	.690
EF2	.679
GenAI Reliance	
REL1	.483
REL2	.922
REL3	.643

Note. All standardized factor loadings were statistically significant at $p < .001$.

In the CFA sample, HTMT was .175, 95% CI [.132, .343], well below the .85 criterion. The EFA sensitivity check gave a similar result (HTMT = .251, 95% CI [.161, .427]). In both samples, items within each scale were more strongly related to one another than to items in the other scale. This pattern supports measuring Autonomous and Efficacious GenAI Use and GenAI Reliance as separate dimensions.



Note. Standardized estimates from the modified two-factor CFA model. Rectangles denote observed items and circles denote latent factors. Single-headed arrows from the latent factors to the observed items represent standardized factor loadings; single-headed arrows entering the observed items from outside the model represent residual variances. The double-headed arrow linking the two latent factors is the latent factor correlation ($r = .11$, 95% CI [−.07, .30]), and the curved double-headed arrow on the left is the residual covariance between AU1 and AU2 (.50). All factor loadings were significant at $p < .001$.

Figure 1. Standardized Two-Factor Measurement Model (Modified CFA, *n* = 211)

5.3 *Reliability*

Internal consistency was examined for the final scales using the validation sample ($n = 211$). The Autonomous and Efficacious GenAI Use scale showed good internal

consistency, Cronbach's α = .86. The GenAI Reliance scale (REL1–REL3) showed acceptable internal consistency, α = .70. Adding REL4 reduced the internal consistency of the GenAI Reliance scale to α = .60, further supporting its exclusion from the final scale (Table 5).

Table 5. Internal-Consistency Reliability of the Scales (*n* = 211)

Scale	No. of items	Cronbach's α
Autonomous and Efficacious GenAI Use	6	.86
GenAI Reliance (REL1–REL3)	3	.70
GenAI Reliance including REL4	4	.60

Note. REL4 was excluded from the final scale and remained only as a supplementary descriptive item.

5.4 *Descriptive Statistics and Correlations*

Item-level descriptive statistics for the final nine items, including means, standard deviations, skewness, response distributions, and potential floor and ceiling effects, are provided in Supplementary Table S1.

Descriptive statistics for the final scale scores are presented in Table 6. Autonomous and Efficacious GenAI Use was relatively high (M = 4.45, SD = 0.99), whereas GenAI Reliance was moderate (M = 3.06, SD = 1.10). The supplementary REL4 item, reflecting concern about reliance, had a relatively high mean score (M = 4.43, SD = 1.29).

Table 6. Descriptive Statistics for Scale Scores and the Supplementary REL4 item (*n* = 211)

Scale	M	SD
Autonomous and Efficacious GenAI Use	4.45	0.99
GenAI Reliance	3.06	1.10
Concern about Reliance (REL4)	4.43	1.29

Note. Scale scores are means of the constituent items, each rated on a 6-point scale. REL4, a supplementary single item excluded from the final nine-item scale and is reported for descriptive purposes only.

Correlations among the scale scores and the supplementary REL4 item are shown in Table 7. Autonomous and Efficacious GenAI Use showed a small, nonsignificant positive association with GenAI Reliance, r = .12, adjusted p = .19. Evidence concerning dimensional distinctness is reported in the section on CFA and dimensional distinctness.

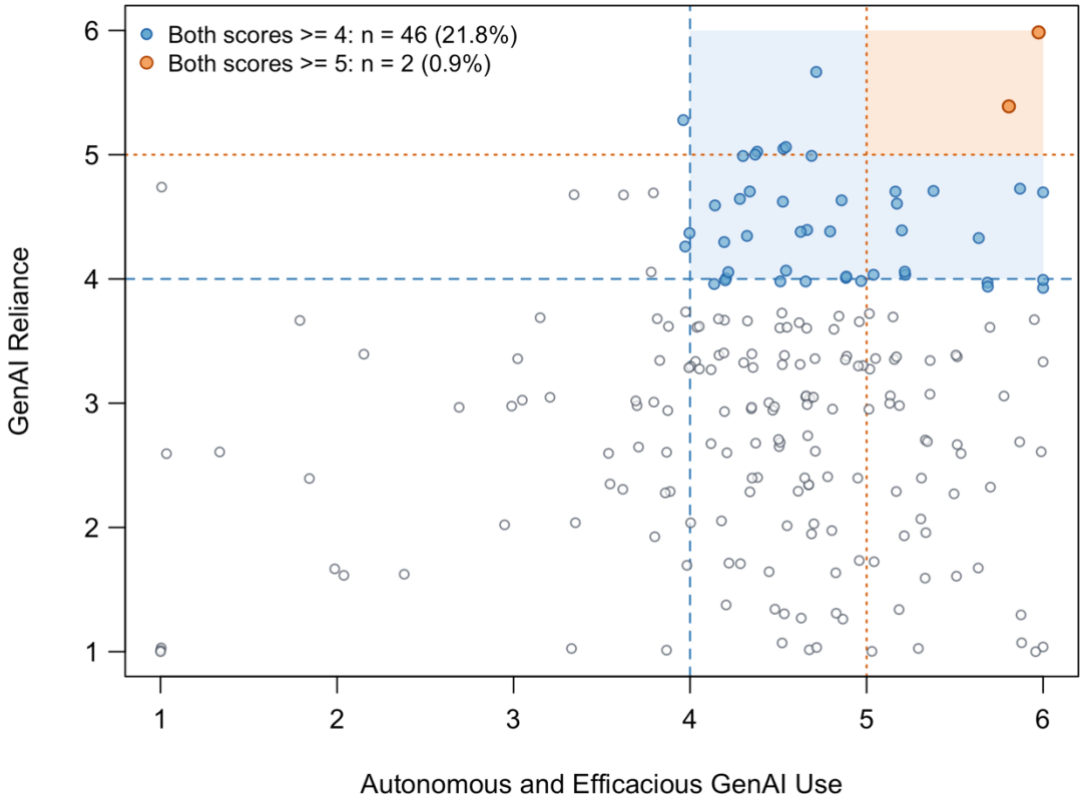
Table 7. Correlations Among Scale Scores and the Supplementary REL4 item (*n* = 211)

Scale	1	2	3
1. Autonomous and Efficacious GenAI Use	—		
2. GenAI Reliance	.12	—	
3. Concern about Reliance (REL4)	.10	.13	—

Note. Pearson correlations computed on the validation sample (n = 211). p-values were adjusted for multiple comparisons using the Holm procedure; none of the correlations reached statistical significance

(all adjusted $p > .05$). Correlations involving REL4 were examined for exploratory purposes only, as REL4 was not included in the final nine-item scale.

Figure 2 plots the two scores for each participant. Forty-six of 211 participants (21.8%, exact 95% CI [16.4%, 28.0%]) scored at least 4 on both scales. In practical terms, about one in five participants indicated that both tendencies applied to them at least slightly. For example, this pattern could describe a learner who decides when to use GenAI but still turns to it before thinking through an assignment independently. Only two participants (0.9%, exact 95% CI [0.1%, 3.4%]) scored at least 5 on both scales, so strong endorsement of both was rare. The EFA sample showed a similar percentage at the 4-point cutoff (45 of 205 participants, 22.0%). These descriptive cutoffs show that the combination occurred in both samples; they do not identify a stable learner type.



Note. Each point represents one respondent. Dashed lines mark scores of 4, and dotted lines mark scores of 5. The shaded upper-right regions indicate simultaneous scores at or above each threshold.

Figure 2. Joint Distribution of Autonomous and Efficacious GenAI Use and GenAI Reliance (*n* = 211)

6. Discussion

This study developed and initially validated a scale measuring Autonomous and Efficacious GenAI Use and GenAI Reliance in EFL learning. The results supported a nine-item, two-factor structure with acceptable fit after a post hoc residual covariance was added. Internal-consistency coefficients were .86 for Autonomous and Efficacious GenAI Use and .70 for GenAI Reliance.

6.1 *RQ1: Factor Structure and Internal Consistency*

In this sample, the autonomy and AI self-efficacy items were not empirically separable. This pattern may reflect conceptual overlap, limited construct coverage, or the small number of AI self-efficacy items and therefore requires replication. Rather than indicating that AI self-efficacy is necessarily a component of learner autonomy, the present results suggest that Autonomous and Efficacious GenAI Use were closely associated in this sample. This aligns closely with technology-acceptance research, in which learners' confidence in using AI and their perceptions of its ease of use underpin their autonomous engagement with the tool (Alsumairi & Madini, 2025).

REL4 ("I worry that relying on generative AI in English learning may interfere with my ability to think for myself") did not load on either factor and was therefore excluded from the final scale. Its relatively high mean score (M=4.43), compared with the moderate mean GenAI Reliance score, whereas average GenAI Reliance was moderate (M=3.06), which may suggest that learners can express concern about reliance without necessarily reporting high levels of reliance themselves. However, this interpretation is exploratory, and REL4 should not be regarded as a validated measure of a distinct construct. Further research using multiple items is needed to examine concern about GenAI reliance.

6.2 *RQ2: Dimensional Distinctness and Co-occurrence*

The findings provide initial evidence supporting two empirically distinguishable dimensions: Autonomous and Efficacious GenAI Use, and GenAI Reliance. The weak association between these dimensions suggests that they should not automatically be treated as opposite ends of a single scale. Across the sample, a higher score on one dimension did not consistently correspond to a higher or lower score on the other, and individual learners could score relatively high on both. However, these findings do not definitively rule out a continuum model, which would require more direct testing using opposing items addressing comparable behaviors.

Some learners may use GenAI purposefully and still rely on it. This combination matters for instruction because reliance may require attention even among learners who report confident and self-directed use. This interpretation is consistent with definitions of autonomy as control over learning rather than learning without support (Holec, 1981; Little, 1991; Reinders & White, 2016). It also fits accounts of AI as helping learners while creating risks of dependence and cognitive offloading (Jose et al., 2025).

Mean scores alone could not show whether the same learners reported both tendencies, so the joint-score analysis examined individual response patterns. About one in five learners met the 4-point cutoff on both scales, whereas scores of 5 or higher on both scales were rare. Because these proportions depend on the chosen cutoffs, they should not be treated as evidence of a stable learner type.

6.3 *Limitations and Implications*

The present study makes several initial contributions. First, it provides a brief scale to measure Autonomous and Efficacious GenAI Use and GenAI Reliance jointly within a single instrument, whereas existing tools have measured learner autonomy, self-regulated learning, or technology acceptance separately (Ayaz et al., 2026; Liu et al., 2025; Murase, 2015). Second, it offers a contextual contribution, as it was developed for and validated with Japanese university EFL learners, a population that has often experienced relatively

teacher-led instruction. Third, in line with open-science practice, the data and reproducible R code are openly available, allowing other researchers to scrutinize, reuse, and extend the instrument.

The findings provide preliminary evidence that the scale may be useful for examining individual differences in Autonomous and Efficacious GenAI Use and GenAI Reliance. Future research could examine the two dimensions in relation to learning outcomes, motivation, self-regulated learning, and behavioral indicators of actual GenAI use. However, further validation is required before the scale can be used for practical identification or instructional decision-making.

This study also has several limitations that point to directions for future research. First, it was conducted at a single institution within a specific EFL context, which limits the generalizability of the findings. Furthermore, including both undergraduate and graduate students may have introduced heterogeneity in learning purposes and GenAI experience. Future work should cross-validate the scale with larger, more diverse samples and test measurement invariance across groups such as year level, faculty, and proficiency.

Second, the modified CFA model included a residual covariance between AU1 and AU2, added post hoc based on modification indices and theoretical overlap in item content. Although this respecification improved model fit, it should be replicated in an independent sample before being treated as a stable feature of the measurement model.

Third, the constructs were assessed solely through self-report at a single time point, so future studies could examine test-retest reliability and gather convergent and criterion-related validity evidence, which can be done, for example, by relating the scale to measures of self-regulated learning, motivation, and proficiency, or to behavioral indicators such as actual AI-use logs.

Fourth, participants' prior experience with GenAI was not systematically measured. Some items (e.g., AU3 and EF2) may have been difficult for inexperienced users to evaluate based on actual experience. Consequently, we could not examine whether prior GenAI experience influenced the factor structure, correlations, or joint-score patterns. Future studies should assess participants' GenAI experience and examine the robustness of the scale across different levels of experience.

Finally, the relatively small initial item pool, particularly the two items representing efficacious GenAI use, may have limited construct coverage and the ability to distinguish autonomy from self-efficacy. GenAI Reliance was measured with only three items, while Concern about Reliance (REL4) was retained as a supplementary single item. Further research should develop a broader item pool and conduct additional content validation to examine these constructs more comprehensively.

7. Conclusion

This study developed and initially validated a brief nine-item scale for assessing Autonomous and Efficacious GenAI Use and GenAI Reliance among Japanese university EFL learners. The modified two-factor model fit the data better than the one-factor model, and the latent factor correlation (r = .11) and HTMT (.175) were low. In the CFA sample, 21.8% of participants scored at least 4 on both scales, indicating that autonomous and efficacious use can coexist with reliance within the same learner. These results support scoring the dimensions separately rather than placing learners on a single continuum from autonomous use to reliance. The scale provides a basis for further research on how these patterns relate to learning processes and outcomes. Replication with more diverse samples and external criteria is needed before the scale is used for individual assessment or instructional decisions.

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest
The authors declare no conflict of interest.

AI use disclosure
Generative AI tools were used to assist with language editing and refinement of English expression during manuscript preparation. The authors reviewed and verified all AI-assisted revisions and take full responsibility for the content of the manuscript.

References

Alamer, A., Sonbul, S., & El-Dakhs, D. A. S. (2025). Revisiting the validity of the vocabulary learning strategies questionnaire using the confirmatory composite analysis (CCA): Setting new directions for the field. *International Journal of Applied Linguistics, 35*(1), 193–217. https://doi.org/10.1111/ijal.12609

Alsumairi, A., & Madini, A. A. (2025). ChatGPT in EFL learning: Technology acceptance and learner autonomy among Saudi university students. *International Journal of English Linguistics, 15*(6), 73–85. https://doi.org/10.5539/ijel.v15n6p73

Anders, A. D., & Speltz, E. D. (2025). Developing generative AI literacies through self-regulated learning: A human-centered approach. *Computers and Education: Artificial Intelligence, 9*, Article 100482. https://doi.org/10.1016/j.caeai.2025.100482

Ayaz, M. A., Karataş, F., & Yüce, E. (2026). Development and validation of the AI-enhanced Self-Regulated Learning (AI-SRL) scale. *RIED-Revista Iberoamericana de Educación a Distancia, 29*(1), 185–216. https://doi.org/10.5944/ried.45452

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), Article 6. https://doi.org/10.3390/soc15010006

Goh, A. Y. H., Hartanto, A., & Majeed, N. M. (2025). Generative artificial intelligence dependency: Scale development, validation, and its motivational, behavioral, and psychological correlates. *Computers in Human Behavior Reports, 20*, Article 100845. https://doi.org/10.1016/j.chbr.2025.100845

Guan, L., Zhang, E. Y., & Gu, M. M. (2025). Examining generative AI–mediated informal digital learning of English practices with social cognitive theory: A mixed-methods study. *ReCALL, 37*(3), 315–331. https://doi.org/10.1017/S0958344024000259

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

Holec, H. (1981). *Autonomy and foreign language learning*. Pergamon Press.

Hong, S., & Shin, Y. K. (2025). Effects of three levels of AI integration on second language academic writing: Evaluating restricted, guided, and free use of ChatGPT. *System, 134*, Article 103820. https://doi.org/10.1016/j.system.2025.103820

Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., S, M., & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. *Frontiers in Psychology, 16*, Article 1550621. https://doi.org/10.3389/fpsyg.2025.1550621

Kartal, G., & Yeşilyurt, Y. E. (2024). A bibliometric analysis of artificial intelligence in L2 teaching and applied linguistics between 1995 and 2022. *ReCALL, 36*(3), 359–375. https://doi.org/10.1017/S0958344024000077

Khanh-Quynh, T. T., & Cong-Lem, N. (2026). The new writing partner? EFL students' real-time use of GenAI tools in academic composing. *Applied Language Sciences, 2,* Article 260107. https://doi.org/10.65553/ALS.260107

Koç, F. Ş., & Savaş, P. (2025). The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL, 37*(1), 4–21. https://doi.org/10.1017/S0958344024000168

Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence, 7*, Article 100266. https://doi.org/10.1016/j.caeai.2024.100266

Little, D. (1991). *Learner autonomy 1: Definitions, issues and problems*. Authentik.

Liu, X., Xiao, Y., & Li, D. (2025). Assessing strategic use of artificial intelligence in self-regulated learning: Instrument development and evidence from Chinese university students. *International Journal of Educational Technology in Higher Education, 22*(1), Article 69. https://doi.org/10.1186/s41239-025-00567-5

Liu, Y., Park, J., & McMinn, S. (2024). Using generative artificial intelligence/ChatGPT for academic communication: Students' perspectives. *International Journal of Applied Linguistics, 34*(3), 1437–1461. https://doi.org/10.1111/ijal.12574

Lyu, B., Lai, C., & Guo, K. (2025). Effectiveness of chatbots in improving language learning: A meta-analysis of comparative studies. *International Journal of Applied Linguistics, 35*(2), 834–851. https://doi.org/10.1111/ijal.12668

Ma, M. (2025). Exploring the acceptance of generative artificial intelligence for language learning among EFL postgraduate students: An extended TAM approach. *International Journal of Applied Linguistics, 35*(1), 91–108. https://doi.org/10.1111/ijal.12603

Maral, S., Naycı, N., Bilmez, H., Erdemir, E. İ., & Satici, S. A. (2026). Problematic ChatGPT Use Scale: AI-human collaboration or unraveling the dark side of ChatGPT. *International Journal of Mental Health and Addiction, 24*(3), 2369–2395. https://doi.org/10.1007/s11469-025-01509-y

Murase, F. (2015). Measuring language learner autonomy: Problems and possibilities. In C. J. Everhard & L. Murphy (Eds.), *Assessment and autonomy in language learning* (pp. 35–63). Palgrave Macmillan.

Nguyen, S. V., & Habók, A. (2021). Designing and validating the learner autonomy perception questionnaire. *Heliyon, 7*(4), Article e06831. https://doi.org/10.1016/j.heliyon.2021.e06831

Reinders, H., & White, C. (2016). 20 years of autonomy and technology: How far have we come and where to next? *Language Learning & Technology, 20*(2), 143–154. https://doi.org/10.64152/10125/44466

Takeuchi, O., & Mizumoto, A. (Eds.). (2023). *Gaikokugo kyoiku kenkyu handobukku: Kenkyu shuho no yori yoi rikai no tame ni* [Handbook of foreign language education research: For a better understanding of research methods] (Expanded ed.). Shohakusha.

Teng, M. F. (2019). *Autonomy, agency, and identity in teaching and learning English as a foreign language.* Springer.

Teng, M. F. (2025). Examining longitudinal development of writing motivation in the GenAI context: A self-determination theory perspective. *Learning and Motivation, 91,* 102157. Https://doi.org/10.1016/j.lmot.2025.102157

Teng, M. F. (2026). Metacognition in using GenAI for language teaching and learning. *Emerging Educational Technologies, 1,* 260101 https://doi.org/10.65553/EET.260101

Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology, 28*(1), 1–19. https://doi.org/10.64152/10125/73597

Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System, 124*, Article 103351. https://doi.org/10.1016/j.system.2024.103351

Zare, J., Al-Issa, A., & Ranjbaran Madiseh, F. (2025). Interacting with ChatGPT in essay writing: A study of L2 learners' task motivation. *ReCALL, 37*(3), 385–402. https://doi.org/10.1017/S0958344025000035

Zou, M., Reinders, H., & Amjad, F. (2026). Understanding the potential role of GenAI-mediated informal digital learning of English (GenAI-IDLE) in the Global South: AI literacy, emotions, and willingness to communicate as outcomes. *ReCALL, 38*(1), 131–149. https://doi.org/10.1017/S0958344025100360

Author biodata

Olga Li is an English instructor at Toyama Prefectural University, Japan, and a PhD student in Foreign Language Education. She holds an MA in TESOL and has extensive experience in English education, including teaching and managing a private English language school, working in elementary school English education, and serving as the director of an English-medium preschool. Her research interests include generative AI in language education, learner and teacher autonomy, AI-mediated writing, and responsible and autonomous GenAI use in EFL contexts.

Email: liolga@pu-toyama.ac.jp
ORCID: 0009-0006-0830-0667

Atsushi Mizumoto, Ph.D. Professor, Faculty of Foreign Language Studies/ Graduate School of Foreign Language Education and Research, Kansai University. His research interests include the educational use of AI and corpora, vocabulary learning strategies, and language testing. Recently, he has focused on learning and instruction using generative AI. He has published widely in international journals such as Applied Linguistics, Language Teaching Research, Language Testing, Language Learning, and System. He has received numerous awards, including the Academic Award from Language Education and Technology (2017) and the Academic Award from Japan Association of English Corpus Studies (2025). In 2025, he was listed in the “World’s Top 2% Scientists” by Stanford University and Elsevier.

Email: mizumoto@kansai-u.ac.jp
ORCID: 0000-0001-6588-4052

Appendix

Questionnaire Items (Japanese with English Translations)

	Statements	Does not apply to me at all 1	Does not apply to me 2	Slightly does not apply to me 3	Slightly applies to me 4	Applies to me 5	Applies to me very much 6
AU1	I can decide for myself when to use generative AI, such as ChatGPT, in my English learning. 英語学習で生成AI(ChatGPTなど)使用するタイミングを自分で判断している。						
AU2	I use generative AI, such as ChatGPT, for specific purposes in my English learning. 英語学習における生成AIは、特定の目的のもと使用している。						
AU3	When generative AI points out mistakes in my English, I pay attention to them and try to use the feedback to improve. 生成AIに間違いが指摘されたとき、それに注意し、自分の上達につなげようとする。						
AU4	When using generative AI for English learning, I am aware of my own strengths and weaknesses. 英語学習の際に生成AIを使用するとき、自分の弱点と強みが分かっている。						

	Statements	Does not apply to me at all 1	Does not apply to me 2	Slightly does not apply to me 3	Slightly applies to me 4	Applies to me 5	Applies to me very much 6
EF1	I can use generative AI effectively in my English learning. 英語学習のために生成AIを使いこなすことができる。						
EF2	When generative AI produces an output that is different from what I expected, I can revise my prompts. 生成AIが期待していたと異なったものを出力したとき、指示の出し方（プロンプト）を変えられる。						
REL1	I feel anxious when I have to complete English assignments, such as writing or presentation preparation, without using generative AI. 生成AIを使わずに英語の課題（作文、スピーキング・発表の準備等）に取り組む必要があるとき、不安を感じる。						
REL2	I often rely on generative AI to generate ideas rather than thinking of ideas myself. 自分で考えるより、生成AIにアイデアを出してもらうことが多い。						

	Statements	Does not apply to me at all 1	Does not apply to me 2	Slightly does not apply to me 3	Slightly applies to me 4	Applies to me 5	Applies to me very much 6
REL3	When I receive an English assignment, I tend to use generative AI first before thinking about it myself. 英語の課題が出ると、まず自分で考える前にとりあえず生成AIを使う。						
REL4	I worry that relying on generative AI in English learning may interfere with my ability to think for myself. 英語学習において生成ＡＩに頼ると、自分で考える力が妨げられるのではないかと心配している。						

Note. The original Japanese items were administered to participants, and the English translations are provided for reference. AU1–AU4 and EF1–EF2 constitute the Autonomous and Efficacious GenAI Use factor, while REL1–REL3 constitute the GenAI Reliance factor. REL4 was excluded from the final nine-item scale and retained only for supplementary exploratory analysis.

Supplementary Table S1. Item-Level Descriptive Statistics and Response Distributions for the Final Nine-Item Scale (*n* = 211)

Item	M	SD	Skewness	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)
AU1	4.80	1.24	-1.41	4.7	0.9	5.2	19.9	37.4	31.8
AU2	4.76	1.36	-1.39	6.2	2.4	4.7	17.1	35.5	34.1
AU3	4.82	1.22	-1.38	3.8	1.9	6.6	16.1	39.8	31.8
AU4	3.76	1.39	-0.10	6.2	11.4	26.5	24.6	18.5	12.8
EF1	3.88	1.32	-0.27	6.6	5.7	22.3	37.4	14.2	13.7
EF2	4.68	1.30	-1.19	4.3	3.8	6.6	19.4	36.5	29.4
REL1	3.72	1.53	-0.28	11.4	13.3	12.8	31.3	17.5	13.7
REL2	3.06	1.39	0.25	16.6	17.1	32.2	17.5	11.4	5.2
REL3	2.41	1.24	0.75	26.5	34.1	19.4	12.8	5.7	1.4

Note. M = mean; SD = standard deviation. Response distributions indicate the percentage selecting each option on the six-point Likert scale (1–6). The percentage at 1 represents the minimum-response (floor) proportion; the percentage at 6 represents the maximum-response (ceiling) proportion. Skewness is the adjusted sample skewness. The table includes the nine items retained for CFA; REL4 was excluded after EFA. Percentages may not sum to 100 due to rounding.