ijonse International Journal on Studies in Education
www.ijonse.net

Efficacy Perception Scale for the Use of Artificial Intelligence in Foreign Language Teaching: Validity and Reliability Study

Fatih Mehmet Cigerci [1] *, Fatih Veyis [2], Fatmanur Cimen [3]

[1] Harran University, Şanlıurfa, Türkiye, 0000-0002-4175-7048
[2] Atatürk University, Erzurum, Türkiye, 0000-0002-4874-7643
[3] Atatürk University, Erzurum, Türkiye, 0000-0001-6068-430X
* Corresponding author: Fatih Mehmet Cigerci (fatihcigerci@harran.edu.tr)

Article Info

Article History

Received:
14 September 2025

Revised:
23 November 2025

Accepted:
12 December 2025

Published:
1 January 2026

Keywords

Foreign language teaching
Artificial intelligence
Scale development

Abstract

This study aims to develop a valid and reliable scale to measure teachers' self-efficacy perception towards the use of artificial intelligence (AI) in foreign language teaching. In the study, the scale development process started with a literature review and a draft of 39 items was created in line with expert opinions. As a result of pilot application and validity and reliability analyses, a final scale with 18 items was obtained. Exploratory and confirmatory factor analyses showed that the scale had a two-dimensional structure (Planning and Instruction and Measurement and Evaluation). According to the EFA results, the scale explained 76.75% of the total variance. Factor loadings ranged between .585 and 1.007 and item-total correlations were between .639 and .879, indicating that the scale items had sufficient discrimination. As a result of all these procedures, a valid and reliable scale was developed to measure the perception of competence in using artificial intelligence in foreign language teaching.

Citation: Cigerci, F. M., Veyis, F., & Cimen, F. (2026). Efficacy Perception Scale for the Use of Artificial Intelligence in Foreign Language Teaching: Validity and reliability study. *International Journal on Studies in Education (IJonSE), 8*(1), 108-125. https://doi.org/10.46328/ijonse.5845

ISSN: 2690-7909 /

Introduction

Foreign language education plays a vital role in fostering communication skills and intercultural understanding, and recent developments in artificial intelligence (AI) have expanded the pedagogical possibilities in this field. While traditional instructional approaches have long supported language acquisition, AI has introduced new opportunities for personalized learning, real-time assessment, and data-driven instructional support .particularly within skill areas such as pronunciation, vocabulary development, and writing accuracy (Luckin et al., 2016). Rather than functioning solely as supplementary tools, AI-driven applications are increasingly reshaping instructional design by offering adaptive and automated learning environments (Holmes et al., 2019).

AI-based technologies—such as natural language processing (NLP), machine learning, and deep learning—now support core elements of language learning, including speech recognition, grammar correction, vocabulary development, and individualized content recommendation (Zawacki-Richter et al., 2019). Intelligent tutoring systems, automated feedback tools, and AI-enhanced chatbots allow learners to engage in interactive, self-regulated activities while enabling teachers to monitor progress and tailor support more efficiently (Lin et al., 2023).

However, the integration of AI into foreign language classrooms is shaped not only by technological affordances but also by pedagogical and psychological factors. Teachers' perceptions of their competence in using AI tools—grounded in the broader construct of self-efficacy (Bandura, 1997)—play a crucial role in determining whether these tools are meaningfully incorporated into instruction (Ertmer & Ottenbreit-Leftwich, 2010). Although research increasingly documents the potential benefits of AI for learners, far fewer studies examine teachers' readiness, confidence, and perceived ability to implement AI-supported practices (Scherer et al., 2019; Tondeur et al., 2016). This gap limits our understanding of the human and pedagogical factors that shape AI adoption.

To address this need, the present study develops and validates a scale measuring teachers' self-efficacy in integrating AI technologies into foreign language instruction. A reliable and valid instrument can provide insights into the factors that influence teachers' adoption of AI, support the development of targeted professional training programs, and contribute to theoretical models of technology acceptance in education. It also has practical value for policymakers seeking to promote AI-supported language instruction through appropriate pedagogical and institutional support structures. Despite this growing body of work, empirical tools that measure teachers' perceived competence in applying AI within foreign language pedagogy remain scarce, limiting both theoretical advancement and practical implementation.

Literature Review

The Role of Technology in Foreign Language Teaching

Technological innovations such as computer-assisted language learning (CALL), mobile-assisted language learning (MALL), online learning platforms, and AI-powered applications have substantially transformed language teaching (Chapelle & Jamieson, 2008). These tools enhance accessibility, personalize learning paths, and increase engagement through interactive digital experiences (Godwin-Jones, 2011). Mobile applications and online platforms extend learning beyond the classroom and enable continuous practice (Kukulska-Hulme & Shield, 2008), while AI-based tools provide real-time feedback through automated speech recognition (ASR) and natural language processing (Derakhshan & Hasanabbasi, 2015).

CALL and MALL research consistently shows positive effects on grammar, vocabulary, pronunciation, and reading comprehension (Heift & Schulze, 2007; Blake, 2013). Digital environments also promote self-paced learning and accommodate diverse learning preferences. Furthermore, technology fosters exposure to authentic language use through virtual exchanges and video-mediated communication, which significantly enhance communicative competence (Golonka et al., 2014).

Despite the documented benefits, challenges remain. Digital inequalities restrict access for some learners, and limited digital literacy among teachers reduces the pedagogical impact of technology (Selwyn, 2021; Ertmer & Ottenbreit-Leftwich, 2010). Overreliance on digital tools may also diminish opportunities for spontaneous interaction, essential for developing communicative competence (Godwin-Jones, 2011). Finally, privacy and data security concerns highlight the need for ethical, informed technology adoption in language classrooms (Sack & Röcker, 2013). These concerns also imply that teachers must possess not only technical but also ethical and evaluative competencies when integrating AI tools.

These developments show that the pedagogical impact of AI in foreign language education depends not only on technological features but also on teachers' ability to interpret, implement, and evaluate AI-supported instructional practices. However, empirical tools that assess teachers' AI-related self-efficacy—particularly within foreign language pedagogy—remain limited, creating a need for valid and reliable measurement instruments.

Artificial Intelligence in Foreign Language Teaching

AI-based tools—ranging from ASR systems to intelligent tutoring technologies—have introduced new possibilities for personalized and interactive language learning (Zawacki-Richter et al., 2019). ASR provides detailed pronunciation feedback (Jiang et al., 2021), chatbots encourage low-pressure conversational practice (Zhang, 2025), and intelligent tutoring systems adapt content based on learner performance data (Paladines & Ramirez, 2020). Automated grammar correction and essay-scoring tools offer immediate feedback on writing quality, promoting more efficient learning (Shermis & Burstein, 2013).

AI is also used to create immersive experiences through augmented and virtual reality (AR/VR), allowing learners to practice communication in realistic sociocultural contexts (Lin & Lan, 2015; Qiu et al., 2024). These tools bridge the gap between theoretical knowledge and authentic language use. Nevertheless, integrating AI into language instruction presents ethical and practical concerns, including issues related to learner autonomy, data security, algorithmic bias, and the pedagogical quality of training data (Selwyn, 2021; European Union Agency for Fundamental Rights, 2022). These challenges underscore the need for teachers who feel confident and competent in making informed decisions about AI integration—reinforcing the importance of assessing teachers'

AI-related self-efficacy.

Method

Development of the Scale

The theoretical framework that forms the basis of the scale was determined by reviewing the literature on the assessment of foreign language teachers' competencies in the use of artificial intelligence. In this process, existing tools and approaches used in the assessment of foreign language teachers' competencies were examined. Thus, to ensure the content validity of the scale, the sub-dimensions to be evaluated and the indicators belonging to these dimensions were determined.

The items to be included in the scale were written in line with the information obtained from the relevant literature and expert opinions. As a result of the relevant literature and expert opinions, 39 items were written and the answers to the items were formed in the form of a five-point Likert (strongly agree, agree, undecided, disagree, strongly disagree). Each item was designed to express the behavior or skill to be assessed in a clear and understandable way.

The following principles were taken into consideration in item writing:

Clarity and Comprehensibility: Care was taken to ensure that the items were clear, understandable and interpretable.

Unidimensionality: Each item is designed to measure only one behavior or skill.

Avoiding the Use of Negative Expressions: By avoiding negative expressions, it is aimed that the participants make the correct interpretation.

Following the item writing, the opinions of academicians and educators who are experts in their fields were obtained to evaluate the content validity of the items in the scale. In line with the feedback received from the experts, necessary corrections and arrangements were made. In particular, it was evaluated whether the items were in compliance with grammar rules and whether they adequately covered the targeted behaviors. A pilot study was conducted to test the validity and reliability of the scale. In line with the data obtained from the participants who participated in the pilot study, item analysis and factor analysis were conducted.

Confirmatory Factor Analysis (CFA) was applied to confirm the factor structure, and the construct validity of the scale was tested (Fornell & Larcker, 1981; Dunn, Baguley, & Brunsden, 2014). In the reliability analysis, Cronbach's Alpha coefficient was calculated, and the internal consistency of the scale was evaluated. As a result of the pilot study and validity-reliability analyses, necessary adjustments were made, and the final scale form was formed as 18 items (see Appendix A and B for Turkish and English versions).

Participants

In this study, an online form was created through Google Forms to collect data. In the introduction of the form, an explanation was added to the participants about the purpose of the research, that the confidentiality of the data would be protected, and that voluntary participation was essential. Subsequently, the form containing demographic information was added and the measurement tool was included after the demographic information. Since each question was marked compulsorily, there were no missing values in the study. Demographic information of the research sample is presented in Table 1.

Table 1 Demographic Information

Variable	Category	N (%)
Gender	Female	217 (79.5)
	Male	56 (20.5)
Branch	English Teacher	215 (78.8)
	German teacher	11 (4.0)
	Turkish Language Teacher	31 (11.4)
	Primary School Teacher	12 (4.4)
	Others	4 (1.5)
Professional experience	1-5 year	50 (18.3)
	6-10 year	94 (34.4)
	11-15 year	59 (21.6)
	16-20 year	70 (25.6)
Faculty	Faculty of Education	156 (57.1)
	Faculty of Arts and Sciences	117 (42.9)
Training Received	Yes	149 (54.6)
	No	124 (45.4)

The research sample consisted of a total of 273 people, 217 women and 56 men.

Data Analysis

Data analysis was conducted in the following sequential phases to ensure the methodological rigor of the scale development process.

Exploratory Factor Analysis (EFA)

Exploratory Factor Analysis (EFA) was conducted to understand the latent structure of the scale and to provide an empirical basis for the subsequent confirmatory analysis. EFA was preferred at this stage because no prior scale existed in the literature that measured competence perceptions regarding the use of artificial intelligence in foreign language teaching; therefore, an empirical exploration of the factor structure was necessary. Before conducting EFA, the Kaiser–Meyer–Olkin (KMO) and Bartlett's Test of Sphericity were examined to determine the suitability of the data. The KMO value exceeding the .60 threshold and the significance of Bartlett's test indicated that the correlation matrix was factorable. The Maximum Likelihood (ML) extraction method was used because it enables statistical testing of factor solutions and is appropriate when data approximate a multivariate normal distribution, a condition met in this study. ML was also preferred because it allows for model comparison and statistical inference (e.g., significance testing), which aligns with our intention to follow EFA with CFA. The Promax rotation was selected as an oblique rotation technique allowing factors to correlate. Eigenvalues greater than 1, a minimum of 5% explained variance per factor, and factor loadings of at least .30 were used as criteria for factor retention (Seçer, 2017). Additionally, a minimum difference of .10 between cross-loadings was applied to determine item specificity. All EFA procedures were performed in SPSS 21.

Confirmatory Factor Analysis (CFA)

A Confirmatory Factor Analysis (CFA) was performed to validate the factor structure identified during EFA. CFA was conducted using the Maximum Likelihood (ML) estimation method. ML was selected due to its robustness in handling continuous Likert-type data and its capacity to provide reliable fit indices when the sample size is adequate, as was the case in this study. A second-order factor model was constructed because the conceptual framework underlying the scale assumes that the first-order factors represent interrelated dimensions of a broader latent construct—competence in using artificial intelligence in foreign language teaching. This hierarchical structure is theoretically supported by prior research indicating that domain-specific competencies often manifest as multidimensional subskills contributing to an overarching ability (Gerbing et al., 1994). A second-order structure also enhances the interpretability of the scale by allowing researchers to examine both subdimension scores and an overall AI-competence score. Model fit was evaluated using widely accepted criteria: a χ^2/df ratio below 5, CFI, TLI, and IFI values of .90 or above, and an SRMR value of .08 or below, as recommended by Marsh et al. (2005). All CFA analyses were conducted using MPLUS 8.10.

Reliability Analysis

Following the validation of the factor structure, reliability analyses were performed using both Cronbach's alpha (α) and McDonald's omega (ω). The inclusion of omega in addition to alpha was intentional, as omega provides a more accurate estimate of internal consistency in multidimensional scales by accounting for factor loadings (Dunn et al., 2014). Values above .70 for both coefficients indicated satisfactory reliability (Kline, 2011). Convergent validity was assessed by examining the Average Variance Extracted (AVE) for each factor. AVE values exceeding .50 demonstrated that the scale items sufficiently captured the theoretical constructs they were intended to measure (Fornell & Larcker, 1981).

Network Analysis

Network analysis was employed to explore the inter-item dynamics of the scale and to determine whether the network structure differed according to gender and years of experience. This approach was chosen because network models allow the visualization of item-level associations and can reveal central items that play a key role in the functioning of the construct—an analytical advantage not provided by traditional factor analyses. Separate networks were estimated for gender and experience groups. Nodes represented scale items, and edges represented partial correlations between items. The thickness of the edges reflected the strength of the associations. To quantify the importance of individual items, expected influence was used as the centrality index. Expected influence was chosen because recent methodological literature suggests it provides more stable and interpretable results than strength or degree centrality in psychological networks (Robinaugh et al., 2016). Networks were estimated using the EBICglasso method with a tuning parameter of $\lambda = 0.50$. EBICglasso was selected because it regularizes small and potentially spurious edges, producing a more interpretable sparse network—an approach recommended for psychological scale items where multicollinearity is common. All analyses were performed using JASP 0.11.1.0 (Epskamp et al., 2018; Bloch et al., 2023). To compare networks across groups, the Network Comparison Test (NCT) was conducted in R. NCT evaluates differences in both Network Structure Invariance (M), which tests whether the pattern of item connections differs across groups, and Global Strength (S), which assesses the overall connectivity of the network. These metrics provide a comprehensive understanding of whether the construct operates similarly across demographic subgroups (van Borkulo et al., 2022).

Results

Before analyzing the data, the data collected through the online form were downloaded, organized and transferred to SPSS. In SPSS, the data were examined in terms of missing data, extreme values, normality and made suitable for analysis (Seçer, 2017; Pallant, 2013). The results obtained are shown in Table 2. When the table was examined, it was seen that the data were normally distributed and there were no outliers (skewness < |3| and kurtosis < |10|; Kline, 2016) and the data were considered suitable for analysis.

Table 2 Descriptive Statistics

	Mean	Sd	Skew.	Kurt.
Total Score	74.31	13.75	-.593	.695

Exploratory Factor Analysis

As a result of the analysis performed for the suitability of the scale for factor analysis, it was seen that the KMO value was .94 and the Barlett test $\chi2$ value was 5951,109 ($p < .001$) and it was concluded that the data were suitable for analysis. As a result of the EFA, the cut-off score criterion was determined as .50 and it was seen that there was no item below this value and the values in the scale were found to be sufficient. However, it was seen that one item (Item 1) in the scale loaded on two factors overlappingly (>.10), so the relevant item was removed from the scale structure. As a result of the repeated analysis, it was seen that the item factor loadings were sufficient (> .30) and there were no overlapping items (< .10). In the Scree Plot graph drawn to determine the number of factors, it was determined that the scale had a two-factor structure (see Figure 1).

When Figure 1 is examined, it is seen that a significant portion of the variance is explained by the first factor, followed by the second factor. As a result, it was concluded that the scale has a two-factor structure. The values of the factor loadings obtained accordingly are presented in Table 3.

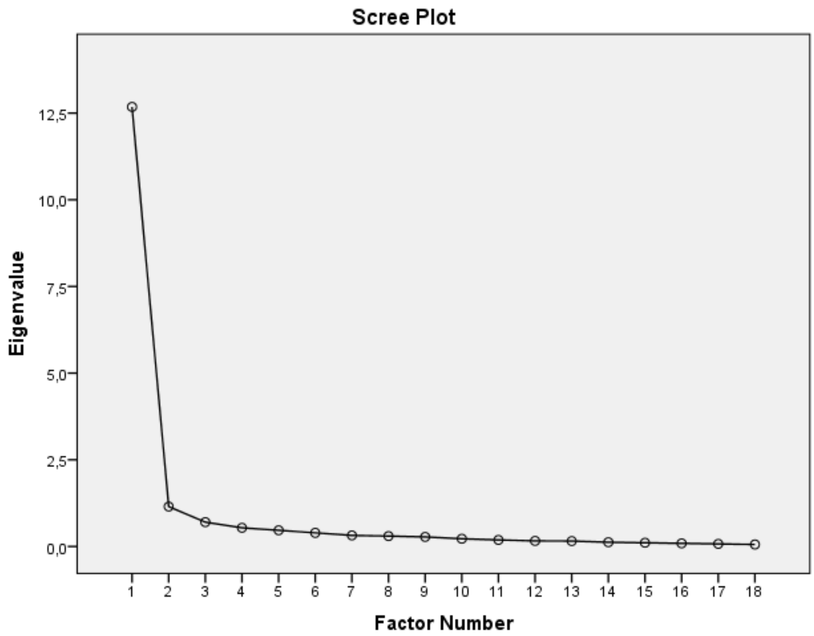


Figure 1. Scree Plot

Table 3. Variances Explained by the Scale of Perception of Efficacy in the Use of Artificial Intelligence in Foreign Language Teaching and Item Analyses

	Factor 1	Factor 2	Item Total Correlation
Item 14	.910		.826
Item 20	.880		.821
Item 16	.876		.727
Item 15	.770		.834
Item 10	.714		.639
Item 4	.694		.795
Item 8	.693		.876
Item 19	.655		.852
Item 18	.642		.847
Item 7	.642		.861
Item 9	.641		.808
Item 2	.585		.827
Item 30		1.007	.835
Item 32		.851	.834
Item 31		.827	.879
Item 38		.821	.825
Item 29		.763	.831
Item 34		.682	.859
Variance Explained	71.07	5.67	
Total Variance explained	76.75		

As a result of the EFA analysis, it was observed that the item factor loadings ranged between .585 and 1.007 and the item total correlations ranged between .639 and .879. As a result, it was seen that the factor loadings were sufficient by meeting the criterion of not being below .30 stated in the literature (Kartal & Bardakçı , 2018). As a result, the scale explains 77% variance, which fulfills the criterion that the total variance ratio should be above 50%. In addition, the first factor of the scale explains 71% variance and the second factor explains 5% variance, and each factor meets the criterion of explaining over 5% variance. As a result, the scale consisting of 18 items and 2 sub-dimensions was evaluated to be adequate and the sub-dimension consisting of 12 items was named as planning and instruction and the sub-dimension consisting of 6 items was named as measurement and evaluation. To evaluate the multicollinearity problem, the relationships between the sub-dimensions of the scale were evaluated by Pearson correlation analysis. As a result of the analysis, it was seen that the relationship between the sub-dimensions was .85 ($p < 0.01$). The fact that this value is not .90 and above indicates that there is no multicollinearity problem.

Confirmatory Factor Analysis (CFA)

The results of the first and second level CFA conducted to examine the fit values of the structure obtained as a result of EFA are given in Figure 2 and Figure 3.

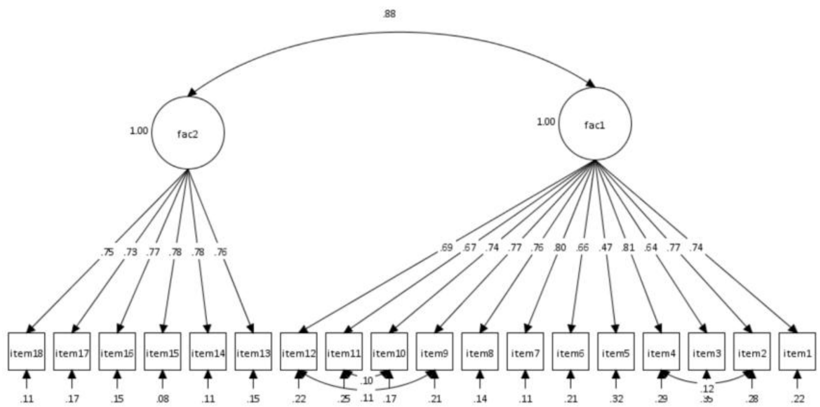


Figure 2 First Level CFA Results

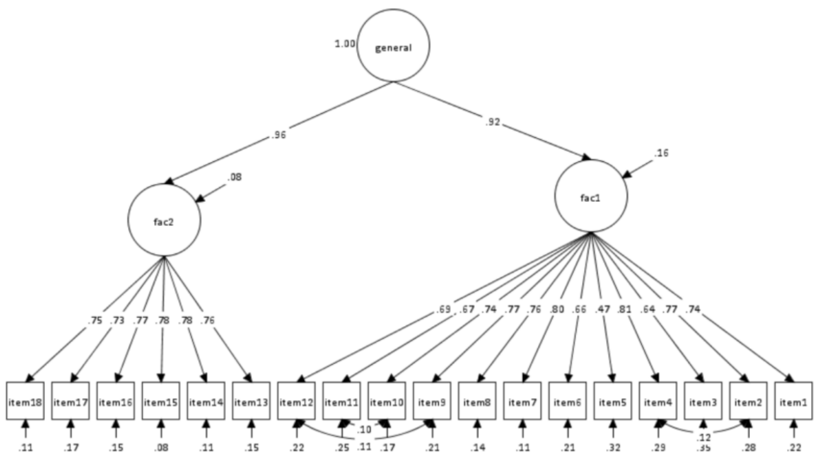


Figure 3 Second Level CFA Results

When the model in Figure 2 was examined, it was seen that the scale consisting of 18 items and 2 sub-dimensions had acceptable fit values (x2/df (628.695/ 131): 4.79; CFI= .918, TLI= .904, IFI= .918, SRMR= 0.03). The second level CFA results of the scale are presented in Figure 3. When the second level model in Figure 3 was examined, it was seen that the scale consisting of 18 items and 2 sub-dimensions had acceptable fit values (x2/df (628.695/130): 4.83; CFI= .917, TLI= .903, IFI= .918, SRMR= 0.03).

Reliability Analysis

The findings obtained from the reliability analysis of the scale are given in Table 4.

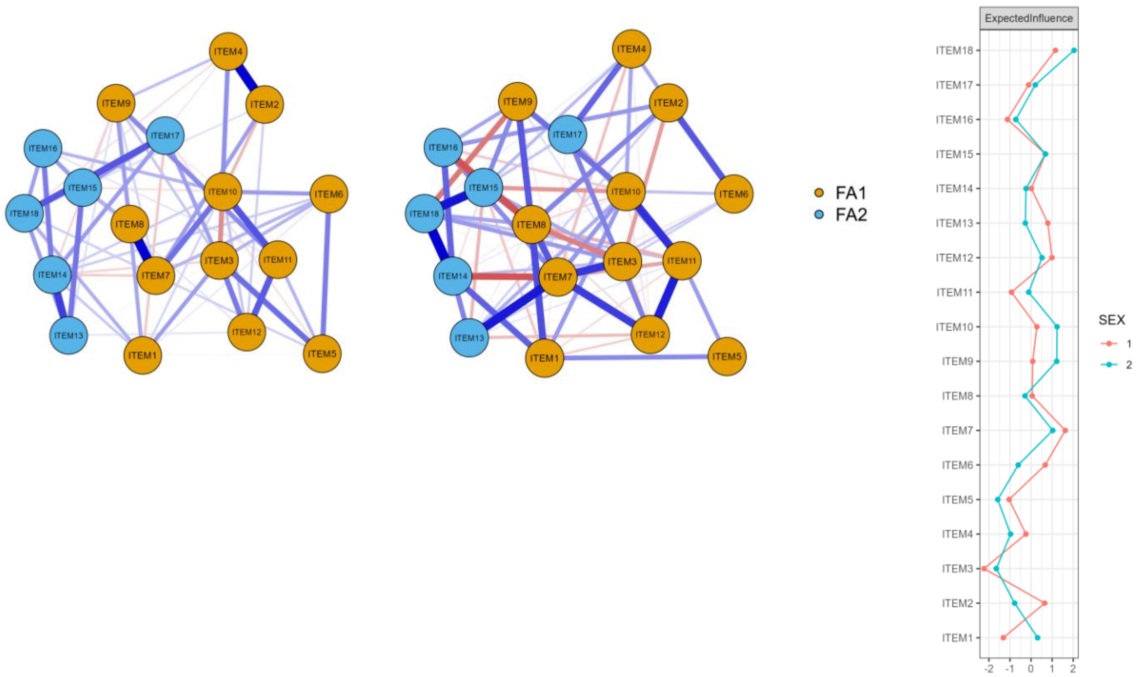
Table 4. Cronbach Alpha, McDonald's Omega Reliability Coefficient and AVE Results of Scale X

Sub-dimensions	α	ω	AVE
Factor 1	.96	.96	.82
Factor 2	.96	.96	.70
Whole scale	.97	.97	-

The reliability coefficients of the scale were found to be at an acceptable level with values above .70 (Seçer, 2017; Pallant, 2013). When the AVE values were examined, it was seen that the values were above .50 and as a result, it was evaluated that the scale had convergent validity.

Network Analysis

The network structures obtained as a result of the network analysis and the visuals of the expected centrality of influence index are presented in Figure 4 and the values are presented in Table 5.



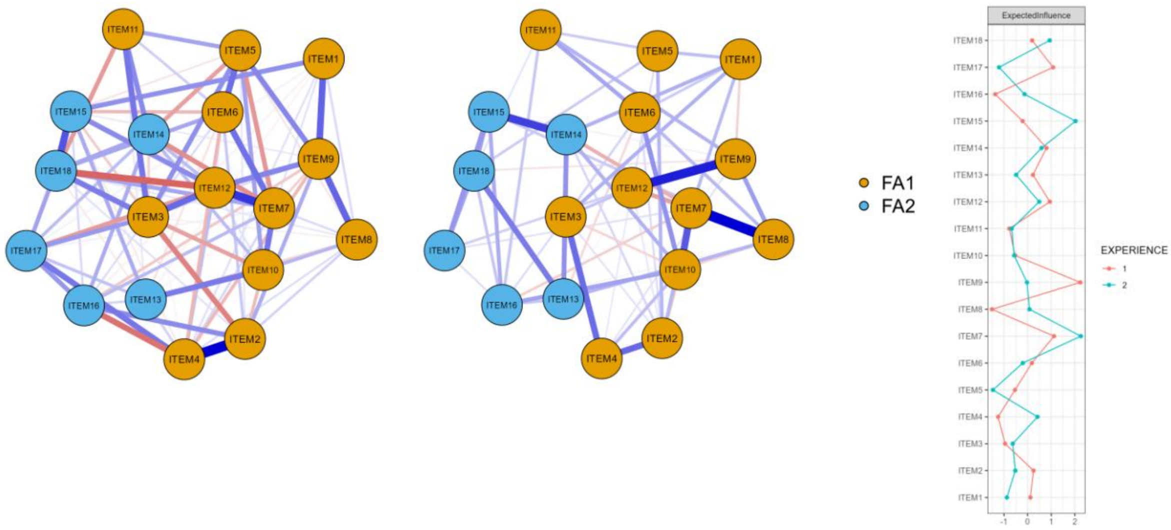


Figure 4. Network Structure (female network on the top left, male network on the top right, network with 1-5 years of experience on the bottom left, network with 5 years of experience or more on the bottom right) and Expected Centrality of Influence Indices

When the network structures of female and male in Figure 4 are examined, it is seen that the items belonging to the sub-dimensions are generally clustered together. However, when centrality indices were examined, it was observed that item 7 was more central in the network of women and item 18 was more central in the network of men (see Table 3). Similarly, when the network structures according to years of experience were examined, it was seen that the items belonging to the sub-dimensions were clustered together, but item 9 was more central in the network of teachers with 1-5 years of experience, while item 7 was more central in the network of teachers with more than 5 years of experience (see Table 5).

Table 5. Expected Influence Values

Variable	Gender		Experience	
	Female	Male	1-5 year	5+ year
ITEM1	-1.230	-0.589	0.120	-0.879
ITEM2	-0.578	-1.026	0.251	-0.520
ITEM3	-0.936	-1.265	-0.956	-0.631
ITEM4	0.535	-1.199	-1.251	0.419
ITEM5	-0.381	-1.607	-0.535	-1.470
ITEM6	-0.057	-0.461	0.178	-0.206
ITEM7	1.374	0.661	1.119	2.254
ITEM8	0.016	0.373	-1.512	0.079
ITEM9	-0.942	1.951	2.233	-0.025
ITEM10	0.229	0.263	-0.530	-0.568
ITEM11	-0.159	0.751	-0.776	-0.674

Variable	Gender		Experience	
	Female	Male	1-5 year	5+ year
ITEM12	1.162	-0.242	0.943	0.492
ITEM13	-0.824	0.150	0.225	-0.488
ITEM14	1.018	0.863	0.801	0.591
ITEM15	1.906	0.059	-0.211	2.031
ITEM16	-1.038	-0.761	-1.373	-0.129
ITEM17	-1.314	0.271	1.081	-1.212
ITEM18	1.220	1.808	0.192	0.936

As a result of the network comparison test applied to examine whether there is a difference between the networks, it was observed that there was a differentiation in the invariance analysis of the networks belonging to experience ($M= 0.532$, $p= 0.012$), but there were no statistically significant differences in the global power invariance test ($S= 0.673$, $p= 0.115$). In parallel with this result, it was observed that there was a differentiation in the network invariance analysis for examining the networks of gender ($M= 0.491$, $p= 0.032$), but there were no statistically significant differences in the global power invariance test ($S= 0.738$, $p= 0.090$).

Discussion

In this study, the validity and reliability analyses of the scale developed to measure the perception of competence in the use of artificial intelligence in foreign language teaching were conducted. In the scale development process, an item pool was created, and after the pilot application, the construct validity of the scale was examined through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Reliability analyses further demonstrated the consistency of the scale, and network analysis provided a detailed examination of item-level relationships. Overall, the findings indicate that the scale is psychometrically adequate. According to the EFA results, the scale had a two-factor structure explaining 76.75% of the total variance. Factor loadings ranged between .585 and 1.007 and item-total correlations between .639 and .879, suggesting strong item discrimination. Considering that factor loadings above .30 are accepted as sufficient (Kartal & Bardakçı, 2018), the values obtained in this study indicate solid construct validity.

The CFA results confirmed the two-factor structure, and the fit indices were at acceptable or good levels. A chi-square/df ratio below 5, CFI, TLI, and IFI values above .90, and an SRMR value below .08 indicate good model fit (Marsh et al., 2005). Accordingly, the measurement model aligns with the theoretical structure, demonstrating satisfactory structural validity. Reliability analyses showed that Cronbach's alpha and McDonald's omega values ranged between .96 and .97, which reflects high internal consistency. Coefficients above .70 are considered adequate for reliability (Kline, 2011; Dunn et al., 2014). Additionally, AVE values above .50 confirm convergent validity (Fornell & Larcker, 1981).

The results of the network analysis revealed that the relationships between the scale items differed according to gender and years of professional experience. Item 7 was more central among female teachers, while item 18 was more central among male teachers. Moreover, item 9 played a more central role for teachers with 1–5 years of experience, whereas item 7 was more central for teachers with more than 5 years of experience. These findings suggest that perceptions of competence in using artificial intelligence technologies vary depending on demographic and professional characteristics. The literature similarly highlights the role of gender and experience in technology-related perceptions (Venkatesh & Morris, 2000; Teo, 2008).

Beyond these findings, an important consideration concerns the cultural and contextual generalizability of the scale. Since the data were collected from a specific educational context operating under one national curriculum, teachers' perceptions of AI-related competencies may reflect the technological infrastructure, institutional priorities, and cultural attitudes prevalent within that system. Educational systems with more established AI integration policies, stronger digital literacy initiatives, or more resource-rich environments may exhibit different patterns of self-efficacy compared with contexts where technological resources are limited or traditional pedagogical approaches dominate. Therefore, future studies should test the scale across diverse cultural, linguistic, and institutional settings to ensure external validity and to better understand how contextual differences shape teachers' AI-related competence perceptions.

Conclusion

In this study, a valid and reliable scale was developed to measure perceptions of competence in using artificial intelligence in foreign language teaching. The two-factor structure—planning and instruction, and measurement and evaluation—was confirmed, indicating that the scale can serve as a comprehensive instrument for evaluating foreign language teachers' competence in using AI technologies. Considering the observed differences by gender and professional experience, differentiated professional development pathways—such as beginner, intermediate, and advanced AI competency strands—would help address diverse teacher needs more effectively.

The findings emphasize the need to strengthen teachers' competencies in using AI-based tools more effectively in education. Importantly, the cultural and contextual specificity of the current sample should be acknowledged. The scale's psychometric properties were established within a single national context, and teachers' perceptions may vary across regions with different technological infrastructures, pedagogical traditions, and policy frameworks. Therefore, future research should examine the validity and reliability of the scale in diverse cultural and linguistic environments. Moreover, exploring the relationship between the scale and additional variables such as digital literacy or attitudes toward technology may provide deeper insight into factors influencing AI-related competence.

In conclusion, the scale developed in this study demonstrates strong scientific validity and reliability for measuring competence perceptions related to AI use in foreign language teaching. Policymakers and teacher education program designers are encouraged to consider these findings when developing strategies to enhance teachers' AI-related competencies, while also recognizing that broader cultural and institutional contexts may influence the effectiveness and applicability of such initiatives.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control.* Freeman.

Blake, R. (2013). *Brave new digital classroom: Technology and foreign language learning.* Georgetown University Press.

Bloch, F., Jackson, M.O., & Tebaldi, P. (2023). Centrality measures in networks. *Social Choice and Welfare, 61*, 413-453. https://doi.org/10.1007/s00355-023-01456-4

De Freitas, S. (2018). Are Games Effective Learning Tools? A Review of Educational Games. *Educational Technology and Society*, *21*, 74-84.

Derakhshan, A., & Hasanabbasi, S. (2015). Social networks for language learning. *Theory and Practice in Language Studies*, *5*(5), 1090-1095, http://dx.doi.org/10.17507/tpls.0505.

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,* 399-412. https://doi.org/10.1111/bjop.12046

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

Ertmer, P. A., & Ottenbreit-Leftwich, A. (2010). Teacher technology change: How knowledge, beliefs, and culture intersect. *Journal of Research on Technology in Education, 42*(3), 255-284. https://doi.org/10.1080/15391523.2010.10782551

European Union Agency for Fundamental Rights. (2022). *Bias in algorithms – Artificial intelligence and discrimination*. https://fra.europa.eu/en/publication/2022/bias-algorithm

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

Gerbing, D. W., Janet G. H., & Elizabeth B. F. (1994). A large-scale second-order structural equation model of the ınfluence of management participation on organizational planning benefits. *Journal of Management* 20 (4), 859–885. https://doi.org/10.1016/0149-2063(94)90033-7.

Ghosh, S. S. (2024). Assessing the Efficacy of Intelligent Tutoring Systems in Language Teacher Education: A Quantitative Study. *Computers in the Schools*, 1–24. https://doi.org/10.1080/07380569.2024.2435308

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70-105. https://doi.org/10.1080/09588221.2012.700315

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning. Parsers and pedagogues*. Routledge, https://doi.org/10.4324/9780203012215

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning.* The Center for Curriculum Redesign.

Jiang, M. Y.-C., Jong, M. S.-Y., Lau, W. W.-F., Chai, C.-S., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian Journal of Educational Technology*, *37*(2), 110–131. https://doi.org/10.14742/ajet.6798

Kartal, S., & Bardakçı, M. (2018). Ölçek geliştirme çalışmalarında dikkat edilmesi gereken istatistiksel kriterler.

Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, *9*(2), 123-140. https://doi.org/10.21031/epod.395937

Karagöz, Y. (2017). *SPSS ve AMOS uygulamalı nitel-nicel-karma bilimsel araştırma yöntemleri ve yayın etiği.* Ankara: Nobel Akademik Yayıncılık.

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

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile-assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL, 20*(3), 271-289. https://doi.org/10.1017/S0958344008000335

Lin, C.C.; Huang, A.Y., & Lu, O.H. (2023). Artificial intelligence inintelligent tutoring systems toward sustainable education: A systematicreview. *Smart Learning Environments*, *10*(1), DOI: 10.1186/s40561-023-00260-y

Lin, T. J., & Lan, Y. J. (2015). Language Learning in Virtual Reality Environments: Past, Present, and Future. EducationalTechnology & Society, 18 (4), 486–497 *Language Learning in Virtual Reality Environments: Past, Present, and Future*.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education.* Pearson Education.

Marsh, H.W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. A. Maydeu-Olivares & J. McArdle (Eds.), *Psychometrics. A festschrift to Roderick P. McDonald. Erlbaum*, Hillsdale, NJ (2005), pp. 275-340

Onesi-Ozigagun, O., Ololade, Y.J., Eyo-Udo, N.L. & Ogundipe, D.O. (2024). Revolutionizingeducation through AI: a comprehensive review of enhancing learning experiences. *International Journal of Applied Research in Social Sciences*, *6*(4), 589-607, DOI: 10.1186/s40561-023-00260-y

Paladines, J., & Ramirez, J. (2020). A systematic literature review of intelligent tutoring systems withdialogue in natural language. *IEEE Access, 8,* 164246–164267. https://doi.org/10.1109/ACCESS.2020.3021383

Qiu, X.b., Shan, C., Yao, J. & Fu, Q.K. (2024). The effects of virtual reality on EFL learning: A meta-analysis. *Educ Inf Technol* 29 (2), 1379–1405 https://doi.org/10.1007/s10639-023-11738-0

Robinaugh, D.J., Millner, A.J., & McNally, R.J., 2016. Identifying highly influential nodes in the complicated grief network. *J Abnorm Psychol 125*, 747–757. https://doi.org/10.1037/abn0000181

Sack, O. & Röcker, C. (2013). Privacy and security issues in technology-enhanced foreign language learning environments: Exlporing users' knowledge about technological processes of diverse user groups. *Universal Journal of Psychology, 1*(2), 72-83. ttps://doi.org/10.13189/ujp.2013.010207

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education, 128*, 13-35. https://doi.org/10.1016/j.compedu.2018.09.009

Selwyn, N. (2021). *Education and Technology: Key Issues and Debates.* Bloomsbury Publishing.

Shadiev R, & Liu J. (2023). Review of research on applications of speech recognition technology to assist language learning. *ReCALL*, *35*(1):74-88. doi:10.1017/S095834402200012X

Shermis, M.D., & Burstein, J. (Eds) (2013). Handbook of Automated Essay Evaluation: Current applications and new directions. (2013). *Journal of Writing Research*, *5*(2), 239-243. https://doi.org/10.17239/jowr-

2013.05.02.4

Teo, T. (2008). Pre-service teachers' attitudes towards computer use: A Singapore survey. *Australasian Journal of Educational Technology, 24*(4), 413-424. https://doi.org/10.14742/ajet.1201

Thi-Nhu Ngo T, Hao-Jan Chen, H. & Kuo-Wei Lai, K. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, *36*(1):4-21. doi:10.1017/S0958344023000113

Tondeur, J., van Braak, J., Ertmer, P. A., & Ottenbreit-Leftwich, A. (2016). Understanding the relationship between teachers' pedagogical beliefs and technology use in education: A systematic review of qualitative evidence. *Educational Technology Research Development, 65*(3), 555-575. https://doi.org/10.1007/s11423-016-9492-z

Van Borkulo, C. D., van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2022). Comparing network structures on three aspects: A permutation test. *Psychological Methods*. https://doi.org/10.1037/met0000476

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly, 24*(1), 115-139. https://doi.org/10.2307/3250981

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zhang J. (2025). Integrating chatbot technology into English language learning to enhance student engagement and interactive communication skills. *Journal of Computational Methods in Sciences and Engineering*. doi:10.1177/14727978241312992

Appendix A. Yabancı Dil Öğretiminde Yapay Zekâ Kullanımına Yönelik Yeterlik Algısı Ölçeği * [Turkish Version]

			Tamamen katılıyorum	Katılıyorum	Kararsızım	Katılmıyorum	Hiç katılmıyorum
Planlama ve Öğretim	1	Yabancı dil öğretim sürecinde yapay zekâ araçlarını kullanarak etkili ve ilgi çekici öğrenme ortamları oluşturabilirim.	☐	☐	☐	☐	☐
	2	Güncel ve çağdaş yabancı dil eğitimi yöntemlerine uygun içerikler oluşturmada yapay zekâdan yardım alabilirim.	☐	☐	☐	☐	☐
	3	Yabancı dil öğretiminde yapay zekâ araçlarını kullanarak öğrencilerin dil öğrenme ihtiyaçlarına uygun öğrenme etkinlikleri oluşturabilirim.	☐	☐	☐	☐	☐
	4	Yapay zekâ tabanlı uygulamaları kullanarak öğretim yabancı dil öğretim stratejilerimi çeşitlendirebilirim.	☐	☐	☐	☐	☐
	5	Yapay zekâyı kullanarak yabancı dil öğretiminde ders materyallerimi güncel ve ilgi çekici bir hâle getirebilirim.	☐	☐	☐	☐	☐
	6	Yapay zekâ tabanlı araçları kullanarak yabancı dil öğretimine yönelik öğretim materyalleri geliştirebilirim.	☐	☐	☐	☐	☐
	7	Yapay zekâ kullanarak öğrencilerime kişiselleştirilmiş öğrenme deneyimleri sağlayacak yabancı dil öğretim materyalleri geliştirebilirim.	☐	☐	☐	☐	☐
	8	Öğrencilerin yabancı dil öğrenme motivasyonunu artırmada yapay zekâdan yardım alabilirim.	☐	☐	☐	☐	☐
	9	Yabancı dil öğretiminde yapay zekâ araçlarını kullanarak öğrencilerin derse katılımını artırabilirim.	☐	☐	☐	☐	☐
	10	Yabancı dil öğretimine yönelik materyalleri hazırlarken yapay zekâ tarafından sunulan önerileri dikkate alabilirim.	☐	☐	☐	☐	☐
	11	Yapay zekâ, benim için dil becerilerinin öğretiminde kullanabileceğim öğrenme ve öğretme materyalleri geliştirmede kullanışlı bir araçtır.	☐	☐	☐	☐	☐
	12	Dil öğretimine yönelik materyal hazırlarken öğrenme çıktılarını belirlemede yapay zekâdan yardım alabilirim.	☐	☐	☐	☐	☐
Ölçme ve Değerlendirme	13	Dil becerilerine yönelik yapay zekâ tabanlı ölçme ve değerlendirme araçlarını kullanarak öğrenci seviyelerini takip edebilirim.	☐	☐	☐	☐	☐
	14	Öğrencilerin yabancı dil becerilerindeki (dinleme, okuma, konuşma, yazma) performanslarını yapay zekâ kullanarak analiz edebilirim.	☐	☐	☐	☐	☐
	15	Yabancı dil öğretiminde yapay zekâ araçlarını kullanarak öğretim sürecinin etkililiğini değerlendirebilirim.	☐	☐	☐	☐	☐
	16	Yabancı dil öğrenen öğrencilerin dil gelişimini izlemede yapay zekânın sunduğu verileri etkin şekilde değerlendirebilirim.	☐	☐	☐	☐	☐
	17	Ölçme ve değerlendirmede yapay zekâ tabanlı araçları etkin bir şekilde kullanabilirim.	☐	☐	☐	☐	☐
	18	Yabancı dil öğretiminde yapay zekâ ile oluşturulan geri bildirim sistemlerini kullanarak öğrencilerime dönüt verebilirim.	☐	☐	☐	☐	☐

* The scale was developed and psychometrically validated in Turkish; therefore, the **original Turkish version** is provided as the final instrument. An **English translation is provided for readers' convenience**. The English version is **for reference only** and has **not** been validated for administration unless explicitly stated.

Appendix B. Efficacy Perception Scale for the Use of Artificial Intelligence in Foreign Language Teaching

			Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
Planning and Instruction	1	I can create effective and engaging learning environments by using artificial intelligence tools in the foreign language teaching process.	☐	☐	☐	☐	☐
	2	I can use artificial intelligence to create content that is appropriate for current and modern foreign language teaching methods.	☐	☐	☐	☐	☐
	3	I can create learning activities appropriate to students' foreign language learning needs by using artificial intelligence tools.	☐	☐	☐	☐	☐
	4	I can diversify my foreign language teaching strategies by using artificial intelligence-based applications.	☐	☐	☐	☐	☐
	5	I can make my course materials up-to-date and engaging in foreign language teaching by using artificial intelligence.	☐	☐	☐	☐	☐
	6	I can develop instructional materials for foreign language teaching using artificial intelligence-based tools.	☐	☐	☐	☐	☐
	7	I can develop foreign language teaching materials that provide personalized learning experiences for my students by using artificial intelligence.	☐	☐	☐	☐	☐
	8	I can use artificial intelligence to increase students' motivation for learning a foreign language.	☐	☐	☐	☐	☐
	9	I can increase students' participation in the lesson by using artificial intelligence tools in foreign language teaching.	☐	☐	☐	☐	☐
	10	I can consider the suggestions provided by artificial intelligence when preparing materials for foreign language teaching.	☐	☐	☐	☐	☐
	11	Artificial intelligence is a useful tool for me in developing learning and teaching materials that I can use in teaching language skills.	☐	☐	☐	☐	☐
	12	I can use artificial intelligence to determine learning outcomes when preparing materials for language teaching.	☐	☐	☐	☐	☐
Assessment and Evaluation	13	I can monitor students' proficiency levels by using artificial intelligence-based assessment and evaluation tools for language skills.	☐	☐	☐	☐	☐
	14	I can analyze students' performance in foreign language skills (listening, reading, speaking, writing) by using artificial intelligence.	☐	☐	☐	☐	☐
	15	I can evaluate the effectiveness of the teaching process by using artificial intelligence tools in foreign language teaching.	☐	☐	☐	☐	☐
	16	I can effectively evaluate the data provided by artificial intelligence in monitoring the language development of foreign language learners.	☐	☐	☐	☐	☐
	17	I can effectively use artificial intelligence-based tools in assessment and evaluation.	☐	☐	☐	☐	☐
	18	I can provide feedback to my students by using artificial intelligence-generated feedback systems in foreign language teaching.	☐	☐	☐	☐	☐