

Eurasian Research Bulletin 	From Cultural Knowledge to Critical Verification: Differential Development of Intercultural Competence in AI-Supported Independent EFL Learning
--	--

Raykhona Gulomova	Scientific researcher, Uzbekistan State World Languages University
------------------------------	---

ABSTRACT	This quasi-experimental study examines how an integrated AI-supported independent-learning methodology affected four dimensions of intercultural competence (IC) among university EFL students. Participants were 316 students from three higher education institutions in Uzbekistan, with 158 students in the experimental condition and 158 in the control condition. IC was operationalised through cognitive-cultural, affective-attitudinal, behavioural-communicative, and critical-reflective dimensions, each assessed by five performance indicators. Pre-test scores did not differ significantly between groups (60.5 vs. 60.2; $t = 0.29$, $p = .773$). At post-test, the experimental group achieved a mean of 78.6 compared with 64.9 in the control group, $t = 14.48$, $p < .001$, Cohen's $d = 1.63$. Large between-group effects were observed across all four dimensions, with the strongest effect in critical-reflective performance ($d = 1.31$), followed by cognitive-cultural ($d = 1.17$), behavioural-communicative ($d = 1.05$), and affective-attitudinal ($d = 1.03$). The proportion of experimental-group students at sufficient or high IC levels increased from 36.7% to 76.6%, compared with 35.4% to 44.9% in the control group. The findings suggest that intercultural development in AI-supported environments is strongest when cultural inquiry includes source verification, stereotype recognition, reflective revision, contextualised communication, and learner-first reasoning. Because the intervention combined task transformation, self-regulation, teacher facilitation, and AI-supported scaffolding, the results should be interpreted as effects of the integrated methodology rather than of AI in isolation.
---------------------	--

Keywords:	intercultural competence; EFL; artificial intelligence; critical reflection; independent learning; source verification; higher education; quasi-experimental research
----------------------	--

1. Introduction

Intercultural competence has become increasingly important in university EFL education because linguistic accuracy alone does not guarantee successful communication across cultural contexts. Learners may know grammar, vocabulary, and factual information about another society while still misinterpreting indirectness, disagreement, silence, institutional expectations, or culturally

embedded meanings. For this reason, contemporary models of intercultural competence emphasise not only knowledge but also attitudes, interpretation, interaction, and critical cultural awareness (Byram, 1997; Deardorff, 2006; Fantini, 2009).

The expansion of generative and conversational artificial intelligence has added a new layer to this problem. AI can provide immediate explanations, alternative interpretations, simulated interaction, and

feedback, but the same affordances can encourage overreliance, cultural simplification, or uncritical acceptance of fluent but weakly supported claims. Recent studies have therefore moved beyond asking whether AI can support language learning toward examining how AI should be pedagogically constrained so that learner reasoning remains central. Hastomo et al. (2025), for example, found that guided interaction with AI conversational agents could support intercultural exploration and critical questioning, while Sun (2025) reported positive intercultural outcomes in an AI-enabled COIL design. Ma and Shi (2026) similarly foreground human-AI collaboration rather than replacement. A 2026 systematic review of digital approaches to intercultural competence concluded that outcomes depend substantially on implementation logic, mechanism, and assessment alignment rather than technology alone (Naranjo-Toro et al., 2026).

A particularly underdeveloped issue is whether different dimensions of intercultural competence respond equally to structured AI-supported learning. This question matters because an intervention may improve factual cultural knowledge without strengthening perspective-taking, behavioural adaptation, or critical evaluation. In AI-rich environments, the critical-reflective dimension is especially important because students must evaluate cultural claims, identify generalisations, distinguish evidence from confident phrasing, and revise interpretations when stronger contextual evidence becomes available.

The present study addresses this issue by examining dimension-specific outcomes from a multi-site quasi-experimental intervention embedded in independent Country Studies learning. The intervention was not designed as unrestricted AI use. It combined structured intercultural tasks, differentiated independent learning, learner self-regulation, teacher facilitation, and selectively applied AI scaffolding. AI support followed a learner-first sequence: learner attempt, AI-supported scaffold, learner reconsideration, and learner decision. The study therefore evaluates an integrated methodology rather than the isolated effect of a digital tool.

The study addresses three research questions: (1) Did the experimental and control groups differ in overall intercultural competence after the intervention? (2) How did the four dimensions of intercultural competence change, and which dimension showed the strongest between-group effect? (3) Did the intervention alter the distribution of students across developmental performance levels, not only mean scores?

2. Theoretical background

2.1. Intercultural competence as a multidimensional construct

Byram's (1997) framework conceptualises intercultural communicative competence through attitudes, knowledge, skills of interpreting and relating, skills of discovery and interaction, and critical cultural awareness. Deardorff (2006) similarly treats intercultural competence as developmental, linking attitudes, knowledge and skills to internal and external outcomes. Fantini (2009) emphasises the interaction of awareness, attitudes, knowledge and skills. Although these frameworks use different terminology, they converge on the view that intercultural competence cannot be adequately represented by cultural knowledge alone.

For the present study, this convergence was operationalised through four dimensions. The cognitive-cultural dimension concerns contextual knowledge, cultural self-awareness, interpretation and comparison. The affective-attitudinal dimension concerns openness, perspective-taking, tolerance of ambiguity, decentring, and willingness to reconsider an initial judgement. The behavioural-communicative dimension concerns context-sensitive response, adaptation, clarification, interaction strategy, and misunderstanding management. The critical-reflective dimension concerns stereotype recognition, evidence evaluation, verification, critical cultural awareness, and reasoned revision.

These dimensions were treated as related but analytically distinct. A student may demonstrate strong factual and contextual knowledge but still make premature cultural generalisations. Another may display openness

to multiple perspectives yet struggle to select an appropriate communicative response. A third may communicate fluently but fail to question the reliability of cultural claims encountered online. A multidimensional assessment is therefore necessary both for diagnosis and for evaluating intervention effects.

2.2. AI-supported intercultural learning and the need for verification

Recent digital and AI-supported intercultural research has shown that technology can extend access to cultural input, interaction, and reflection. Sun's (2025) AI-enabled COIL intervention reported gains in intercultural competence among Chinese English majors, while Hastomo et al. (2025) showed that students used conversational agents not only to obtain cultural facts but also to explore sensitive questions, compare perspectives, and verify information. Ma and Shi (2026) further argue for a human-AI collaboration framework in which AI support is integrated with pedagogical design.

At the same time, AI-generated cultural information creates a distinctive epistemic risk. Fluent output can appear authoritative regardless of whether a claim is representative, contextualised, or well supported. This makes critical verification part of intercultural learning rather than an external digital-literacy add-on. Learners need opportunities to compare sources, identify overgeneralisation, recognise missing context, and explain why a revised interpretation is stronger than an initial assumption. In the present methodology, these processes were built into the critical-reflective dimension through claim audits, source comparison, media analysis, stereotype challenge, and initial-versus-revised interpretation tasks.

Independent learning also changes the distribution of responsibility. When students work outside direct classroom supervision, teachers cannot immediately correct every stereotypical interpretation or weak cultural claim. Structured self-regulation and bounded pedagogical support are therefore essential. The intervention combined planning, monitoring, help-seeking, self-assessment and adaptation with teacher monitoring and selective AI

scaffolding. This arrangement was intended to preserve learner judgement while increasing opportunities for practice beyond classroom time.

3. Method

3.1. Design and participants

The study used a quasi-experimental pre-test-post-test design with intact university groups. Individual students were not randomly reassigned across institutions or scheduled classes. A total of 316 students participated: 158 in the experimental condition and 158 in the control condition. The participants were drawn from Uzbekistan State World Languages University, Samarkand State Institute of Foreign Languages, and Namangan State Institute of Foreign Languages. The experimental groups completed the structured AI-supported independent-learning methodology, whereas the control groups continued the existing organisation of Country Studies independent work.

At Uzbekistan State World Languages University, 54 students participated in each condition; at Samarkand State Institute of Foreign Languages, 52 participated in each condition; and at Namangan State Institute of Foreign Languages, 52 participated in each condition. The use of existing groups retained the normal curricular context but limits strong causal inference because unmeasured group differences and classroom clustering cannot be excluded.

3.2. Intervention

The experimental methodology integrated intercultural competence development with structured independent learning and controlled AI support. Students completed interpretive, comparative, perspective-taking, behavioural, communicative, verification and reflective tasks. Diagnostic profiles identified relative strengths and weaknesses across the four IC dimensions, after which required tasks ensured common curricular coverage and recommended tasks provided additional practice in weaker areas.

The technological environment was the Intercultural Learning Hub ZRIT. AI was treated as one support component within a broader

learning environment rather than as the intervention itself. Learners normally produced an initial interpretation before receiving AI-supported prompts, contextual hints, alternative perspectives, simulated dialogue, verification cues, or formative feedback. The final interpretation and decision remained the learner's responsibility.

Critical-reflective work included source comparison, AI Claim Audit activities, media analysis, stereotype challenge, verification tasks, and comparison of initial and revised judgements. Behavioural-communicative development used situational judgement, branching scenarios, clarification requests, and simulated dialogue. Affective-attitudinal development relied on ambiguity-based cases and competing perspectives, while cognitive-cultural development transformed factual Country Studies content into contextual interpretation and comparison.

3.3. Assessment framework

The pre-test and post-test were study-specific criterion-referenced performance assessments rather than a commercial intercultural questionnaire or general English proficiency test. Each dimension contained five indicators scored from 0 to 5, yielding a maximum of 25 points per dimension and 100 points overall. Parallel forms used comparable reasoning demands while varying surface contexts.

The assessment framework deliberately separated factual recall from demonstrated intercultural reasoning. Cognitive-cultural evidence came from interpretation, comparison, and source-based cases; affective-attitudinal evidence came from ambiguous situations, perspective-taking, justification and revision; behavioural-communicative evidence came from situational judgement and simulated interaction; critical-reflective evidence came from media analysis, source comparison, claim verification, and initial-versus-revised judgement.

Table 1. Operational dimensions and diagnostic evidence

Dimension	Core indicators	Primary diagnostic evidence
Cognitive-cultural	Contextual knowledge; cultural self-awareness; interpretation; comparison; contextual variation	Cultural interpretation tasks; comparative analysis; source-based cases
Affective-attitudinal	Openness; perspective-taking; tolerance of ambiguity; decentring; reconsideration	Ambiguous situations; competing perspectives; justification and revision
Behavioural-communicative	Context-sensitive response; adaptation; clarification; interaction strategy; misunderstanding management	Situational judgement; role-based response; simulated interaction
Critical-reflective	Stereotype recognition; evidence evaluation; verification; critical cultural awareness; revision	Source comparison; media analysis; claim verification; initial and revised judgement

3.4. Statistical analysis

Group means and standard deviations were compared using independent-samples t

tests at the pre-test and post-test stages, and Cohen's d was used to describe effect magnitude. The main post-test comparison was interpreted together with dimension-level

effects and developmental-level distributions. Statistical significance was not treated as sufficient evidence of educational importance; effect sizes and the direction of change across multiple indicators were also considered (Lakens, 2013).

Because the available analysis was based on aggregated group summaries, the study does not report a fitted student-level repeated-measures model or cluster-adjusted analysis. A non-significant baseline difference is evidence of comparability, not proof of equivalence, and the pooled student-level calculations assume independence. These limitations are retained in the interpretation.

4. Results

4.1. Baseline comparability and overall outcome

At pre-test, the experimental group had a mean overall IC score of 60.5 (SD = 9.4) and the control group had a mean of 60.2 (SD = 9.1). The 0.3-point difference was not statistically significant, $t = 0.29$, $p = .773$. No dimension-level pre-test comparison was significant: cognitive-cultural $p = .573$, affective-attitudinal $p = .778$, behavioural-communicative $p = .785$, and critical-reflective $p = .791$.

At post-test, the experimental group mean was 78.6 and the control group mean was 64.9. The final between-group difference was 13.7 points, $t = 14.48$, $p < .001$, with Cohen's $d = 1.63$. The experimental group gained 18.1 points from pre-test to post-test, whereas the control group gained 4.7 points. These figures indicate a much stronger pattern of development under the experimental methodology, while still showing some improvement under regular Country Studies instruction.

Table 2. Overall intercultural competence outcomes

Indicator	Experimental group	Control group
N	158	158
Pre-test mean (SD)	60.5 (9.4)	60.2 (9.1)
Post-test mean	78.6	64.9
Mean gain	+18.1	+4.7
Post-test t	14.48	
Post-test p	< .001	
Cohen's d	1.63	

4.2. Differential development across the four IC dimensions

Development was observed across all four dimensions, but the magnitude was not identical. The experimental group improved by

4.6 points in cognitive-cultural performance, 4.1 in affective-attitudinal performance, 4.2 in behavioural-communicative performance, and 5.2 in critical-reflective performance. Control-group gains ranged from 1.1 to 1.4 points.

Table 3. Development across intercultural competence dimensions

Dimension	Experimental pre	Experimental post	Experimental gain	Control pre	Control post	Control gain
Cognitive-cultural	15.8	20.4	+4.6	15.6	17.0	+1.4
Affective-attitudinal	15.2	19.3	+4.1	15.1	16.2	+1.1
Behavioural-communicative	14.9	19.1	+4.2	14.8	15.9	+1.1

Critical-reflective	14.6	19.8	+5.2	14.7	15.8	+1.1
---------------------	------	------	------	------	------	------

The post-test between-group effects were large for every dimension. The critical-reflective dimension produced the largest effect, $d = 1.31$, followed by cognitive-cultural, $d = 1.17$, behavioural-communicative, $d = 1.05$, and affective-attitudinal, $d = 1.03$.

Table 4. Post-test between-group comparisons by dimension

Dimension	Experimental Mean $\pm$ SD	Control Mean $\pm$ SD	t	p	Cohen's d
Cognitive-cultural	20.4 $\pm$ 2.8	17.0 $\pm$ 3.0	10.41	< .001	1.17
Affective-attitudinal	19.3 $\pm$ 2.9	16.2 $\pm$ 3.1	9.18	< .001	1.03
Behavioural-communicative	19.1 $\pm$ 3.0	15.9 $\pm$ 3.1	9.32	< .001	1.05
Critical-reflective	19.8 $\pm$ 2.9	15.8 $\pm$ 3.2	11.64	< .001	1.31

4.3. Movement across developmental levels

The intervention altered not only group means but also the distribution of students across developmental levels. In the experimental group, the proportion at the low level fell from 26.6% to 5.1%, while the high level rose from 8.2% to 31.0%. In the control group, low-level performance fell from 27.2% to

19.6% and high-level performance rose from 7.6% to 10.1%.

When sufficient and high levels were combined, the experimental group increased from 36.7% to 76.6%, a rise of approximately 39.9 percentage points. The control group increased from 35.4% to 44.9%, a rise of approximately 9.5 percentage points.

Table 5. Distribution of students by IC level before and after the intervention

Level	Experimental pre %	Experimental post %	Control pre %	Control post %
Low	26.6	5.1	27.2	19.6
Developing	36.7	18.4	37.3	35.4
Sufficient	28.5	45.6	27.8	34.8
High	8.2	31.0	7.6	10.1

5. Discussion

The first finding is that the experimental methodology was associated with substantially stronger overall intercultural performance than the existing organisation of independent Country Studies work. This result is consistent with recent research showing that AI-supported environments can contribute to intercultural and communicative development when embedded within purposeful pedagogical designs (Sun, 2025; Hastomo et al., 2025; Ma & Shi, 2026). The present study extends that literature by showing a large multi-dimensional

effect in a structured independent-learning context rather than in a purely collaborative, conversational, or teacher-led AI intervention.

The second and more theoretically important finding is that the largest dimension-level effect occurred in critical-reflective performance. This dimension was also the weakest at baseline, suggesting that conventional Country Studies experience had provided students with substantial cultural content but fewer formal opportunities to verify claims, detect stereotyping, evaluate sources, and revise interpretations. The intervention explicitly targeted these processes through

source triangulation, AI Claim Audit activities, stereotype challenge, media analysis, and initial-versus-revised judgement tasks. The resulting gain from 14.6 to 19.8 in the experimental group, compared with 14.7 to 15.8 in the control group, suggests that verification can be taught as an intercultural practice rather than treated only as generic digital literacy.

This result has particular relevance for AI-supported education. Generative systems can produce fluent cultural explanations that appear authoritative even when context is missing or a statement is overgeneralised. The pedagogical response should therefore not be to avoid AI-generated content entirely, but to make its evaluation part of the learning process. This interpretation aligns with Hastomo et al. (2025), whose participants reported questioning and verifying AI-generated cultural information, and with broader reviews warning that AI-supported learning is most productive when human judgement remains active.

The large cognitive-cultural effect also indicates that transforming Country Studies content from topic reproduction into contextual inquiry can strengthen the use of cultural knowledge. Students were not simply asked to remember facts; they had to determine which contextual information was relevant, compare representations, and distinguish observations from interpretations. Constructive alignment is useful here because assessment tasks mirrored the kind of reasoning expected during learning (Biggs, 1996).

Behavioural-communicative development supports the value of simulated interaction and scenario-based practice. Independent learning often lacks live interlocutors, yet students still need opportunities to request clarification, respond to misunderstanding, adapt language, and make context-sensitive decisions. AI-supported scenarios can extend opportunities for rehearsal, provided that simulation is not misrepresented as an authentic embodiment of a whole culture.

The affective-attitudinal effect was slightly smaller than the other dimensions but remained large. This may reflect the difficulty of inferring private attitudes from observable task behaviour. The study therefore treated this

dimension cautiously: scores represented evidence of openness, decentring, tolerance of ambiguity, and willingness to reconsider within tasks, not a comprehensive psychological measurement of students' internal beliefs.

The shift in developmental-level distribution is pedagogically important because it shows that the intervention did not benefit only a small group of high-performing students. The sufficient-or-high proportion more than doubled in the experimental condition. At the same time, the control group also improved, which is reasonable because students continued normal instruction. The relevant comparison is therefore the stronger trajectory under the structured methodology, not an unrealistic assumption that conventional teaching should produce no change.

6. Limitations

The study has several limitations. First, intact university groups were used and students were not randomly reassigned, so unmeasured differences may remain despite non-significant baseline comparisons. Second, the reported pooled analyses assume student-level independence; fuller inference would require classroom identifiers and cluster-adjusted modelling. Third, the intervention combined task redesign, differentiated independent learning, self-regulation, teacher facilitation, platform use, and AI-supported scaffolding. The study therefore cannot identify the separate causal contribution of AI. Fourth, the assessment was a study-specific criterion-referenced performance framework. Replication should report item-level reliability, inter-rater agreement, complete rubric descriptors, and student-level longitudinal data. Finally, the participating institutions were all Uzbek EFL programmes with comparable Country Studies contexts, so broader international generalisation requires additional studies.

7. Conclusion

The findings indicate that intercultural competence can develop across cognitive, affective, behavioural, and critical-reflective dimensions when independent learning is redesigned around contextualised performance

rather than factual reproduction. The strongest outcome occurred in critical-reflective competence, suggesting that source verification, stereotype recognition, evidence evaluation, and reflective revision deserve a central place in AI-supported intercultural pedagogy.

The study also supports a restrained interpretation of AI's educational role. The observed outcomes should be attributed to the integrated pedagogical condition rather than to AI alone. Within that condition, AI was most defensible when it supplied prompts, alternative perspectives, simulated interaction, verification cues, and feedback after an initial learner attempt. The central implication is that AI-supported intercultural learning should increase the amount and quality of reasoning performed by the learner, not transfer that reasoning to the system.

References

1. Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347–364.
2. Byram, M. (1997). Teaching and Assessing Intercultural Communicative Competence. *Multilingual Matters*.
3. Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, Teaching, Assessment – Companion Volume. Council of Europe Publishing.
4. Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, 10(3), 241–266.
<https://doi.org/10.1177/1028315306287002>
5. Fantini, A. E. (2009). Assessing intercultural competence: Issues and tools. In D. K. Deardorff (Ed.), *The SAGE Handbook of Intercultural Competence* (pp. 456–476). SAGE.
6. Hastomo, T., Widiati, U., Ivone, F. M., Zen, E. L., Hasbi, M., & Khulel, B. (2025). AI-powered conversational agents and intercultural learning: Insights from Indonesian EFL students. *Intercultural Communication Education*, 8(1), 103127.
<https://doi.org/10.29140/ice.v8n1.103127>
7. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Article 248.
8. Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.
<https://doi.org/10.3389/fpsyg.2013.00863>
9. Ma, G., & Shi, Y. (2026). Toward a human-AI collaboration framework for intercultural competence development: Evidence from a mixed-methods study in Chinese EFL context. *Journal of China Computer-Assisted Language Learning*.
<https://doi.org/10.1515/jccall-2025-0032>
10. Naranjo-Toro, M., Basantes-Andrade, A., Meneses, M., & Guerra, J. (2026). Digital approaches to intercultural competence in higher education—a PRISMA systematic review with bibliometric and SWiM evidence. *Frontiers in Education*, 11, 1725080.
<https://doi.org/10.3389/feduc.2026.1725080>
11. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin.
12. Sun, L. (2025). Enhancing intercultural competence of Chinese English majors through AI-enabled Collaborative Online International Learning (COIL) in the digital era. *Education and Information Technologies*, 30, 7995–8027.
<https://doi.org/10.1007/s10639-024-13143-7>
13. Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *BMJ*, 323(7321), 1123–1124.

<https://doi.org/10.1136/bmj.323.7321.1123>

14. Zhang, S., Zou, B., Liu, Y., Chen, Z., & Yan, R. (2026). Impacts of artificial intelligence on willingness to communicate and its key antecedents in EFL contexts: A systematic review (2015–2026). *Frontiers in Education*, 11, 1927432.
<https://doi.org/10.3389/feduc.2026.1927432>