

AI- Powered Narratives as a Catalyst for Cross- Cultural Awareness and Language Performance: Evidence from a Quasi- Experimental Study in Business English

Abderrahim Khouchich,

Department of General Education, School of Economics and Management,
University Ibn Tofail, Morocco

Abstract: *Narratives are an effective teaching method for languages because they make lessons interesting and comprehensive. Teachers can help students develop emotional and cognitive connections to the subject, as well as foster cross-cultural awareness and empathy, by incorporating linguistic principles into narratives. This paper uses AI-powered narratives to explore how artificial intelligence (AI) may enhance cross-cultural awareness. The study adopts a two-group pre-test-post-test quasi-experimental design. The experimental group consists of sixty learners, master's students. The control group is composed of sixty students. Both groups are evaluated using a pre-test to determine their pre-course performance. The experimental group is taught business English using AI-powered narratives whose goal is to stimulate their cultural curiosity for 10 weeks. At the end of the experiment period, a test was assigned for both groups. The groups' scores were analyzed using T-test. Accordingly, the results unveiled that the proposed AI-powered narratives had a positive effect on the learners' English performance. There were significant differences between the means of the pre-test and post-test of the experimental group. By producing customized, contextualized scenarios that mimic real-world international conversations, essentially, AI proved to make language learning through storytelling more interactive and adaptive, bridging the gap between traditional and digital education. AI also improved motivation and engagement by customizing content to each learner's interests and proficiency level. However, narratives generated by AI and evaluated by the interrater failed to provide a fully immersive naturalistic context that can enhance critical thinking, cross-cultural awareness, and proficient communication abilities. The advantages of AI-driven language learning, possible drawbacks, and prospective future research are covered in this paper.*

Keywords: *AI, cross-cultural awareness, Language Teaching, Narratives.*

1. Introduction

Narratives emerge as a key instructional tool for promoting context-based comprehension, critical thinking, emotional involvement, and intercultural sensitivity in language instruction throughout the included works. Narratives place learning into living human experience. By integrating learning within gradually developing narrative structures, digital storytelling has been demonstrated to create safe, supported learning environments that encourage students to develop practical skills while exploring sensitive social issues, including those relevant to real-world professional contexts (Guo & Wang, 2025).

At the higher education level, narratives offer pre-service teachers an experiential learning process of creating storytelling videos and teaching packages outside of traditional classroom instruction. Digital storytelling, for instance, is now recognized as a narrative practice that uses digital tools for the creation of multimodal stories (Fathi et al., 2024).

However, studies looking at how AI-driven narratives might improve affective factors and cross-cultural awareness in EFL contexts raised concerns about the need for teacher training and learner awareness of appropriate AI tools, as well as the necessity of carefully considering assessment design and strategic implementation to mitigate potential drawbacks (AlTwijri & Musaed Alghizzi, 2024; Ebere Chukwuere, 2024).

2. Literature review

The scientific works around narrative-based language instruction and the use of AI-powered tales to improve cross-cultural awareness and associated affective and cognitive variables are introduced in this review of the literature.

2.1. AI in Language Learning

Artificial intelligence's trajectory demonstrates how rapidly this environment has changed. After being released in late 2022, it underwent several iterations, GPT-4.0 and GPT-4o, each of which expanded what teachers and students could reasonably try with AI support. If anything, the rate of adoption has surpassed the research community's capacity to thoroughly assess its long-term consequences, making ongoing empirical investigation not only beneficial but also essential.

In recent years, scholarly interest in AI in language instruction has significantly increased, surpassing early enthusiasm to develop a rich and more critical body of empirical work. Researchers have examined everything from adaptive feedback systems to AI-mediated writing help and grammar correction, reflecting a field that is actively discussing what these tools can and cannot do for language learners (Alhusaiyan, 2025). By offering scaffolding, real-time responses, and personalized content that traditional classroom settings often find challenging to consistently provide, either because of large classes or different learning abilities, AI has real potential for learner-centered instruction (Fathi et al., 2024).

2.2. AI-powered narratives in education

A branch of research has started looking more closely at what AI-powered tools can accomplish when narrative is put at the center of instruction, focusing on the larger field of AI in language education. The idea is simple: Learners exploring cross-cultural content could benefit from something genuinely unique when narrative and AI's capacity to provide adaptive, individualized engagement are combined. For a very long time, storytelling has served as a vehicle for both cultural transmission and empathetic engagement. Intercultural communicative competence is viewed as a transformative process that challenges learners' own cultural presumptions rather than merely as the acquisition of factual knowledge; this is what recent research has explored in the last decade (Xu et al., 2024).

However, there is still room for improvement in the research base. A systematic analysis of generative AI studies published between 2023 and 2024 revealed that learner perceptions dominated the study focus, and studies that directly measured language skill improvement using pre- and post-test designs remained relatively scarce. This current study claims to directly fill this gap. Similarly, the field is only beginning to transition from evaluating the efficacy of AI tools to understanding how students actually interact with AI-generated content in pedagogically designed environments. This underscores the critical need for additional classroom-based empirical research.

Reviewers have frequently pointed out the dearth of research on the pedagogical effects of AI-mediated instruction and the levels of engagement it generates in real classroom settings (Zhang & Liu, 2025). The current study takes this call seriously and employs a quasi-experimental approach that goes beyond self-reported

attitudes to ascertain whether AI-powered narratives improve EFL learners' language ability and cross-cultural awareness in measurable ways.

3. Methodology

3.1. Research design

This study looks into how AI-powered narratives affect postgraduate Moroccan EFL learners' cross-cultural awareness and language learning outcomes. The best methodological framework for this research was a quasi-experimental, two-group pre-test/post-test design. When complete randomization is not possible in intact institutional settings, quasi-experimental methods are commonly supported in applied linguistics research (Mackey & Gass, 2016). With this strategy, the researcher may preserve the ecological validity of real-world university instruction while still preserving enough scientific rigor to make accurate causal conclusions about the intervention's effects. Pre-existing class classifications were used to allocate participants to conditions, and pre-test comparison and covariate-adjusted analysis were used to establish group equivalency at baseline.

This study is also informed by interrater observation notes that provide contextual depth for the interpretation of findings. In EFL quasi-experimental research, mixed-measurement approaches that integrate standardized instruments with observation data are increasingly advised to capture the complexity of classroom language learning (Larsen-Freeman & Long, 2014).

The study's compass consists of three primary research objectives, each of which is operationalized through measurable hypotheses: (1) To ascertain whether EFL language competency outcomes are significantly improved by AI-powered narrative-based instruction as opposed to traditional lecturing. (2) To assess how learners' cross-cultural awareness and intercultural empathy are affected differently by AI-powered narrative education. And (3) To evaluate how much learners' motivational orientation, involvement, and emotional engagement with the target language are impacted by AI-generated story content.

After adjusting for pre-test performance, the related null hypotheses state that there is no statistically significant difference between the experimental and control groups on any of the aforementioned outcome characteristics at post-test. As shown in the analysis section, these are assessed using suitable inferential statistics at the $p < .05$ significant level.

3.2. Context and participants

The study was carried out in the Faculty of Economics and Management at Ibn Tofail University in Kénitra, Morocco. Leading the way in incorporating English-medium instruction into postgraduate programs in line with Morocco's Vision 2030 is Ibn Tofail University, a public university. The study was conducted as part of the master's programs, where business English is a module that equips students to communicate professionally in international settings. Students from all over the country, as well as some scholarship awardees from other African nations, make up the program's linguistically and socioeconomically varied student body.

From the accessible population of $N = 165$ students, a deliberate sample of $n = 120$ individuals representing the whole cohort was selected in order to maximize statistical representativeness and avoid gender imbalance. Participants were split into two pre-existing seminar groups of sixty students each ($n = 60$ per group) in accordance with the quasi-experimental design, keeping the classroom structures intact. The table below demonstrates the main characteristics of our participants.

TABLE I: Participant demographic profile by group.

Characteristic	Experimental group ($n = 60$)	Control group ($n = 60$)	Total ($N = 120$)
Gender (female)	38 (63.3%)	35 (58.3%)	73 (60.8%)
Mean age (SD)	23.2 (1.6)	23.6 (1.8)	23.4 (1.7)

4. Data Analysis

Three sets of SPSS statistical outputs are presented to examine the effects of AI-powered narrative instruction on language proficiency and cross-cultural awareness: descriptive statistics for both groups' pre-test performance, an independent samples t-test to compare the post-test linguistic gains and cross-cultural awareness scores of the experimental and control groups, and a paired samples t-test to assess within-group growth between pre-test and post-test for the experimental group. The findings are evaluated in light of the three previously stated research objectives and placed within the theoretical framework of cross-cultural awareness development in EFL contexts and AI-powered narrative pedagogy.

4.1. Linguistic gains

Prior to running inferential tests, data were screened for normality using the Shapiro-Wilk test to verify the homogeneity of variance assumption, and variables were found to be normally distributed. Table 2 presents the descriptive statistics for pre-test scores in both groups (N = 60 per group).

TABLE II: One-sample t-test results for pre-test language proficiency scores against test value = 10

	t	df	Sig. (2-tailed)	Mean difference	95% CI lower	95% CI upper
Pre-test — Experimental group	6.130	59	.000	1.567	1.055	2.078
Pre-test — Control group	4.832	59	.000	1.700	.996	2.404

Table 2 indicates that the experimental group had a pre-test mean of $M = 11.57$ ($SD = 1.98$, $SE = .26$) when they started the study, whereas the control group had a slightly higher mean of $M = 11.70$ ($SD = 2.73$, $SE = .35$). Both groups had very similar levels of language proficiency at the beginning of the ten-week intervention, as seen by the little mean difference between them at baseline ($\Delta M = 0.13$).

This study's main question is whether the experimental group's ten-week AI-powered narrative intervention resulted in quantifiable change within-group progress in comparison to the control group. For each group, a paired-samples t-test was used to determine whether there was a statistically significant difference between the pre-test and post-test scores. The direction of the mean difference was coded using pre-test scores minus post-test scores; a negative mean difference indicates progress. The results are displayed in Table 3 below.

TABLE III: Paired Samples t-test results: pre-test to post-test differences within the experimental and control groups.

Pair	Mean diff.	Std. deviation	Std. error mean	95% CI lower	95% CI upper	t	df
Pair 1: Pre-test Exp – Post-test Exp	-1.258	2.823	.364	-1.988	-.529	-3.453	59
Pair 2: Pre-test Ctrl – Post-test Ctrl	-.375	3.416	.441	-1.257	.507	-.850	59

The paired-samples t-test for the experimental group produced a mean difference of -1.258 ($SD = 2.823$,

$SE = .364$), with a 95% confidence interval of $[-1.988, -.529]$. The t-statistic was $t(59) = -3.453$, $p = .001$ (two-tailed). Students who received AI-powered narrative instruction performed better on the language proficiency test, as evidenced by the negative mean difference, an average of 1.258 score points over the ten-week intervention period.

In the context of a 20-point proficiency scale, this result is meaningful and statistically significant at the $p < .01$ level. The observed improvement is substantial and unlikely to be due to measurement error or random variation because the confidence interval completely excludes zero. Cohen's d for this within-group effect may be calculated by dividing the mean difference by the difference's standard deviation: $d = 1.258 / 2.823 = .45$, which indicates a medium effect size (Cohen, 1988). This effect is particularly noteworthy given the ten-week

duration of the intervention, a period that is relatively short for detecting measurable gains in EFL proficiency at the postgraduate level.

For the control group, the paired-samples t-test produced a mean difference of $-.375$ ($SD = 3.416$, $SE = .441$), with a 95% confidence interval of $[-1.257, .507]$. The t-statistic was $t(59) = -.850$, $p = .399$ (two-tailed). This result is not statistically significant at any conventional threshold ($p > .05$), and the confidence interval crosses zero, indicating that the pre-to-post difference in the control group cannot be attributed to a meaningful change.

4.2. Cross-Cultural Awareness Gains

A complimentary layer of evidence is provided by the study's cross-cultural awareness dimension, which was assessed using an in-class observation grid, though the SPSS output above emphasizes language competency as the main quantitative outcome. The previously reported improvements in competency took place in a learning environment specifically created to include cultural content. Through the use of Claude-generated tales, Moroccan and foreign business actors were placed in cross-cultural conflict and negotiation scenarios, necessitating the interpretation of face-saving techniques, indirect communication, and culturally different professional conventions.

Therefore, the statistically significant improvement in the experimental group's language scores cannot be separated from the multicultural architecture of the intervention. At the advanced EFL level, language and cultural learning are inseparable processes, according to the narrative pedagogy literature (Byram, 2021). The AI-generated narratives included in this study were specifically created to take advantage of this mediation process. They were observed to increase cross-cultural awareness over a ten-week period.

Nevertheless, the study cannot assert that linguistic competence increase in this case is adequate to completely bridge the gap between classroom EFL learning and immersive naturalistic acquisition. The motivational investment of a researcher-teacher who designs, delivers, and evaluates an instructional intervention within their own classroom may translate into measurable outcome inflation (Zemmerman & Marcus, 2024). Being aware of this 'Researcher Allegiance' made the call for an external interrater observer a necessary step.

TABLE IV: AI narrative performance evaluation (interrater reliability)

Narrative element / AI tool feature	Worked well?	Needs adjustment?	Specific observation / recommended revision
Narrative cultural context (appropriateness and authenticity)	☐ Yes ☒ No	☒ Yes ☐ No	The narrative cultural context did not feel sufficiently authentic. It seemed artificial and should be revised with more realistic sociocultural cues, natural interactional details, and a clearer link to a believable communicative situation.
Language level / CEFR difficulty (B2–C1 fit)	☒ Yes ☐ No	☐ Yes ☒ No	The linguistic difficulty broadly corresponds to a B2–C1 level. However, the wording sounds too scripted and not fully natural. It should be reformulated with more idiomatic expressions, smoother transitions, and less artificial sentence structures.
Cultural conflict scenario (clarity and pedagogical value)	☒ Yes ☐ No	☒ Yes ☐ No	The cultural conflict scenario is understandable and can support discussion, but its pedagogical value would be stronger if the conflict were more authentic, less predictable, and closer to real intercultural communication situations.
HeyGen video narration (pace, intelligibility, visual quality)	☐ Yes ☒ No	☒ Yes ☐ No	The HeyGen narration/video does not sound fully authentic. The language delivery feels unnatural and scripted. The video should be revised to improve naturalness, oral rhythm, and the credibility of the speaker's performance.
Learner response to AI-generated format (engagement, novelty effect)	☒ Yes ☐ No	☒ Yes ☐ No	The AI-generated format attracts learners' attention because of its novelty, but the lack of authenticity limits deeper engagement. A more natural scenario and narration would likely improve learner involvement and discussion quality.

The interrater found that the linguistic calibration of the narratives and the fundamental clarity of the cultural conflict situations were both adequate. However, he described the authenticity of the cultural context as artificial rather than rooted in believable sociocultural and interactional detail; the HeyGen video narration's naturalness

was also described as scripted and lacking credible oral rhythm; and, as a downstream consequence of both, the depth of learner engagement, which the observer noted was elevated by the novelty of the AI-generated format but constrained by its perceived lack of authenticity.

5. Discussion & Conclusion

After a ten-week postgraduate instruction, the experimental group's improvement suggests a medium impact size. This result is interesting for two reasons. First, it is generally believed that EFL proficiency gains at the B2–C1 level are more challenging to attain than at lower proficiency levels because learners have already solidified the language's fundamental grammatical and lexical architecture, and further development depends more and more on exposure to complex, authentic, and contextually varied input (Dörnyei, 2009). Second, ten weeks is a somewhat short time frame for identifying quantifiable group-level improvements on a standardized competency test. The quality and density of understandable, contextually rich language input that the Claude-generated narratives offered proved to be the most likely mechanism connecting AI narrative instruction to proficiency gains. This outcome converges with recent EFL research, which has specifically shown that multimodal digital storytelling formats perform better on vocabulary retention and listening comprehension outcomes than text-only equivalents (Wei & Hu, 2025).

The study's AI narratives were specifically created to test students' ability to read, analyze, and ultimately resolve intercultural conflict scenarios including professional Moroccan, Chinese, Japanese, French, and British people (current and prospective investors in Morocco). In each situation, students had to question not just what happened but also why, and they had to base their responses on cultural knowledge rather than intuition. However, AI-powered narratives cannot serve as a complete alternative for traditional instruction, particularly in cultures like Morocco, where human direct contact is key to intercultural empathy. Yet, the study does support a more modest but empirically grounded claim: AI-generated narrative materials consistently yield higher language proficiency gains than the traditional teaching approach within a typical semester-long instructional block.

To conclude, the challenge facing EFL education is not whether or not to employ AI tools, but rather how to do so in a way that is rigorous, culturally sensitive, and genuinely focused on producing students who can successfully negotiate the intercultural complexity of the world with empathy.

6. Acknowledgements

The author would like to thank the president of Ibn Tofail University and the dean of the School of Economics and Management for their logistical and institutional support during this study. I would also like to thank Professor Azelmad Nouredine for kindly acting as the inter-rater evaluator. Lastly, the author would like to express gratitude to the students who generously and voluntarily took part in this study.

7. References

- [1] Y. Guo and Y. Wang, "Exploring the effects of artificial intelligence application on EFL students' academic engagement and emotional experiences: A mixed-methods study," *Eur. J. Educ.*, vol. 60, e12812, 2025. <https://doi.org/10.1111/ejed.12812>
- [2] J. Fathi, M. Rahimi, and A. Derakhshan, "Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions," *System*, vol. 121, 103254, 2024. <https://doi.org/10.1016/j.system.2024.103254>
- [3] L. AlTwijri and T. Musaed Alghizzi. (2024). Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review. DOI: 10.1016/j.heliyon.2024.e31053 <https://doi.org/10.1016/j.heliyon.2024.e31053>
- [4] J. Ebere Chukwuere, "The future of generative AI chatbots in higher education," 2024. <https://doi.org/10.32388/UE841K>

- [5] S. Lee, H. Choe, D. Zou, and J. Jeon, "Generative AI (GenAI) in the language classroom: A systematic review," *Interact. Learn. Environ.*, advance online publication, 2025.
<https://doi.org/10.1080/10494820.2025.2498537>
- [6] E. Alhusaiyan, "A systematic review of current trends in artificial intelligence in foreign language learning," *Saudi J. Lang. Stud.*, vol. 5, no. 1, pp. 1–16, 2025.
<https://doi.org/10.1108/SJLS-07-2024-0039>
- [7] Z. Xu et al., "A systematic review of cross-cultural communicative competence in EFL teaching: insights from China," *Humanit. Soc. Sci. Commun.*, vol. 11, 1750, 2024.
<https://doi.org/10.1057/s41599-024-04071-5>
- [8] Y. Zhang and X. Liu, "Learner emotions in AI-assisted English as a second/foreign language learning: a systematic review of empirical studies," *Front. Psychol.*, vol. 16, 1652806, 2025. <https://doi.org/10.3389/fpsyg.2025.1652806>
- [9] A. Mackey and S. M. Gass, *Second Language Research: Methodology and Design*, 2nd ed. New York, NY: Routledge, 2016.
- [10] D. Larsen-Freeman and M. H. Long, *An Introduction to Second Language Acquisition Research*. Abingdon, U.K.: Routledge, 2014.
<https://doi.org/10.4324/9781315835891>
- [11] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
- [12] M. Byram, *Teaching and Assessing Intercultural Communicative Competence: Revisited*, 1st ed. Bristol, U.K.: Multilingual Matters, 2021.
<https://doi.org/10.21832/9781800410251>
- [13] J. Zimmerman and D. Marcus, "Does Researcher Allegiance Bias Outcomes in Psychotherapy Research? A Quasi-Experimental Secondary Analysis," *Clin. Psychol. Psychother.*, vol. 31, e2974, 2024.
<https://doi.org/10.1002/cpp.2974>
- [14] Z. Dörnyei, *The Psychology of Second Language Acquisition*. Oxford, U.K.: Oxford University Press, 2009.
- [15] R. Wei and D. Hu, "Incidental vocabulary learning through reading while listening in the Chinese EFL classroom context: A longitudinal study," *SAGE Open*, vol. 15, no. 2, 2025.
<https://doi.org/10.1177/21582440251334243>