

Does VR Benefit Speaking Proficiency? An Experimental Study

Manar Baabbad

Princess Nourah bint Abdulrahman University

Background

Speaking competence refers to the ability to produce fluent, accurate, and appropriate spoken language, utilizing various knowledge, skills, and strategies (Goh & Burns, 2012). Speaking can be developed if L2 learners have opportunities to interact with others and are pushed to produce the target language (Long, 1996; Swain, 2005). Involving L2 learners in such cases helps activate their attention and mental linguistic resources, which gradually promotes the automaticity of the language production process (Long, 1996; Swain, 2005). However, the flow of spoken production can be hindered by constraints, including linguistic variables (Goh & Burns, 2012) or unwillingness to communicate (MacIntyre et al., 1998).

Previous studies have discussed the essentiality of willingness to communicate (WTC) as a prerequisite for speaking development and interaction (MacIntyre et al., 1998; Kang, 2005; Peng, 2022). WTC is “readiness to enter into discourse at a particular time with a specific person or persons, using L2” (MacIntyre et al., 1998, p. 547). However, WTC is affected by several contextual factors (Kang, 2005; Peng, 2022). Kang (2005) suggested that providing a safe environment that obtains security (i.e., lack of fear or punishment), excitement (i.e., entails feelings of joy to talk), and responsibility (i.e., a sense of duty to understand or deliver the message) can likely increase the WTC in the classroom.

Recent studies have demonstrated that virtual reality (VR) can enhance speaking proficiency and promote a sense of security due to the absence of negative evaluation and judgment, as well as anonymity (Ebadi & Ebadijalal, 2022; Chen, 2021; Dhimolea et al., 2022). Furthermore, learners become more engaged in VR due to the immersive language use experience (Ebadi &

E-mail: mkbaabbaad@pnu.edu.sa

Ebadijalal, 2022; Dhimolea et al., 2022). However, Dhimolea et al. (2022) found contradictory findings among 32 studies regarding language gain, including speaking.

The inconclusive findings may be attributed to research variables, including the study's duration, experimental design, and type of VR. Additionally, little is known about the instructional approaches used in previous studies, which may have affected the studies' conclusions. According to Chen et al. (2021), integrating VR with an appropriate instructional approach is a significant step that should be considered to evaluate the effectiveness of VR. Although the research agenda emphasizes integrating Task-based language teaching (TBLT) with VR due to their affordances (i.e., authenticity and interactivity) (Chong & Reinders, 2020), further research is needed to examine the impact of integrating technology with TBLT on language development and WTC. To fill this gap, this study aims to investigate the following questions:

1. To what extent does VR-mediated TBLT promote the speaking proficiency of Arabic language learners when compared with face-to-face TBLT?
2. To what extent does VR-mediated TBLT promote the situated L2 WTC of Arabic language learners when compared with face-to-face TBLT?
3. What contextual factors appear to impact the learners' situated L2 WTC in each context?

Methodology

A quasi experimental study using a mixed-methods approach was conducted, involving two intact classes ($n_1 = 13$ and $n_2 = 14$) assigned randomly. The participants were female students from diverse countries. They were studying Arabic as a second language at a tertiary level. Their ages ranged from 18 to 35 years. No one had used VR before this study.

Both groups (VR and non-VR) were taught by the same instructor and performed the same task (reasoning or opinion gap tasks), in 2-hour sessions each week. Fifteen tasks were designed based on TBLT criteria (Ellis, 2003) and aligned with the course syllabus (i.e., topics, levels and language knowledge). The study implemented high-immersive (i.e., using head-mounted devices) and multi-user VR that allows synchronous interaction. The VR group was trained (over 8 hours, distributed across different days) on how to use the technology after the pre-test and before starting the intervention. The study lasted five months, including the testing period.

The participants performed the OPI test (i.e., Oral Proficiency Interview, a reliable, valid, and proficiency-based standard test developed by ACTFL) before and after the intervention, with two proficient examiners (see Table 1).

Table 1. Descriptive results of the three tests for each group

	Pretest			Posttest		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
VR group	3.92	.954	13	6.31	1.182	13
Non-VR group	4.07	1.072	14	6.21	1.424	14

The Mann-Whitney test was used, since the assumption of normality was rejected, to compare the means of speaking proficiency.

All the sessions were recorded and analyzed using an observation scheme adopted from Al-Murtadha (2019) and Coa (2014). After transcribing the recordings, the WTC indicators (i.e., voluntarily answering, asking, declaring, or responding to an opinion, and explaining) for each participant were identified and then calculated to determine the sum of WTC indicators in each group when performing the same task. The word quantity produced by each participant was also considered in this stage. The means of WTC and word quantity were compared using the Mann-Whitney test.

Journals were also collected immediately after each session to measure WTC. The journals (336 journals) were analyzed to identify the high and low WTC moments reported by the participants, along with the associated factors. The factors were classified into three categories: task variables, VR variables, and other variables (i.e., those not related to the task design or technology features).

Results and Discussion

The findings showed that the VR group improved their speaking, but no statistically significant difference ($U = 80.500$, $z = -.526$, $p = .599$) with a very small effect size ($r = 0.01$). This finding aligns with Nicolaidou et al. (2023), who demonstrated a similarity between their study groups (VR vs. non-VR groups) in terms of vocabulary improvement. On the other hand, this finding contrasts with Ebadi and Ebadijalal (2022), who found that the VR group outperformed the control group in their oral skills.

The inconsistent findings may be attributed to two key reasons. First, while this study employed a holistic test that provides an overall assessment of speaking proficiency, Ebadi and Ebadijalal (2022) used an analytical test that assigns a score to the spoken language components, encompassing fluency, vocabulary, pronunciation, and grammar. The improvement of these components can be observed in a short period, compared to developing speaking as a whole, which

requires a longer time. Different types of tests draw different conclusions (Luoma, 2004).

The second key point is the novelty of the instructional approach experienced by the non-VR group in this study. TBLT was not a regular approach used in their classes before conducting the study, as the participants in the non-VR group reported, which possibly affected their outcome. In contrast, the non-VR group in Ebadi and Ebadijalal's (2022) study typically gave presentations about different locations, a common practice in language classrooms.

The findings also showed no statistically significant difference between the groups regarding WTC indicators and word quantity. This is demonstrated by the results for WTC indicators ($U = 65.50$, $z = -1.239$, $p = .215$, $r = .05$) and word quantity ($U = 66.50$, $z = -1.189$, $p = .234$, $r = .05$), with a small effect size. The finding suggests that the situated WTC of both groups were similar. It is worth noting that this study lost some data (only 11 sessions were included, lasting approximately 15 minutes each), which might have uncovered different findings, due to technical errors and participant absences. This is considered a limitation in the study.

Although the quantitative findings suggest that VR may not offer benefits to language development and WTC compared to the classroom, the qualitative analysis revealed some positive results. First, the findings show that VR enhanced security conditions, as learners used avatars that hid their real personalities, also increased their feeling of social presence. This is supported by Chen (2021), who found that masked identities, represented as avatars, heightened learners' confidence. Second, integrating TBLT, as an instructional approach, with VR kept learners' engagement, which enhanced the second physiological condition (i.e. excitement) (Kang, 2005).

However, the qualitative results show that the fluctuation of WTC was more dynamic, especially at the beginning of the intervention, in the VR group than in the non-VR group, due to physical (i.e., headache and eye strain) and technical issues. This finding suggests that the non-significant results may have been affected by the novel features of VR, which could negatively impact the learners' outcomes. Despite the participants being trained, it seems that reaching normalization requires a considerable amount of time. Consequently, it is vital to move VR toward normalization, allowing for reliable examination of the potential benefits of the technology (Bax, 2011). Additionally, the small scale of the study (i.e., small sample size and short duration) might hinder the discovery of potential positive results related to speaking development. Therefore, further research is necessary in a large-scale setting.

References

- Al-Murtadha, M. A. (2019). Enhancing EFL learners' willingness to communicate with visualization and goal-setting activities. *TESOL Quarterly*, 53(1), 133–157. <https://doi.org/10.1002/tesq.474>
- Bax, S. (2011). Normalisation revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching*, 1(2), 1–15. <https://doi.org/10.4018/ijcallt.2011040101>
- Cao, Y. K. (2014). A socio-cognitive perspective on second language classroom willingness to communicate. *TESOL Quarterly*, 48(4), 789–814. <https://doi.org/10.1002/tesq.155>
- Chen, J. C. C. (2021). The interplay of avatar identities, self-efficacy, and language practices. *Australian Review of Applied Linguistics*, 44(1), 65–81. <https://doi.org/10.1075/aryl.19032.che>
- Chen, C. H., Hung, H. T., & Yeh, H. C. (2021). Virtual reality in problem-based learning contexts: Effects on the problem-solving performance, vocabulary acquisition and motivation of English language learners. *Journal of Computer Assisted Learning*, 37(3), 851–860. <https://doi.org/10.1111/jcal.12528>
- Chong, S. W., & Reinders, H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning & Technology*, 24, 70–86. <http://hdl.handle.net/10125/44739>
- Goh, C., & Burns, A. (2012). *Teaching speaking: A holistic approach*. Cambridge University Press.
- Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends*, 66(5), 810–824. <https://doi.org/10.1007/s11528-022-00717-w>
- Ebadi, S., & Ebadijalal, M. (2022). The effect of Google Expeditions virtual reality on EFL learners' willingness to communicate and oral proficiency. *Computer Assisted Language Learning*, 35(8), 1975–2000. <https://doi.org/10.1080/09588221.2020.1854311>
- Ellis, R. (2003). *Task-based language teaching and learning*. Oxford University Press.
- Kang, S. J. (2005). Dynamic emergence of situational willingness to communicate in a second language. *System*, 33(2), 277–292. <https://doi.org/10.1016/j.system.2004.10.004>
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
- MacIntyre, P. D., Dörnyei, Z., Clément, R., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562. <http://dx.doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- Nicolaidou, I., Pissas, P., & Boglou, D. (2023). Comparing immersive virtual reality to mobile applications in foreign language learning in higher education: A quasi experiment. *Interactive Learning Environments*, 31(4), 2001–2015. <https://doi.org/10.1080/10494820.2020.1870504>
- Peng, J. (2022). Willingness to communicate. In S. Li, P. Hiver & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 159–171). Routledge.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Routledge.