Virtual Reality Serious Games and Language Learning: A Quasi-Experimental Study in an Ivorian School Setting

Saho Venance Zoh and Konan Yao
Université Virtuelle de Côte d'Ivoire (UVCI), Ivory Coast

Saho.zoh@uvci.edu.ci

Abstract: Background and objective : A lack of motivation towards traditional lecture-based teaching methods remains a persistent obstacle in African education systems. This study aims to assess the impact of a virtual reality (VR) serious game, enhanced with adaptive artificial intelligence, on the foreign language performance of Ivorian students. Method : A quasi-experimental single-group study (pre-test–post-test) was conducted with 68 secondary school pupils (in the second and first years of the literary stream, aged 15–17) at the Collège Moderne Mami Faitai in Abidjan. Language acquisition was measured using a validated test (Cronbach's $\alpha = 0.78$) and self-assessment (5-point Likert scale). Reaction (Level 1) and behavioural transfer (Level 3) were assessed according to the Kirkpatrick model. Pre-test–post-test differences were analysed using the Wilcoxon test for paired samples ($\alpha = 0.05$). Results: The median score rose from 52% in the pre-test to 67% in the post-test ($W = 189$, $p < 0.001$, $d = 0.75$ [95% CI: 0.47–1.02]). Overall satisfaction stood at 4.45/5 (≈84%). Self-assessment of learning outcomes (Level 2) revealed more marked perceived gains in speaking than in writing. Behavioural transfer (Level 3) presents a mixed picture: motivation to continue is high (CE4 = 4.57/5) but frequency of use in real-life situations remains modest (FU3–FU5: 2.33–2.84/5). Conclusion : These results suggest that VR immersion combined with adaptive AI promotes language acquisition and fosters high engagement. However, the gap between immediate satisfaction and actual transfer calls for further investigation into the conditions for the sustainable use of these tools.

Keywords: Serious games, Virtual reality, Language learning, Motivation, Quasi-experimental study, Ivorian context

1. Introduction

The digital transformation of education systems is opening up new pedagogical possibilities. Serious games, defined as educational applications designed to serve a practical purpose (Djaouti, 2011), are increasingly being used as active learning methods aimed at boosting motivation and pedagogical effectiveness. Several meta-analyses indicate that, on average, serious games improve learners' outcomes compared to traditional methods (Connolly et al., 2012 ; Hamari et al., 2016), although their direct effect on intrinsic motivation remains mixed.

Virtual reality (VR) is defined as an immersive digital environment that facilitates sensory-motor and cognitive activities (Slater & Wilbur, 1997). It offers a rich framework for language learning by enhancing contextual interaction and the learner's sense of presence. In the Ivorian context, where traditional lecture-based methods struggle to foster sustained engagement, the integration of these technologies represents a potential lever for educational transformation.

The research question is as follows: 'Does an immersive VR serious game, equipped with adaptive artificial intelligence, improve the language skills of Ivorian secondary school pupils, and which skills are most enhanced? ' In line with the active learning framework (Bandura, 1997) and self-determination theory (Deci & Ryan, 2000), we hypothesise that this immersive environment will yield measurable linguistic gains, particularly in oral skills, high levels of satisfaction, and motivation to continue using it, whilst presenting limitations in transfer to authentic situations.

2. Theoretical Framework

2.1 Multimedia Learning and Virtual Reality

According to Mayer's (2021) Cognitive Theory of Multimedia Learning (CTML), individuals learn more effectively when information is presented in a multimodal manner, simultaneously engaging both visual and auditory channels (dual coding). This model is based on three principles : (a) dual coding, which promotes long-term retention; (b) the limited capacity of the cognitive system, which requires optimal load management; and (c) active processing, which involves the learner's engagement in the selection and integration of information. By combining three-dimensional visual stimuli with contextualised auditory interactions, VR creates an ideal environment for this type of learning. In particular, it enhances speaking skills by immersing the learner in simulated dialogues with conversational agents, which is consistent with the self-assessment results observed in this study.

2.2 Serious Games: Definition and Educational Potential

Serious games, as conceptualised by Djaouti (2011), refer to game-based applications designed with a practical purpose that goes beyond mere entertainment. They combine an explicit educational aim with an engaging game structure, thereby bringing together enjoyment and learning. This approach is consistent with the principles of game-based learning (Gee, 2003) and draws on the principle of learning by doing (Schank, 1999). The integration of VR amplifies this potential by generating a sense of presence defined as the sensation of being 'in' the environment (Slater & Wilbur, 1997) which enhances long-term memory. Systematic reviews document the positive effects of educational VR on learning outcomes and creativity.

2.3 Motivation and Engagement

Deci and Ryan's (2000) Self-Determination Theory (SDT) posits that intrinsic motivation is based on the fulfilment of three needs: autonomy, competence (reinforced by immediate feedback, Bandura, 1997) and social connection. Immersive serious games simultaneously meet these needs. Csikszentmihalyi (1990) complements this perspective with the concept of flow, a state of total immersion generated when the challenge matches the learner's skills. Hamari et al. (2016) and Connolly et al. (2012) confirm that serious games increase emotional and behavioural engagement. However, as several authors point out, this reported engagement does not automatically translate into autonomous usage behaviours outside the gaming environment a tension documented by our own Level 3 results.

2.4 Adaptive Artificial Intelligence

Adaptive artificial intelligence (AI) represents a breakthrough in educational design. Through reinforcement learning (Sutton & Barto, 2018), these systems dynamically adjust the level of difficulty, feedback and content based on the user's performance. Conati and Kardan (2013) show that adaptive AI infers the learner's emotional and cognitive state to offer personalised interventions, reducing extrinsic cognitive load (Sweller, 2011) and placing the learner within their zone of proximal development (Vygotsky, 1978). Applications such as Mondly VR illustrate this approach. Our study provides a nuanced perspective: whilst real-time AI feedback is well received (AI3 – Useful feedback = 4.32/5), level adaptation is the lowest-rated aspect of satisfaction (EG6 = 3.91/5), suggesting room for algorithmic improvement.



Figure 1: Adaptive serious game system in VR

3. Methodology

3.1 Study Design

This study employs a quasi-experimental, single-group design with repeated measures, following a pre-test–post-test approach. The evaluation framework is based on Kirkpatrick's (2006) four-level model, which allows for a sequential assessment of learners' reactions, learning outcomes, observable behaviors, and the effects of the training. The study took place at Collège Moderne Mami Faitai in Abidjan between February and April 2025,

or approximately (3 months), involving learners who participated in virtual reality (VR) language learning activities.

The experiment utilized Mondly Languages VR, a language-learning app designed for virtual reality. The program places learners in virtual environments that simulate everyday communication situations, where they are encouraged to interact verbally with virtual characters and complete various language activities. Learning is thus based on a combination of immersion, interaction, and role-playing, allowing learners to apply the vocabulary and expressions they have studied in simulated conversational contexts. The activities are organized around situations such as social interactions, getting around, and interactions in various everyday settings.

Particular attention is paid to the adaptive nature of the system. Artificial intelligence plays a key role in conversational interactions and in processing the learner's responses, enabling the system to react to the learner's input and suggest activities or content based on their learning path. This adaptive aspect distinguishes the system under study from a simple immersive environment: virtual reality provides the spatial and interactional context, while artificial intelligence mechanisms help personalize the learning experience based on the user's interactions. These characteristics are an important factor in interpreting the results related to motivation, engagement, and learning.

The protocol followed a pre-test–intervention–post-test design to observe changes in the indicators studied after using the system. The absence of a control group constitutes a methodological limitation of this quasi-experimental design. This limitation is taken into account in the interpretation of the results and discussed in Section 5.

3.2 Participants

The sample comprises 68 pupils in the first and second years of lower secondary school (mean age = 15.5 years, SD = 0.9; 47% girls). Sample sizes vary slightly depending on the level of analysis (n = 65 for Level 2, n = 66 for Level 1, n = 64 for Level 3) due to missing data on some questionnaires. Recruitment was convenience-based (whole classes). Consent was obtained from parents and the school management prior to the study.

3.3 Measurement Tools

Level 1: Reaction. A questionnaire comprising 14 items across three sub-dimensions (Overall Experience, EG; Feedback, IA; Pedagogical Aspects, AP) using a 5-point Likert scale was used to measure immediate satisfaction.

Level 2 : Language proficiency. Two complementary instruments were used. An objective test (30 items : contextual vocabulary and functional grammar, aligned with CEFR levels A2–B1, $\alpha = 0.78$, validated by two expert teachers) measured observable performance. A 25-item self-assessment questionnaire across 5 dimensions – Self-Assessment (SA), Reading Comprehension (RC), Listening Comprehension (LC), Writing (W), Speaking (S) – supplemented by 5 items specific to the serious game (SG), captured the students' own perceptions of their gains.

Level 3: Transfer behaviour. A 12-item questionnaire covering three sub-dimensions (Frequency of Use FU, Confidence/Engagement CE, Real-Life Situations SR) was administered two weeks after the intervention. This measure is self-reported and should be interpreted with caution.

3.4 Statistical Analyses

Pre-test–post-test differences were analysed using the Wilcoxon test for paired samples ($\alpha = 0.05$). The effect size was estimated using Cohen's d with a 95% CI calculated via bootstrapping (1,000 resamples). For each learner, the gain is defined as $D_i = X_i^{post} - X_i^{pre}$. The composite Likert scores were aggregated by dimension. The analyses were carried out using R (version 4.3).

4. Results

4.1 Participants' Responses- Level 1

Overall satisfaction is very high (M = 4.45/5, SD = 0.38, or ≈84% after standardisation). Figure 2 details the 14 items by sub-dimension. The item 'Enjoyable gameplay' (EG1 = 4.62) received the highest score, reflecting the hedonic engagement of the experience. The AI feedback was well received (IA3 – Useful feedback = 4.32; IA1 – Adaptive AI = 4.29). Conversely, the item 'Appropriate difficulty level' (EG6 = 3.91) is the only one not to reach the 4/5 agreement threshold, indicating an imperfection in the dynamic adaptation of the difficulty level, which warrants further investigation in future iterations of the system.

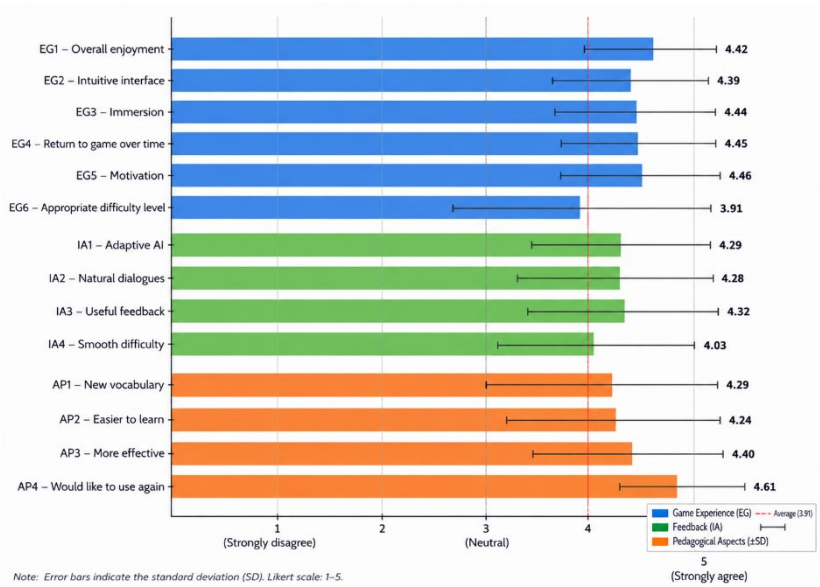


Figure 2: Means ± standard deviation by item – Level 1: Reaction (n = 66)

Note. EG = Overall Experience ; IA = IA Feedback; AP = Pedagogical Aspects. Error bars = standard deviation (SD). Likert scale 1–5.

4.2 Language Skills – Level 2

Table 1 presents the inferential statistics from the objective test. A median gain of 15 percentage points is observed (52% → 67%), which is statistically significant and represents a moderate-to-large effect size.

Table 1: Comparison of pre-test and post-test language scores (N = 68)

	Pretest	Post-test	Difference
Médiane (%)	52	67	+15
EIQ (%)	18	16	—
Wilcoxon W	—	—	189
p-value	—	—	< 0,001
d de Cohen	—	—	0,75
IC 95 % (d)	—	—	[0,47 ; 1,02]

Note. IQR = interquartile range. 95% CI calculated using bootstrapping (1,000 resamples). d = 0.75 = moderate-to-large effect (Cohen, 1988).

The self-assessment of learning outcomes (Figure 2) reveals a varied profile across different skills. Items relating to listening comprehension (CO3 – Identifying the context = 4.25; CO2 – Recognising key words = 4.09) and speaking (EO1 – Introducing oneself without hesitation = 4.26) achieved the highest scores. Conversely, the written expression items are consistently lower (EE3 – Correct conjugation = 3.35; EE1 – Writing a sentence without help = 3.38), suggesting that the VR system favours oral skills more than written skills.

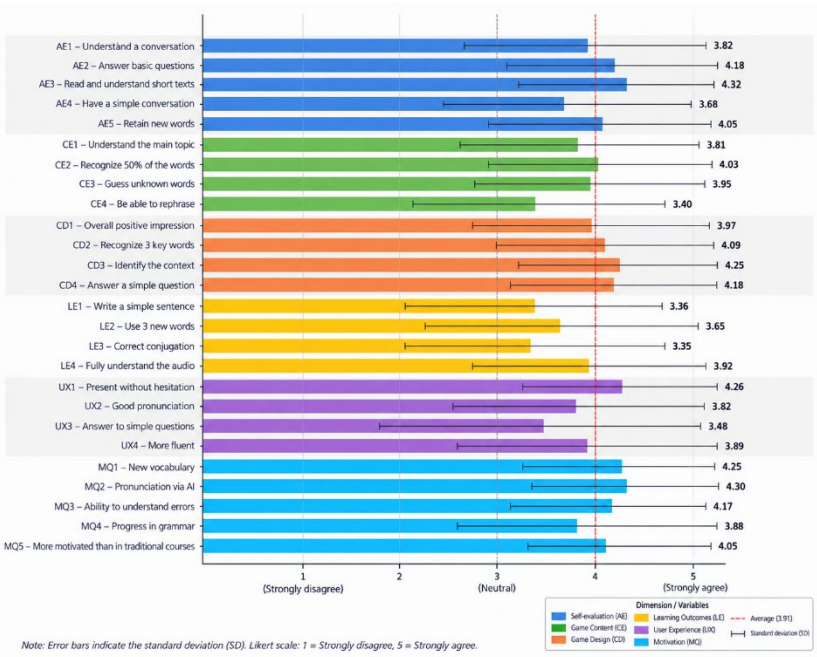


Figure 3: Means ± standard deviation of self-assessment scores by item – Level 2 (n = 65)

Note. SA = Self-Assessment; RC = Reading Comprehension; LC = Listening Comprehension; WE = Written Expression ; LE = Spoken Expression; SG = Serious Game. Error bars = standard deviation (SD). Likert scale 1–5.

The radar plot (Figure 3) confirms this asymmetry: Oral Comprehension (M = 4.12, SD = 0.78) and the Serious Game as a motivational tool (M = 4.19, SD = 0.74) dominate, whilst Written Expression has the lowest score (M = 3.58, SD = 0.93) and the greatest individual variability.

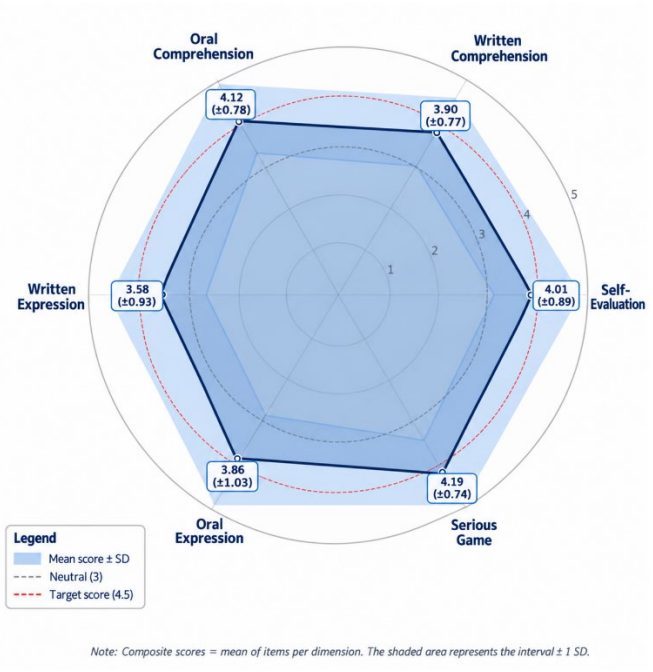


Figure 4: Radar plot of mean scores by dimension – Level 2 (n = 65)

Note: Composite scores = mean of items per dimension. The coloured band represents the ±1 SD interval. Agreement threshold (4/5) shown as a red dotted line.

4.3 Transfer of Learning – Level 3

The behavioural transfer profile (Figure 4) is mixed. The Confidence/Engagement sub-dimension shows the highest scores : CE4 – Motivation to continue the serious game scores 4.57/5, the highest score in the entire

Level 3 questionnaire, and CE2 – Thinking in the target language = 3.52. In contrast, the Frequency of Use sub-dimension reveals modest scores: FU3 – Conversation outside class (M = 2.33), FU4 – Watching videos in the original language (M = 2.36) and FU5 – Everyday use (M = 2.84) all remain below the neutral score of 3/5. The Real-Life Situations sub-dimension is also low (RS1 – Talking to a classmate = 2.35; RS3 – Games/films in the original language = 2.60).

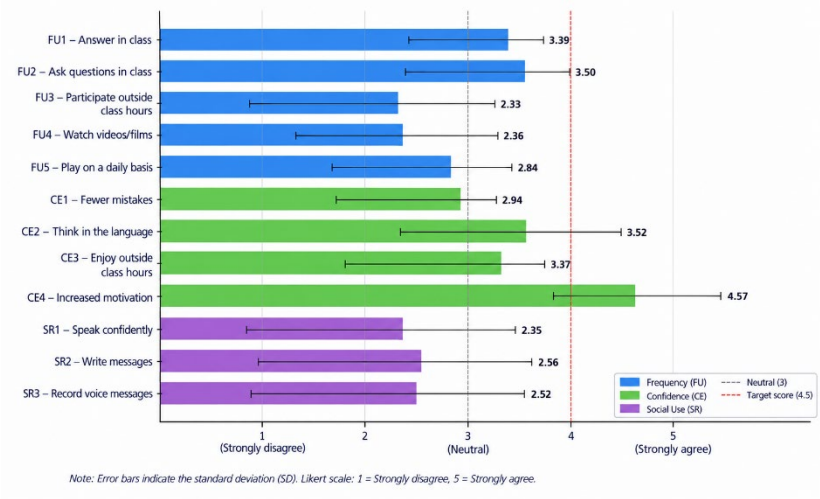


Figure 5: Means ± standard deviation by item – Level 3: Transfer behaviour (n = 64)

Note. FU = Frequency of Use ; CE = Confidence/Commitment ; SR = Real-life situations. Error bars = standard deviation (SD). Frequency/Confidence 1–5.

The radar chart (Figure 5) summarises all the dimensions. Levels 1 and 3. It illustrates the gap between satisfaction (Level 1 : scores 4.23–4.41) and observed be3 :viours (Level 3: Frequency of Use = 2.88; Real-life Situations = 2.49). Only the Trust/Commitment dimension at Level 3 approaches the neutral threshold (3.60/5).

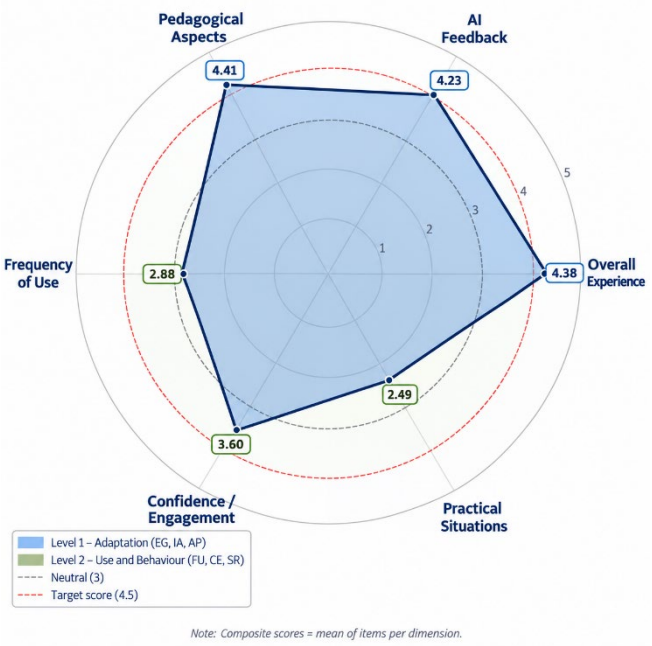


Figure 6: Radar plot of average scores by dimension – Levels 1 and 3 (n = 66/64)

Note : Composite scores = average of items per dimension. Neutral threshold (3/5) shown as a grey dotted line; agreement threshold (4/5) shown as a red dotted line.

5. Discussion

The results confirm that the immersive multimedia approach is associated with significant language gains among Ivorian secondary school pupils. The observed effect size (d = 0.75) is comparable to those reported in the meta-

analysis by Connolly et al. (2012, mean d = 0.66) and Hamari et al. (2016), which reinforces the plausibility of the results despite the absence of a control group.

The analysis differentiated by skill (Figures 2 and 3) constitutes the main empirical contribution of this study. The perceived gains are significantly higher for oral skills (OS: M = 4.12; EO: M = 4.19) than for written skills (EE: M = 3.58), which is consistent with the nature of the VR system: dialogued interactions with conversational agents in the target language primarily engage oral comprehension and expression, whilst grammar and formal writing remain under-practised. This asymmetry, which is absent from the comparative literature on general-purpose serious games, suggests a need to develop supplementary teaching materials specifically targeting writing skills.

The study's second key finding concerns the discrepancy between reported satisfaction and actual transfer, as shown in Figure 6. Motivation to continue (CE4 = 4.57) is remarkably high, but the frequency of independent use outside the school setting remains low (FU3–FU5: 2.33–2.84). This gap suggests that engagement with the game is primarily extrinsic, linked to the institutional framework of the experiment, rather than intrinsic. It echoes the warnings of Deci and Ryan (2000) regarding the fragility of motivation when it depends on a constrained context. Longitudinal studies with open access to the system would be necessary to assess the sustainability of engagement.

Item EG6 – Adapted Level (3.91/5), the only item below the satisfaction threshold, highlights a limitation of adaptive AI: the algorithm does not yet perceive the diverse individual needs of the class with sufficient accuracy. This result serves as a warning of the risk of overestimating the potential for personalisation in a resource-constrained environment.

Several limitations must be explicitly acknowledged. The absence of a randomised control group prevents us from isolating the effect of the game from the Hawthorne effect and the effect of repetition. The convenience sample (N = 68, a single institution) limits the generalisability of the findings. The Level 3 measure is self-reported and susceptible to social desirability bias. The short duration of the intervention may have overestimated the gains due to the novelty effect. Studies with a full experimental design (active control group, randomisation, longitudinal follow-up, direct observation of transfer) are required.

6. Conclusion

This quasi-experimental study provides preliminary data suggesting that a virtual reality serious game enhanced with adaptive AI is associated with significant language gains (d = 0.75) and very high satisfaction (84%) among Ivorian secondary school pupils. It highlights two new findings compared to the existing literature: (1) a differentiated profile of gains, more pronounced in oral than in written skills, which should guide the design of future versions of the programme; (2) a gap between stated engagement and actual independent use outside the school setting, which raises the question of the conditions for transfer.

In practical terms, these findings suggest a gradual integration into the curriculum : training teachers in VR, adapting content to Ivorian cultural references, developing supplementary materials focused on written expression, and longitudinal monitoring to assess the sustainability of the effects. Future research incorporating an active control group, an objective measure of transfer, and monitoring over several months would enable these encouraging trends to be causally validated.

AI Declaration: In preparing this study, the author used the artificial intelligence tool Claude for text restructuring and data processing. The content and results generated by this tool were critically reviewed, verified, and revised by the author prior to their inclusion in the manuscript. The author assumes full responsibility for the final content of this publication as well as for the accuracy of the analyses and interpretations presented herein.

References

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

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. Computers and Education: Artificial Intelligence, 1, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Conati, C., & Kardan, S. (2013). Student modeling: Supporting personalized instruction, from problem solving to exploratory open-ended activities. AI Magazine, 34(3), 13–26. https://doi.org/10.1609/aimag.v34i3.2483

Connolly, T., Boyle, E., MacArthur, E., Hainey, T., & Boyle, J. (2012). A systematic literature review of empirical evidence on computer games and serious games. Computers & Education, 59(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. Harper & Row.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Djaouti, D. (2011). Serious Game Design: Considérations théoriques et techniques sur la création de jeux vidéo à vocation utilitaire [Doctoral dissertation, Université Toulouse III – Paul Sabatier].

Gee, J. P. (2003). What video games have to teach us about learning and literacy. Palgrave Macmillan.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. Computers in Human Behavior, 54, 170–179. https://doi.org/10.1016/j.chb.2015.07.045

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). Evaluating training programs: The four levels (3rd ed.). Berrett-Koehler Publishers.

Koster, R. (2004). A theory of fun for game design. Paraglyph Press.

Mayer, R. E. (2020). Multimedia learning (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. Computers & Education, 70, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Schank, R. C. (1999). Dynamic memory revisited. Cambridge University Press.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Presence, 6(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory. Springer.

Vygotsky, L. S. (1978). Mind in society: Development of higher psychological processes. Harvard University Press.