Enhancing English Speaking Vocabulary through an Augmented Reality Game-Based Learning Framework Guided by Merrill's Principles of Instruction

Sasini Randunuge

Faculty of Computing, ESU, Kandy, Sri Lanka

DOI: https://dx.doi.org/10.47772/IJRISS.2026.1026EDU0120

Received: 15 February 2026; Accepted: 21 February 2026; Published: 11 March 2026

ABSTRACT

This study investigates the pedagogical effectiveness of incorporating Augmented Reality (AR) and gamified learning into an extended instructional framework based on Merrill's First Principles of Instruction to improve English speaking vocabulary acquisition among certificate-level learners in Sri Lanka. Traditional language training has persistent constraints, such as poor learner engagement, limited contextual immersion, and inadequate personalization, necessitating creative pedagogical approaches that promote active speaking growth.

To fill this gap, the study creates and empirically tests an AR-enhanced gamified framework that operationalizes problem-centered learning, prior knowledge activation, demonstration, application, and integration, while also incorporating structured peer collaboration, guided facilitation, and sensitivity to learners' linguistic backgrounds. A positivist, deductive research approach was used, combining a structured 5-point Likert-scale survey delivered to 276 students from two metropolitan campuses with a six-week longitudinal pretest-posttest experimental design.

Data was examined using SPSS, which included reliability testing (Cronbach's $\alpha > 0.7$), correlation analysis, multiple regression, and paired sample t-tests. The study found a strong correlation (adjusted $R^2 = 0.928$) between AR-supported learning effectiveness and instructional factors like tool quality, technical competence, peer collaboration, teacher guidance, instructional relevance, and family linguistic background. Age had no significant influence.

The post-intervention results show statistically significant gains in speaking fluency and vocabulary accuracy. The study provides a theoretically informed and scalable instructional strategy for advancing immersive language learning methods in technology-enhanced educational settings.

Keywords: Augmented Reality, Gamified Learning, Merrill's First Principles of Instruction, English Speaking Vocabulary, Technology-Enhanced Learning

INTRODUCTION

The fast use of digital technology has transformed educational methods, allowing for smart learning environments that mix computing, telecommunications, and interactive systems to provide adaptable, personalized, and context-aware instruction. Smart technology has a significant impact on a variety of sectors, including finance, transportation, agriculture, healthcare, and education, by allowing for seamless interaction and automation [10]. These breakthroughs in language instruction address long-standing hurdles to successful acquisition, notably in terms of speaking vocabulary and fluency.

Traditional language teaching methods, such as grammar-translation, direct, audiolingual, and community language learning, provide structured foundations but have significant limitations, including low learner engagement, limited real-world application, rote memorization, a lack of personalization, and insufficient opportunities for active speaking practice [6]. These limitations frequently impede vocabulary retention,

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue XXVI February 2026 | Special Issue on Education

pronunciation correctness, and communicative confidence, particularly in non-native English settings such as Sri Lanka.

Smart learning combines sophisticated technologies such as cloud computing, augmented reality (AR), virtual reality (VR), Web 2.0 tools, social networks, and gamification to create dynamic, interactive ecosystems that bypass traditional limits [10].

Gamified Augmented Reality (AR) appears to be particularly promising, providing low-cost, time-efficient simulation of real-world scenarios (e.g., conversational interviews or immersive experiences) that promote "learning by doing" with reported recall rates of up to 75% when compared to traditional methods [6].

AR improves focus, motivation, and experiential practice across a wide range of learner demographics, while reducing distractions and enabling personalized feedback.

This study focuses on gamified AR applications like Mondly AR (which supports over 30 languages with interactive simulations, proficiency assessment, and adaptive activities for speaking, reading, listening, pronunciation, and grammar), as well as comparative tools like Virtual Speech, Fulldive VR, Assembler Edu, and Talk (Table 1).

These solutions solve educational inadequacies by offering personalized, interesting skill development pathways.

Table 1. Factors identified to build the conceptual framework.

Topic	**Summary**	**References**
Quality of the Tool	Tool effectiveness relies on performance, usability, accessibility, hedonic motivation, and monetary worth. Primary attributes include ease of use, performance, and availability.	[11, 8, 26]
Impact of Social Support on Language Learning	Social influence, including support from teachers, peers, and institutions, significantly affects user intention in adopting new tools or technologies. Teachers and peers play significant roles.	[2, 16, 5, 25]
Relevance of the Tool in Language Learning	Tools must be interesting, practical, and aligned with learning approaches to equip students for tasks. Relevance enhances learning experience.	[18, 8, 21, 25]
Impact of Technical Knowledge on Learning	Stronger technical knowledge enhances navigation and utilization of digital platforms for language learning. Technology usage positively impacts language learning outcomes.	[12, 1, 7, 13]
Impact of Family Background on Language Learning	Exposure, attitudes, parental involvement, cultural context, socioeconomic status affect language acquisition. Family support is crucial for improving language learning outcomes.	[15, 25, 26]
Impact of Age Group on Language Learning	Critical Period Hypothesis (CPH) debates optimal language acquisition periods. Social robots can aid language learning but age groups may differ in their receptivity to new technology.	[24, 20,14]

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue XXVI February 2026 | Special Issue on Education

Merrill's First Principles of Instruction (MPI) are central to this research, as they emphasize five core elements: problem-centered learning through real-world challenges, activation of prior knowledge, demonstration of new skills, application in authentic tasks, and integration into learners' lives [18] (Figure 1).

MPI has shown beneficial in situations such as MOOCs, increasing engagement, mastery, satisfaction, and completion rates through task-centered and active techniques [3, 24].

Despite its applicability to AR systems, there are significant gaps in extending MPI to gamified AR for language learning, particularly in terms of contextual factors like tool quality, technical knowledge, social support (peer collaboration, teacher guidance), tool relevance, family background, and age (Table 2; Figure 2).

Figure 1. Source: Merrill's Principles of Instructions framework, [17]

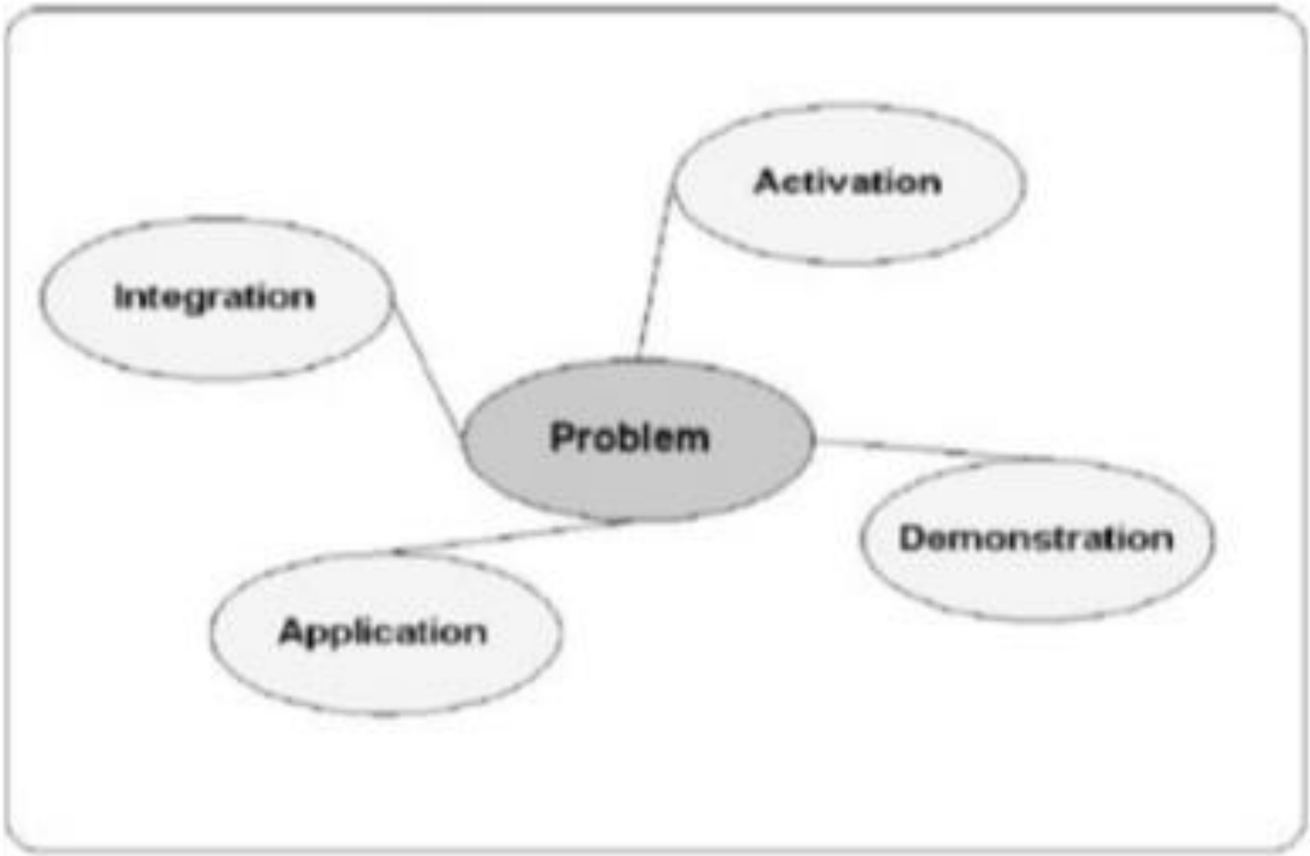
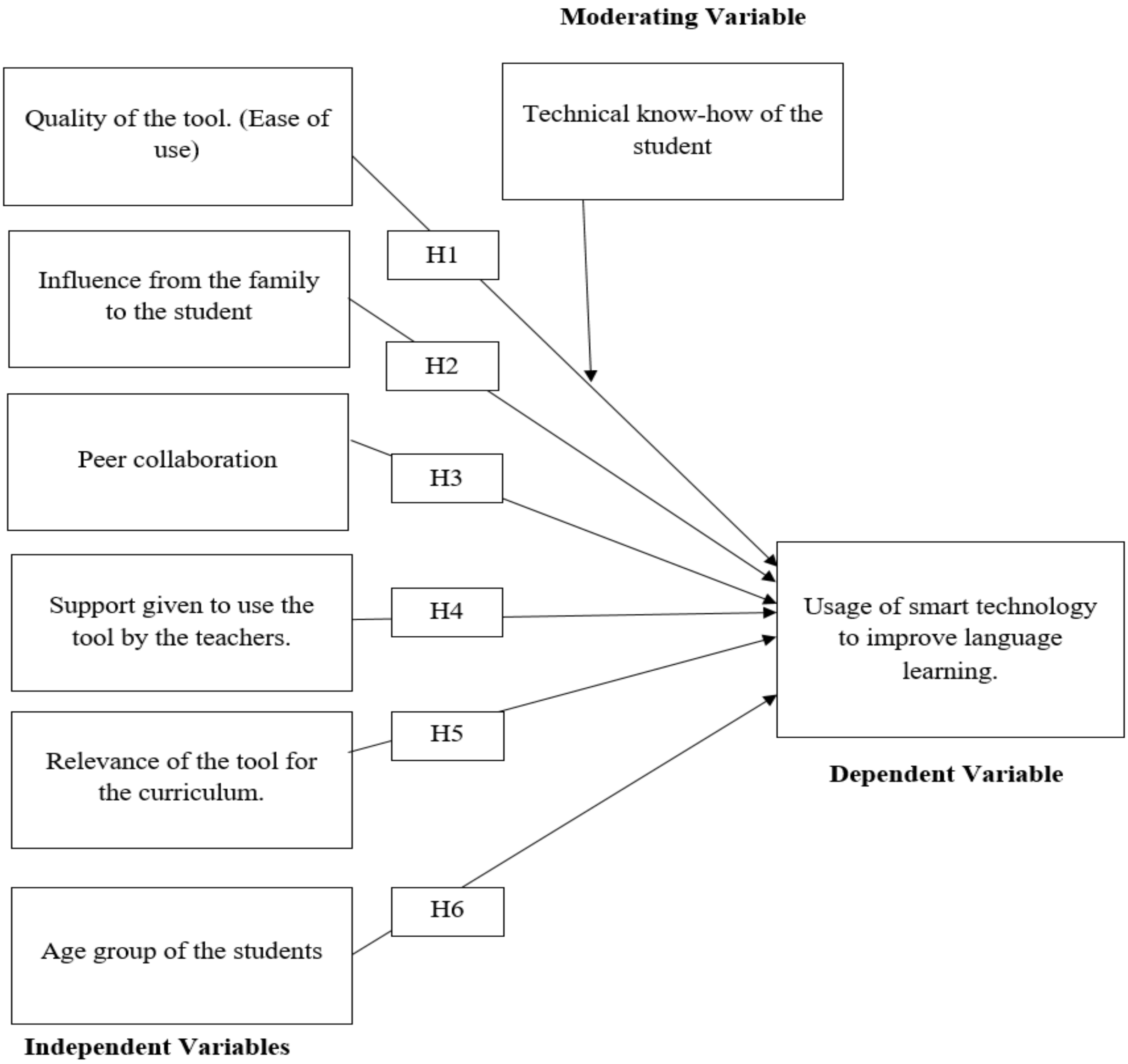


Table 2. Summary of the tools that can be used in similar type of research studies.

Tool	Areas that can improve
Mondly AR	Speaking, Reading, Listening, Pronunciation, Grammar
Virtual Speech, VR Speech	Speaking, Reading, Listening, Pronunciation, Grammar
Fulldive VR	Speaking, Reading, Listening, Pronunciation, Grammar, Writing
Assembler Edu	Speaking, Reading, Listening, Pronunciation, Grammar, Writing
Talk	Speaking, Listening

Figure 2. Conceptual Framework of the AR based language learning research.



This empirical study fills these gaps by looking into the essential success elements that influence AR uptake and efficacy in improving English speaking vocabulary among certificate-level learners. Participants use gamified AR tactics to get engaging, real-world experience. The primary goal is to create and validate an extended MPI-based framework that combines traditional instructional principles with modern technologies like gamification, immersive AR simulations, collaborative elements, and personalized paths to create more effective, inclusive smart language learning environments.

By addressing traditional restrictions through evidence-based technological integration, this study extends instructional design theory and provides a scalable approach for technology-enhanced language education in resource-constrained environments.

METHODOLOGY

This study employs a positivist paradigm to ensure an objective, quantitative, and replicable investigation of the phenomena under consideration. A deductive research strategy was used, with existing theoretical frameworks particularly Merrill's First Principles of Instruction (MPI) [17] guiding hypothesis formulation and structuring empirical inquiry. The research design includes a cross-sectional survey component to investigate affecting factors and a longitudinal quasi-experimental pretest-posttest component to assess the influence of gamified Augmented Reality (AR) interventions on English speaking vocabulary growth.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue XXVI February 2026 | Special Issue on Education

Research Design

A mixed-methods longitudinal methodology was used to gather both quantitative and qualitative data on AR adoption and learning outcomes. Data were collected in two phases, a baseline (pretest) assessment of speaking proficiency and contributing factors, followed by intervention delivery and a posttest evaluation 1.5 months later. This approach enabled a direct comparison of speaking performance before and after the implementation of AR-supported MPI features (guidance, peer cooperation, and integrated components).

Target Population, Sample size and Sampling Technique

The target population consisted of certificate-level English language learners registered at ESOFT METRO CAMPUS's Colombo and Kandy branches. A two-stage cluster sampling technique [4] was used to choose a sample of 276 pupils from an estimated population of 800. To ensure geographic representation, clusters (classes) were selected proportionally at random from both campuses in the first stage. In the second stage, all students from the selected clusters were invited to participate, resulting in a final sample with sufficient statistical power for correlation, regression, and paired-sample analyses.

Data Collection Instruments and Procedures

Two main instruments were used:

Structured Questionnaire

A self-administered questionnaire was created to assess attitudes toward key aspects impacting AR adoption (tool quality, technical expertise, family background, peer collaboration, instructor assistance, tool relevance, and age). Items were scored using a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). The instrument was derived from previously validated scales in the literature and pilot tested for clarity and reliability [19].

Speaking Assessments: Pretest and Posttest

Speaking proficiency was assessed using a performance-based rubric focused on accuracy and fluency, developed in [9] (Figure 3).

Figure 3. Criteria on checking the accuracy and fluency.

Accuracy	Score	Fluency	Score
Little or no language produced.	1	Little or no communication.	1
Poor vocabulary, a mistake in basic grammar, may have a very strong foreign accent.	2	Very hesitant and brief utterances, sometimes difficult to understand.	2
Adequate but not rich vocabulary, occasional grammar slips, slight foreign accent.	3	Gets ideas across, but hesitantly and briefly.	3
Good range of vocabulary, occasional grammar slips.	4	Effective communication in short turns.	4
Wide vocabulary appropriately used, virtually no grammar mistakes.	5	Easy and effective communication uses long turns.	5

Pretest

In a controlled situation, individual oral tasks were used to assess baseline speaking proficiency. Participants were exposed to gamified AR applications (mainly Mondly AR) for 1.5 months. Three MPI-aligned elements were presented sequentially and in combination:

- Teacher assistance (organized scaffolding and feedback).
- Peer collaboration (pair/group engagement in augmented reality scenarios).

- Integrated application (whole MPI cycle, including all pieces)

Posttest:

To measure progress, identical speaking activities were redone [27]. Real-world conversational events were reproduced using AR (Figure 4), allowing for interactive practice with virtual interlocutors.

Figure 4. Given scenario to the students.

Questionnaires and tests were conducted in classroom settings under supervision to assure response quality and reduce external interference.

Data Analysis

The quantitative data were analyzed using IBM SPSS Statistics (version 25 or later). The analysis proceeded in stages.

Checks for reliability and validity: Cronbach's alpha for internal consistency (>0.7 threshold) and the Kaiser-Meyer-Olkin (KMO) measure with Bartlett's factorability test.

Descriptive statistics: Means, standard deviations, frequencies, and cross-tabulations of demographic and variable profiles.

Inferential statistics:

- One-sample t-tests and one-way ANOVA to analyze group differences (e.g., age, family background).
- Pearson correlation analysis is used to examine correlations between variables.
- Multiple linear regression was used to assess the predictive power of independent variables on AR usage and learning outcomes (adjusted R^2 provided).
- A paired-samples t-test was used to compare fluency and accuracy results before and after the test, respectively.

Moderation Analysis: A hierarchical regression was used to determine whether technical expertise acted as a mediator between tool quality and AR adoption.

All statistical tests were performed with a significance level of $p < 0.05$. Ethical considerations included gaining informed consent, ensuring voluntary participation, protecting anonymity, and obtaining institutional approval before data collection.

This methodological framework allowed for a rigorous, evidence-based evaluation of gamified AR's effectiveness inside an MPI-extended model, while also accounting for contextual factors in a real-world educational setting.

RESULTS

This section summarizes the empirical findings from the survey data (n = 276) and the longitudinal pretest-posttest speaking evaluations. IBM SPSS Statistics was used for all analyses, and findings were published at a p-value of less than 0.05.

Reliability and Validity of Instruments

Cronbach's alpha reliability study revealed significant internal consistency across all dimensions, with values higher than 0.70 for each scale (tool quality, technical competence, family background, peer collaboration, instructor guidance, tool relevance, and AR usage). This criterion demonstrates acceptable reliability and strengthens the questionnaire's trustworthiness.

The Kaiser-Meyer-Olkin (KMO) score (> 0.70) and Bartlett's Test of Sphericity ($p < 0.001$) confirmed the dataset's suitability for multivariate analysis.

Descriptive and Inferential Statistics

Kolmogorov-Smirnov and Shapiro-Wilk tests were used to test normality assumptions, and the findings indicated that the majority of data followed an approximate normal distribution. Mean ratings for key constructs differed considerably from the neutral point ($p < 0.05$) using one-sample t-tests, demonstrating non-random response patterns.

A one-way ANOVA indicated significant variations in mean scores across family background categories ($p < 0.05$), indicating that linguistic and socioeconomic family context influenced perceptions and results. In contrast, there were no significant differences between age groups ($p > 0.05$).

Correlation Analysis

Pearson correlation coefficients showed significant positive relationships ($p < 0.01$) between the independent variables (tool quality, technical knowledge, family background, peer collaboration, teacher guidance, tool relevance) and the dependent variable (use of AR tools for English language learning). The strongest relationships were found between tool relevance and teacher guidance.

Regression Analysis

Multiple linear regression was used to calculate the combined explanatory power of the independent factors on AR tool utilization. The model had an adjusted R^2 of 0.928, explaining 92.8% of the variance in AR adoption (Figure 5). The whole model showed statistical significance (F-statistic $p < 0.001$). Regression results (Figure 6) revealed positive and significant beta values for all variables except age group ($\beta \approx -0.011$, $p = 0.522$).

Tool relevance had the most beneficial impact ($\beta = 0.432$, $p < 0.001$), followed by instructor assistance, peer collaboration, family background, and technical competence.

Moderation analysis found that technical expertise strongly moderated the link between tool quality and AR usage (interaction term $p = 0.009$), with higher levels of technical competence having a stronger effect on tool quality.

Figure 5. Model summary of the regression analysis.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.964[a]	.930	.928	.24570	.930	506.631	7	268	.000

a. Predictors: (Constant), Age, Quality, Technical_ability, Techers_guidance, Relevance, Family_support, Peer_collaboration

b. Dependent Variable: Use_of_AR_tool_in_language_learning

Figure 6. Coefficients of the variables.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.092	.076		1.212	.227
	Quality	.317	.054	.322	5.832	.000
	Family_support	.032	.064	.031	.496	.020
	Techers_guidance	.099	.060	.094	1.650	.000
	Peer_collaboration	.085	.084	.080	1.018	.010
	Technical_ability	.200	.068	.190	2.929	.004
	Relevance	.424	.059	.432	7.177	.000
	Age	-.019	-.029	-.011	.642	.522

a. Dependent Variable: Use_of_AR_tool_in_language_learning

Hypotheses Testing

The six hypotheses were investigated using the regression results and related p-values (Table 3).

Table 3. Summary of the analysis.

Factor	H0
H1: Quality of the tool (Ease of use)	Rejected
H2: Family background of the student	Rejected
H3: Peer collaboration	Rejected
H4: Teachers' support to use the tool	Rejected
H5: Relevance of the tool	Rejected
H6: Age group of the student	Accepted

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue XXVI February 2026 | Special Issue on Education

H1: Tool quality leads to AR usage (significant positive effect; rejected null hypothesis).

H2: Family background to AR usage (significant positive effect, rejected null)

H3: Significant positive influence of peer collaboration on AR usage (rejected null hypothesis).

H4: Teacher supervision leads to increased AR usage (significant positive effect; null hypothesis rejected).

H5: Tool relevance leads to AR usage (significant positive effect; rejected null).

H6: There is no significant effect between age group and AR consumption (null hypothesis was not rejected).

Pretest- Posttest Speaking Performance

The paired-samples t-test results showed a substantial improvement in total speaking scores from pretest to posttest (mean difference = -4.993, $p < 0.001$; Figures 7 and 8).

Figure 7. Overall paired sample output.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	OB	13.8623	276	4.81164	.28963
	OA	18.8551	276	4.81294	.28970

Figure 8. Overall mean difference.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	OB - OA	-4.99275	2.30414	.13869	-5.26579	-4.71972	-35.999	275	.000

Detailed frequency distributions (Table 4) revealed clear increases in all three intervention elements:

Table 4. Summary of the results.

<table>
<tr><th>Element</th><th>Fluency</th><th>Accuracy</th></tr>
<tr><td rowspan="2">Guidance</td><td>Before
BeforeDF
<table>
<tr><th></th><th></th><th>Frequency</th><th>Percent</th><th>Valid Percent</th><th>Cumulative Percent</th></tr>
<tr><td>Valid</td><td>Very Low</td><td>77</td><td>27.9</td><td>27.9</td><td>27.9</td></tr>
<tr><td></td><td>Low</td><td>140</td><td>50.7</td><td>50.7</td><td>78.6</td></tr>
<tr><td></td><td>Medium</td><td>38</td><td>13.8</td><td>13.8</td><td>92.4</td></tr>
<tr><td></td><td>High</td><td>21</td><td>7.6</td><td>7.6</td><td>100.0</td></tr>
<tr><td></td><td>Total</td><td>276</td><td>100.0</td><td>100.0</td><td></td></tr>
</table>
As mentioned in the outcome, for the fluency in the test 1, most of the students have got a low mark. The second highest score was for very low. Which depicts the students are below the average in fluency in speaking.</td>
<td>Before
BeforeDA
<table>
<tr><th></th><th></th><th>Frequency</th><th>Percent</th><th>Valid Percent</th><th>Cumulative Percent</th></tr>
<tr><td>Valid</td><td>Very Low</td><td>77</td><td>27.9</td><td>27.9</td><td>27.9</td></tr>
<tr><td></td><td>Low</td><td>140</td><td>50.7</td><td>50.7</td><td>78.6</td></tr>
<tr><td></td><td>Medium</td><td>38</td><td>13.8</td><td>13.8</td><td>92.4</td></tr>
<tr><td></td><td>High</td><td>21</td><td>7.6</td><td>7.6</td><td>100.0</td></tr>
<tr><td></td><td>Total</td><td>276</td><td>100.0</td><td>100.0</td><td></td></tr>
</table>
As mentioned in the outcome, for the accuracy in the test 1, most of the students have got a low mark. The second highest score was for very low. Which depicts the students are below the average in accuracy.</td></tr>
<tr><td>After
AfterDF
<table>
<tr><th></th><th></th><th>Frequency</th><th>Percent</th><th>Valid Percent</th><th>Cumulative Percent</th></tr>
<tr><td>Valid</td><td>Low</td><td>86</td><td>31.2</td><td>31.2</td><td>31.2</td></tr>
<tr><td></td><td>Medium</td><td>149</td><td>54.0</td><td>54.0</td><td>85.1</td></tr>
<tr><td></td><td>High</td><td>29</td><td>10.5</td><td>10.5</td><td>95.7</td></tr>
<tr><td></td><td>Very High</td><td>12</td><td>4.3</td><td>4.3</td><td>100.0</td></tr>
<tr><td></td><td>Total</td><td>276</td><td>100.0</td><td>100.0</td><td></td></tr>
</table>
After providing the guidance for a total 1.5 months of time, the most number of students have got a medium level mark and the below medium students has been reduced from 216 to 86 which proves that there is a clear improvement of the students. Even 12 students have managed to score the highest mark as well.</td>
<td>After
AfterDA
<table>
<tr><th></th><th></th><th>Frequency</th><th>Percent</th><th>Valid Percent</th><th>Cumulative Percent</th></tr>
<tr><td>Valid</td><td>Low</td><td>86</td><td>31.2</td><td>31.2</td><td>31.2</td></tr>
<tr><td></td><td>Medium</td><td>132</td><td>47.8</td><td>47.8</td><td>79.0</td></tr>
<tr><td></td><td>High</td><td>47</td><td>17.0</td><td>17.0</td><td>96.0</td></tr>
<tr><td></td><td>Very High</td><td>11</td><td>4.0</td><td>4.0</td><td>100.0</td></tr>
<tr><td></td><td>Total</td><td>276</td><td>100.0</td><td>100.0</td><td></td></tr>
</table>
After providing the guidance for a total 1.5 months of time, the most number of students have got a medium level mark and the below medium students has been reduced from 216 to 86 which proves that there is a clear improvement of the students. Even 11 students have managed to score the highest mark as well.</td></tr>
<tr><td>Peer collaboration</td><td>Before
BeforePCFluency
<table>
<tr><th></th><th></th><th>Frequency</th><th>Percent</th><th>Valid Percent</th><th>Cumulative Percent</th></tr>
<tr><td>Valid</td><td>Low</td><td>86</td><td>31.2</td><td>31.2</td><td>31.2</td></tr>
<tr><td></td><td>Medium</td><td>149</td><td>54.0</td><td>54.0</td><td>85.1</td></tr>
<tr><td></td><td>High</td><td>29</td><td>10.5</td><td>10.5</td><td>95.7</td></tr>
<tr><td></td><td>Very High</td><td>12</td><td>4.3</td><td>4.3</td><td>100.0</td></tr>
<tr><td></td><td>Total</td><td>276</td><td>100.0</td><td>100.0</td><td></td></tr>
</table>
The previous results from after the guidance was taken to measure the impact of the peer collaboration. Before incorporating peer collaboration, 149 of the highest number of students have obtained a medium mark. 86</td>
<td>Before
BeforePCAccuracy
<table>
<tr><th></th><th></th><th>Frequency</th><th>Percent</th><th>Valid Percent</th><th>Cumulative Percent</th></tr>
<tr><td>Valid</td><td>Low</td><td>86</td><td>31.2</td><td>31.2</td><td>31.2</td></tr>
<tr><td></td><td>Medium</td><td>132</td><td>47.8</td><td>47.8</td><td>79.0</td></tr>
<tr><td></td><td>High</td><td>47</td><td>17.0</td><td>17.0</td><td>96.0</td></tr>
<tr><td></td><td>Very High</td><td>11</td><td>4.0</td><td>4.0</td><td>100.0</td></tr>
<tr><td></td><td>Total</td><td>276</td><td>100.0</td><td>100.0</td><td></td></tr>
</table>
The previous results from after the guidance was taken to measure the impact of the peer collaboration. Before incorporating peer collaboration, 132 of the highest number of students have obtained a medium mark. 86</td></tr>
</table>

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue XXVI February 2026 | Special Issue on Education

has a low mark and the rest of 41 students have a mark above the average.

has a low mark and the rest of 58 students have a mark above the average.

After

AfterPCFluency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low	48	17.4	17.4	17.4
	Medium	142	51.4	51.4	68.8
	High	54	19.6	19.6	88.4
	Very High	32	11.6	11.6	100.0
	Total	276	100.0	100.0	

After incorporating the peer collaboration element, the students who got a low score has been decreased from 86 to 48. There is no much difference in the medium score but the above medium score has been increased from 41 to 86.

After

AfterPCAccuracy

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low	48	17.4	17.4	17.4
	Medium	139	50.4	50.4	67.8
	High	57	20.7	20.7	88.4
	Very High	32	11.6	11.6	100.0
	Total	276	100.0	100.0	

After incorporating the peer collaboration element, the students who got a low score has been decreased from 86 to 48. There is no much difference in the medium score but the above medium score has been increased from 58 to 89.

background

Before

	BeforeBF				
	Very Low	Low	Medium	High	Total
Sinhala	68	136	5	0	209
English	0	0	32	16	48
Tamil	9	4	1	5	19
	77	140	38	21	276

These scores (before) were taken before incorporating the guidance and peer collaboration elements. As this mentions, both Sinhala and Tamil students has low and very low scores. But the ones with an English background, have obtained medium to high scores. Also there are few (5) students from Tamil background that have obtained high marks.

Before

	BeforeBA				
	Very Low	Low	Medium	High	Total
Sinhala	68	136	5	0	209
English	0	0	32	16	48
Tamil	9	4	1	5	19
	77	140	38	21	276

These scores (before) were taken before incorporating the guidance and peer collaboration elements. As this mentions, both Sinhala and Tamil students has low and very low scores. But the ones with an English background, have obtained medium to high scores. Also there are few (5) students from Tamil background that have obtained high marks.

After

	AfterBF				
	Low	Medium	High	Very High	Total
Sinhala	42	133	31	3	209
English	0	0	21	27	48
Tamil	6	9	2	2	19
	48	142	54	32	276

After incorporating the above three elements, it was tested that students from which background have improved more. The students who are having a Sinhala background, have decreased the below average level by 26. And in the medium level there in no much difference. But the

After

	AfterBA				
	Low	Medium	High	Very High	Total
Sinhala	42	132	33	2	209
English	0	0	20	28	48
Tamil	6	7	4	2	19
	48	139	57	32	276

After incorporating the above three elements, it was tested that students from which background have improved more. The students who are having a Sinhala background, have decreased the below average level by 26. And in the medium level there in no much difference. But the above

	above average limit has been also increased by 29 which shows an improvement of the speaking skills. 11 students who has an English background who has got high mark has moved from high to very high level. And the ones who has a Tamil background, also have reduced the count in below average by 3. Also has improved the average value by 8.	average limit has been also increased by 30 which shows an improvement of the speaking skills. 12 students who has an English background who has got high mark has moved from high to very high level. And the ones who has a Tamil background, also have reduced the count in below average by 3. Also has improved the average value by 3.

Guidance: Following intervention, the proportion of children scoring below medium reduced from 216 to 86 in both fluency and accuracy; medium-level scores became the norm, with 11-12 kids earning the highest category.

Peer Collaboration: Low scores dropped from 86 to 48, while above-medium levels climbed significantly.

Family/Linguistic Background: Students from Sinhala and Tamil backgrounds had the greatest reduction in below-average performance (Sinhala: -26; Tamil: -3). Students with an English background advanced to the extremely high levels.

These trends were constant across the efficiency and accuracy dimensions, indicating that the combined MPI elements (guidance + peer collaboration + background consideration) resulted in the greatest increases.

DISCUSSION

The findings give strong empirical support for the efficacy of gamified Augmented Reality (AR) combined with an extended Merrill's First Principles of Instruction (MPI) paradigm in increasing English speaking vocabulary among certificate-level students.

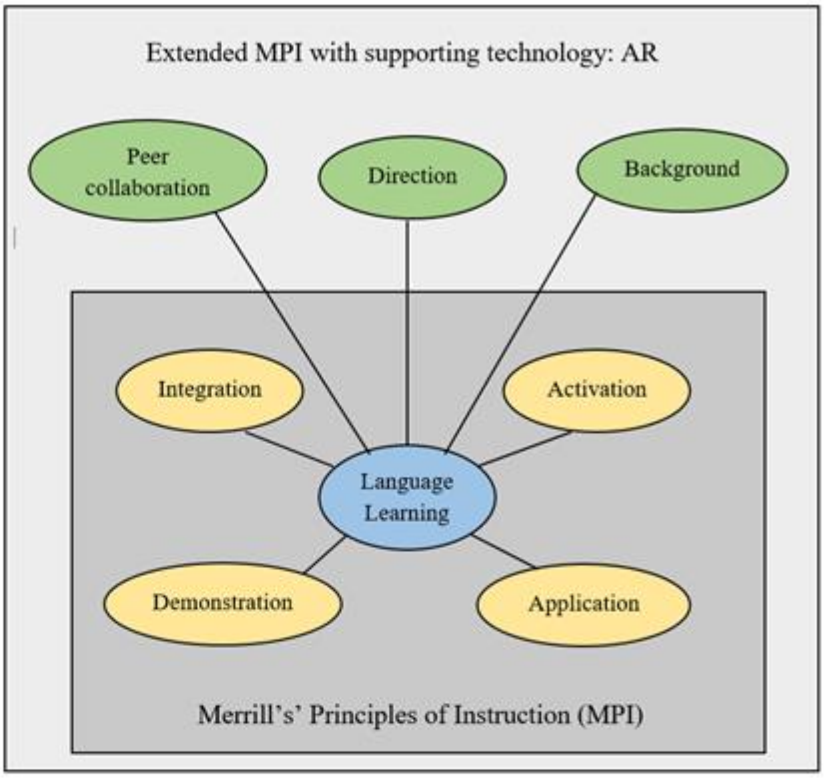
The high adjusted R^2 value (0.928) indicates that the identified factors tool quality, technical expertise, family background, peer collaboration, teacher assistance, and tool relevance explain almost all variance in AR adoption. This finding is consistent with previous research highlighting the relevance of perceived utility, convenience of use, and social support in technology acceptability for language acquisition [7-9]. Notably, tool relevance emerged as the best predictor, indicating that alignment with curriculum goals and practical speaking tasks is critical for inspiring long-term participation.

The considerable moderating influence of technical expertise in the tool quality-AR usage connection shows that learners with higher levels of digital competence benefit more from high-quality AR interfaces, which is consistent with previous research on technology-mediated learning [22, 23]. In contrast, the lack of a significant effect of age supports the claim that modern AR tools are accessible to people of all ages when appropriately scaffolded, calling into question the Critical Period Hypothesis in technology-assisted contexts [12].

The pre-posttest results are especially convincing. AR-facilitated MPI elements increase speaking skills, as evidenced by the considerable mean improvement ($p < 0.001$) and consistent gains in fluency and accuracy. The sequential and combined introduction of guidance, peer collaboration, and background-sensitive adaptation resulted in the most significant adjustments, particularly among Sinhala- and Tamil-background learners who began with lower baselines. This highlights the need of socially and contextually appropriate interventions in closing equity inequalities in language instruction.

These findings expand on existing MPI applications (already verified in MOOCs and hybrid contexts [5,6]) by establishing their effectiveness in immersive, gamified AR environments for speaking-focused language acquisition. The suggested extended framework (Figure 9) successfully integrates instructional design theory and modern technology, eliminating traditional limitations such as a lack of real-world experience, restricted feedback, and one-size-fits-all instruction.

Figure 9. Extended MPI framework.



Limitations include the specific context (certificate-level learners on two metropolitan campuses), the possibility of self-report bias in survey data, and the intervention's brief duration (1.5 months). Nonetheless, the data provide compelling evidence that including peer participation, teacher direction, and family background awareness into AR-supported MPI considerably enhances speaking outcomes.

Future research should replicate this design in a variety of linguistic and socioeconomic contexts, extend the intervention period for long-term retention study, and include qualitative learner perspectives to improve understanding of motivational factors.

This study confirms the transformative power of gamified AR when embedded in a theoretically grounded, contextually sensitive instructional framework, offering a realistic model for promoting smart language learning in developing educational situations.

CONCLUSION

This study looked into the use of gamified Augmented Reality (AR) as a new intervention to improve English speaking vocabulary acquisition among certificate-level students at ESOFT METRO CAMPUS in Sri Lanka. The research addressed key limitations of traditional language teaching methods, such as limited engagement, insufficient real-world practice, and a lack of personalization, by incorporating AR tools within an extended framework based on Merrill's First Principles of Instruction (MPI).

The empirical findings give compelling evidence that various contextual and instructional elements have a substantial impact on the acceptance and effectiveness of AR-based language acquisition. Positive predictors of AR usage include tool quality, technical competence, family language background, peer collaboration, teacher guidance, and curricular relevance, which account for 92.8% of the variation (adjusted $R^2 = 0.928$).

However, there was no significant link between age group and AR intervention effectiveness, demonstrating that correctly designed AR treatments can be effective across learner age ranges when technical scaffolding is provided.

The longitudinal pretest-posttest results showed evident and statistically significant gains in both fluency and accuracy after incorporating MPI-aligned aspects (teacher assistance, peer cooperation, and background-sensitive adaptation) into gamified AR apps. These benefits were particularly noticeable when aspects were combined, demonstrating the synergistic usefulness of social support, structured facilitation, and contextual relevance in technology-enhanced speaking practice. Learners from non-English-dominant families (Sinhala and Tamil) made particularly impressive gains, implying that such interventions can help bridge equity disparities in language instruction.

Building on these findings, the study offers an expanded MPI-based framework that includes gamification, immersive AR simulations, real-time interactive feedback, peer collaboration, instructor scaffolding, and consideration for learners' familial and cultural backgrounds (Figure 9). This framework combines proven instructional design concepts with modern digital resources to provide a theoretically solid and practically realistic paradigm for overcoming traditional constraints such as passive learning, a lack of experience possibilities, and one-size-fits-all training.

The findings add to the growing discussion about smart learning environments by demonstrating that AR, when integrated into a structured, context-responsive instructional design, can significantly improve speaking vocabulary outcomes in resource-constrained educational contexts. They affirm the revolutionary power of technology integration in language education and advocate for a change toward more engaging, inclusive, and successful instructional techniques.

Future research should build on this work by conducting longitudinal studies to assess long-term retention and transfer of speaking skills, replicating the framework across different age groups, proficiency levels, and cultural contexts, and investigating the use of artificial intelligence for additional personalization. Additionally, teacher training programs and accessibility concerns for students with diverse levels of digital literacy will be required to ensure scaled adoption.

This study provides empirical support for the usage of gamified AR within an MPI-extended framework as an effective technique for improving English language speaking proficiency. By encouraging the integration of instructional theory and emerging technologies, it lays the groundwork for more dynamic, equitable, and future-ready language teaching paradigms.

ACKNOWLEDGEMENTS

I am grateful to my supervisors and mentors at Kingston University UK and ESOFT METRO CAMPUS for their invaluable assistance and support throughout this project. I am grateful to the management and personnel of ESOFT METRO CAMPUS in Colombo and Kandy for allowing participant access, as well as to the 276 certificate-level English language learners who willingly participated in the study and provided valuable input. Special thanks to my family and friends for their unwavering encouragement and understanding throughout this journey. Finally, I thank the developers of the Augmented Reality tools utilized in this study for making it possible to put the treatments into practice. This work would not have been feasible without the collaborative efforts of everyone involved.

REFERENCES

1. Ahmadi, D. M. R. (2018). The use of technology in English language learning: A literature review. *International Journal of Research in English Education*, 3(2), 115–125. https://doi.org/10.29252/ijree.3.2.115

2. Argo, J. J., & Dahl, D. W. (2020). Social influence in the retail context: A contemporary review of the literature. *Journal of Retailing*, 96(1), 25–39. https://doi.org/10.1016/j.jretai.2019.12.005
3. Badali, M., Hatami, J., Farrokhnia, M., & Noroozi, O. (2022). The effects of using Merrill's First Principles of Instruction on learning and satisfaction in MOOC. *Innovations in Education and Teaching International*, 59(2), 216–225. https://doi.org/10.1080/14703297.2020.1813187
4. Biggs, H. M., Harris, J. B., Breakwell, L., Dahlgren, F. S., Abedi, G. R., Szablewski, C. M., Drobeniuc, J., Bustamante, N. D., Almendares, O., Schnall, A. H., & Gilani, Z. (2020). Estimated community seroprevalence of SARS-CoV-2 antibodies — two Georgia counties, April 28–May 3, 2020. *Morbidity and Mortality Weekly Report*, 69(29), 965–970. https://doi.org/10.15585/mmwr.mm6929a5
5. Cho, J. H. (2018). A study on the factors influencing the satisfaction and effectiveness of smart learning in the view of HRD in company. *Journal of the Korea Academia-Industrial Cooperation Society*, 19(3), 468–478. https://doi.org/10.5762/KAIS.2018.19.3.468
6. Demir, K. A. (2021). Smart education framework. *Smart Learning Environments*, 8(1), 1–36. https://doi.org/10.1186/s40561-021-00170-x
7. Evans, M. (Ed.). (2009). *Foreign language learning with digital technology*. A&C Black.
8. Gambo, Y., & Shakir, M. Z. (2022). Evaluating students' experiences in self-regulated smart learning environment. *Education and Information Technologies*, 27(1), 1–34.
9. Ho, P. V. P., Thien, N. M., An, N. T. M., & Vy, N. N. H. (2020). The effects of using games on EFL students' speaking performances. *International Journal of English Linguistics*, 10(1), 183–193. https://doi.org/10.5539/ijel.v10n1p183
10. Hosseindoost, S., Khan, Z. H., & Majedi, H. (2022). A shift from traditional learning to e-learning: Advantages and disadvantages. *Archives of Neuroscience*, 9(2), e122804.
11. Kang, N. (2015). The comparison between regular and flipped classrooms for EFL Korean adult learners. *Multimedia-Assisted Language Learning*, 18(3), 1–28.
12. Ko, M. H. (2017). Learner perspectives regarding device type in technology-assisted language learning. *Computer Assisted Language Learning*, 30(8), 844–863. https://doi.org/10.1080/09588221.2017.1367310
13. Lai, C. (2013). A framework for developing self-directed technology use for language learning. *Language Learning & Technology*, 17(2), 100–122.
14. Lee, H., & Lee, J. H. (2022). The effects of robot-assisted language learning: A meta-analysis. *Educational Research Review*, 35, 100425. https://doi.org/10.1016/j.edurev.2021.100425
15. Li, Z., & Qiu, Z. (2018). How does family background affect children's educational achievement? Evidence from Contemporary China. *The Journal of Chinese Sociology*, 5(1), 1–21. https://doi.org/10.1186/s40711-018-0083-8
16. Liew, B. T., Kang, M., Yoo, E., & You, J. (2013). Investigating the determinants of mobile learning acceptance in Korea. In *Proceedings of EdMedia + Innovate Learning 2013* (pp. 1424–1430). Association for the Advancement of Computing in Education (AACE).
17. Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, 50(3), 43–59. https://doi.org/10.1007/BF02505024
18. Omidinia, S., Masrom, M., & Selamat, H. (2012). Determinants of smart school system success: A case study of Malaysia. *International Journal of Academic Research*, 4(1), 17–25.
19. Qin, Y. (2022). Analysis of online learning questionnaire based on network information technology. In *Proceedings of the International Conference on Innovative Computing* (pp. 395–401). Springer.
20. Singleton, D., & Leśniewska, J. (2021). The critical period hypothesis for L2 acquisition: An unfalsifiable embarrassment? *Languages*, 6(3), 149. https://doi.org/10.3390/languages6030149
21. Thomas, L. J., Parsons, M., & Whitcombe, D. (2019). Assessment in smart learning environments: Psychological factors affecting perceived learning. *Computers in Human Behavior*, 95, 197–207. https://doi.org/10.1016/j.chb.2018.11.037

22. Thompson, P. (2021). Learning by doing. In *Handbook of the Economics of Innovation* (Vol. 1, pp. 429–476). Elsevier.
23. Truong, M. T., Elen, J., & Clarebout, G. (2019). Implementing Merrill's first principles of instruction: Practice and identification. *Journal of Educational and Instructional Studies in the World*, 9(2), 14–28.
24. Tryfonova, Y. (2021). Age and critical period hypothesis. [Unpublished manuscript or source as cited; replace with full publication details if available].
25. Ventura, R., Roca-Cuberes, C., & Corral-Rodríguez, A. (2018). Interactive digital communication: Assessment of academic competencies and professional profiles by communication professionals, faculty and students. *Revista Latina de Comunicación Social*, 73, 331–351.
26. Wang, S., Shi, G., Lu, M., Lin, R., & Yang, J. (2021). Determinants of active online learning in the smart learning environment: An empirical study with PLS-SEM. *Sustainability*, 13(17), 9923. https://doi.org/10.3390/su13179923
27. Wu, X. V., Chi, Y., Panneer Selvam, U., Devi, M. K., Wang, W., Chan, Y. S., Wee, F. C., Zhao, S., Sehgal, V., & Ang, N. K. E. (2020). A clinical teaching blended learning program to enhance registered nurse preceptors' teaching competencies: Pretest and posttest study. *Journal of Medical Internet Research*, 22(4), e18604. https://doi.org/10.2196/18604