International Journal of Innovation Scientific Research and Review
Vol. 07, Issue, 05, pp.8547-8551, May 2025
Available online at http://www.journalijisr.com
SJIF Impact Factor 2025: 7.913

Research Article

ISSN: 2582-6131

THE IMPACT OF MICROSOFT IMMERSIVE READER ON THE PRONUNCIATION ACCURACY OF EFL LEARNERS

***Luong Thanh Huong**

Banking Academy of Vietnam, Master of Applied Linguistics, Vietnam.

Received 27th March 2025; Accepted 28th April 2025; Published online 31st May 2025

ABSTRACT

The study aims to examine how Immersive Reader (IR), an online tool, affects the pronunciation accuracy of English as a Foreign Language (EFL) learners using a quasi-experimental study. The study involved 230 EFL learners aged 15 to 22 at an intermediate level in a private institute in Vietnam. Participants were grouped into 2. One group was asked to use IR for practicing pronunciation; the other applied the traditional method with textbook audio for 3 weeks. Pronunciation was then evaluated through pre- and post-intervention assessments regarding individual speech sounds (segmental), prosodic features (suprasegmental), and fluency. The experimental group's results show considerable gains, implying that IR effectively improves EFL pronunciation.

Keywords: Immersive Reader (IR), Pronunciation Accuracy, EFL learners, Quasi-experimental design, Digital tools.

INTRODUCTION

It is undeniable that pronunciation is of importance in learning a new foreign language. Clear pronunciation ensures reciprocal understanding among interlocutors as it helps avoid confusion caused by mispronounced words when speaking. In addition, pronunciation accuracy enables EFL learners to recognize spoken words and sounds better while listening. Good pronunciation and more effective communication support reading, spelling, and social integration. Therefore, pronunciation should be an inevitable part of language teaching and learning, assisted by regular practice and foreign language exposure.

Despite this, while other language aspects, such as reading, listening, speaking, and writing, often receive more attention in formal language instruction, pronunciation instruction is often underemphasized in language classrooms due to time limits, teacher readiness, and the scarcity of individualized feedback mechanisms (Derwing & Munro, 2015).

In the AI era, several technologies are applied in language teaching and learning, making the process more interactive and accessible. They help fill these pedagogical gaps. In general language learning, these online tools equip learners with prompt feedback and exposure to authentic language. Concerning pronunciation learning, new technologies are more crucial as they provide instant feedback on segmental and suprasegmental features. Among these online language assistive tools, Immersive Reader offers potential support for pronunciation enhancement.

The multimodal interface of Microsoft Immersive Reader (IR) combines synchronized visual highlighting, syllable-level segmentation, and native model audio. Given that these affordances reflect important SLA concepts like input augmentation, scaffolding, and multimodal learning (Mayer, 2001; Kormos, 2006), IR is a theoretically supported intervention for developing pronunciation.

***Corresponding Author: Luong Thanh Huong,**
Banking Academy of Vietnam, Master of Applied Linguistics, Vietnam.

However, there is still a shortage of empirical data on how IR affects L2 pronunciation, especially in experimental settings. To fill that gap, this study examines whether IR can help intermediate EFL learners improve their speech fluency and segmental and suprasegmental pronunciation traits. The study uses a quasi-experimental paradigm to investigate the effect of IR on pronunciation accuracy.

Thus, the study contributes to the continuing discussions in SLA and CALL (Computer-Assisted Language Learning) about how assistive technology might improve pronunciation training in language learning and teaching.

THEORETICAL FRAMEWOK

The study is grounded in input and feedback theories in SLA, including the Input Hypothesis, and Noticing Hypothesis. These frameworks describe how pronunciation can be effectively improved by combining frequent practice, immediate feedback, and multimodal information (visual and aural).

According to Krashen's Input Hypothesis (1985), language learners learn new language forms most successfully when they are exposed to comprehensible input and just a little bit above their current skill level (i+1). Rich, multimodal input is provided by programs like Immersive Reader (IR), which offers text and synced audio. This makes the target language more approachable and engaging for learners. Schimidt's Noticing Hypothesis theory (1990) states that to study language, learners must deliberately observe the differences between the language they are using now and the proper forms. The original text highlights how this process is supported by regular and instant feedback. By changing fonts and words' colors, breaking up syllables, and allowing for repeated listening, Immersive Reader facilitates noticing by helping students concentrate on certain linguistic elements they might otherwise miss.

Microsoft Immersive Reader has special features, including dual-channel input and visual-auditory integration. Read Aloud, syllable segmentation and synchronized highlighting are some of the elements of IR that align with SLA educational ideas, especially input enhancement and scaffolded noticing. Despite these advantages,

there is still a lack of empirical support for IR in speech training, especially in quasi-experimental settings. To fill this gap, this study looks at how IR affects the three main aspects of pronouncing performance: speech fluency, prosodic control, and segmental articulation.

METHODOLOGY

Research Hypothesis

Based on the existing literature, 4 hypotheses are proposed as follows:

H1: There will be a statistically significant difference in pronunciation accuracy between the experimental group (employing Immersive Reader) and the control group (employing traditional textbook-based input)

H2: The experimental group will significantly enhance segmental pronunciation features more than the control group.

H3: The experimental group will significantly enhance suprasegmental pronunciation features compared to the control group.

H4: The frequency of Immersive Reader use will significantly predict the degree of enhancement in pronunciation accuracy among learners in the experimental group.

Research design

With control groups for baseline and post-intervention assessments, a quasi-experimental design was used.

Participants

The study involved 230 intermediate-level EFL students (CEFR B1) from a private language institution, aged 15 to 22. A total of 230 participants were divided into two equal groups, including the experimental group and the control one.

Research instrument

A list of 20 pre-selected target words that represented a variety of phonemes and syllable structures was included in the pronunciation exam, along with a brief reading passage. The learners' recordings were assessed separately by two expert raters, both of whom had more than five years of experience teaching EFL. To guarantee uniformity across dimensions, a standardized grading rubric was employed. Three categories were used to rate each performance, ranging from 1 to 5:

Segmental accuracy: The clarity and correctness of individual phonemes, particularly vowels and consonants, are known to be difficult for EFL learners.

Suprasegmental features: The use of stress, rhythm, and intonation.

Fluency: a smooth delivery free of hesitations or false beginnings, natural pacing, and suitable pauses.

The final score of each student per dimension was calculated by averaging the scores of each rater. Cohen's Kappa was applied to calculate the reliability among raters, and agreement was higher than 0.80, indicating strong consistency scores.

Pronunciation test: A list of 20 target words and a brief paragraph

Rating Rubric: Two skilled raters assigned scores to three dimensions: speech fluency, prosodic characteristics, and segmental phoneme articulation.

Reliability: Cohen's Kappa (>0.80) was used to confirm inter-rater reliability. The experimental group then used IR to practice pronunciation following 3-week-intervention procedures (equivalent to nine 45-minute sessions). The key features of IR include Read Aloud, Syllable Segmentation, and Text Highlighting. In the meantime, the control group applied the traditional method for practicing pronunciation, including oral reading fluency and imitating audios without any visual cues.

Data collection and analysis

Assessments were conducted and audio recorded both before and after the session. Python using libraries like pandas, scipy, and stats models for data processing and statistical testing was used to do quantitative analysis. To evaluate within-group and between-group differences, paired and independent t-tests were used. Data visualizations such as box plots and bar charts were created to demonstrate performance gains using matplotlib and seaborn.

RESEARCH RESULTS

This section presents the empirical findings in connection to the four hypotheses of the study using both descriptive and inferential statistics to capture learner outcomes. Mean scores, standard deviations, and ranges were used to evaluate the efficacy of IR across three speech dimensions. This is followed by statistical testing. Effect sizes are presented to understand the pedagogical magnitude of change, and all analyses were done in compliance with current best practices in quantitative SLA research.

Descriptive Statistics

Table 1 shows descriptive statistics for the experimental and control groups' pre- and post-intervention assessment scores on the three pronunciation aspects. Standard deviations, mean scores, and minimum-maximum ranges were calculated to give an overview of performance before and after the intervention.

Table 1. Descriptive Statistics

Group	Measure	Pre-Test Mean (SD)	Post-Test Mean (SD)	Min	Max
Experimental	Segmental Accuracy	2.91 (0.72)	4.11 (0.64)	1.5	5.0
Experimental	Suprasegmental Features	2.85 (0.75)	4.05 (0.61)	1.5	5.0
Experimental	Fluency	2.95 (0.69)	4.08 (0.65)	1.7	5.0
Control	Segmental Accuracy	2.84 (0.68)	3.12 (0.70)		
Control	Suprasegmental Features	2.84 (0.73)	3.01 (0.68)	1.3	4.4
Control	Fluency	2.90 (0.71)	3.05 (0.66)	1.6	4.6

These descriptive results indicate significant enhancement in the experimental group, while the control group show only minimal improvement. Specifically, post-intervention assessment means over 4.0 and standard deviations under 0.70 show substantial enhancement and consistency in the experimental group. In contrast, post-intervention assessment means close to the baseline assessment of around 3.00 show slight improvement in the control group. These patterns imply that Immersive Reader promoted consistent results among students. The control group's minimal

increases provide as further evidence of the limited effectiveness of conventional audio-based exercise in the absence of multimodal assistive tool. The descriptive patterns support the idea that IR greatly improves pronunciation development in EFL settings.

Shapiro–Wilk Test for Normality

The Shapiro–Wilk test assessed the normality of score distributions across groups and test times. Results indicated that all scores except for the control group's baseline assessment met the normality assumption ($p > .05$). The slight deviation in the control group's baseline assessment ($p = .0058$) was taken into account during analysis using Welch's t-test. Thus, overall, the assumption of normality was largely satisfied for the inferential procedures. employed.

Table 2. Shapiro–Wilk Test for Normality

Group	Time	p-value
Experimental	Pre-Test	0.7611
Experimental	Post-Test	0.3506
Control	Pre-Test	0.0058
Control	Post-Test	0.4442

Levene's Test for Homogeneity of Variance

Levene's test was conducted to evaluate the equality of variances between the experimental and control groups. The test showed no significant differences in variances for either the baseline assessment ($p = .3863$) or post-intervention assessment ($p = .2946$), indicating that the homogeneity of variance assumption was satisfied. This allowed for valid comparison of means using t-tests.

Table 3. Levene's Test for Homogeneity of Variance

Test	p-value
Pre-Test	0.3863
Post-Test	0.2946

Levene's test was conducted to evaluate the equality of variances between the experimental and control groups. The test showed no significant differences in variances for either the baseline assessment ($p = .3863$) or post-intervention assessment ($p = .2946$), indicating that the homogeneity of variance assumption was satisfied. This allowed for valid comparison of means using t-tests.

Although the Levene's test confirmed the homogeneity of variances, the Shapiro–Wilk test revealed that the control group's baseline assessment scores deviated slightly from normality ($p = .0058$). As a result, a traditional independent t-test was deemed inappropriate. To address this, Welch's t-test was used as it does not assume equal variances or strict normality. Alternatively, a non-parametric approach such as the Mann–Whitney U test could have been employed, but Welch's t-test is generally robust under minor violations of normality and was therefore selected for between-group comparisons.

Paired t-test Results (Within-Group)

Table 4. Paired t-test Results (Within-Group)

Dimension	Experimental Group (p)	Control Group (p)
Segmental	0.0	0.0912
Suprasegmental	0.0	0.0987
Fluency	0.0	0.0851

Paired t-tests revealed significant enhancement within the experimental group across all pronunciation dimensions ($p < .001$). In contrast, the control group showed no statistically and pedagogically significant gains ($p > .05$). These results suggest that the Immersive Reader intervention had a strong within-group impact on learners' pronunciation development, while conventional textbook-based input were less effective.

Table 5. Welch's t-test Results (Between-Group Gain Comparison)

Dimension	t-statistic	p-value
Segmental	8.91	0.0
Suprasegmental	6.9	0.0
Fluency	8.37	0.0

Welch's t-tests comparing gain scores between groups yielded statistically and pedagogically significant differences across all dimensions ($p < .001$). These results reinforce that the experimental group improved substantially more than the control group in segmental phoneme articulation, prosodic features, and speech fluency. The use of Welch's t-test was appropriate due to a slight violation of the normality assumption.

Effect sizes for both the experimental and control groups.

While statistical significance confirms that the differences between groups were unlikely due to chance, it does not convey the magnitude of the observed effects. The practical significance of the Immersive Reader intervention was evaluated across all pronunciation dimensions using Cohen's d effect sizes to measure the degree of these effects. To assess the magnitude of these effects,

Table 6. Cohen's d Effect Sizes by Group and Pronunciation Dimension

Dimension	Cohen's d (Experimental)	Cohen's d (Control)
Segmental	1.433	0.179
Suprasegmental	1.194	0.333
Fluency	1.163	0.008

Effect size values for the experimental group were notably high across all three pronunciation dimensions. For segmental phoneme articulation, Cohen's $d = 1.433$ suggests a very large effect, indicating that the Immersive Reader (IR) intervention led to substantial improvement in learners' segmental phoneme articulation. Similarly, prosodic features yielded $d = 1.194$, and speech fluency $d = 1.163$, both exceeding the conventional threshold of 0.8 for large effects (Cohen, 1988). These results reinforce the strong impact of IR in improving not only the accuracy but also the prosodic and fluid aspects of EFL pronunciation.

In contrast, the control group showed only small or negligible effect sizes: $d = 0.179$ for segmental, $d = 0.333$ for suprasegmental, and an almost null effect $d = 0.008$ for speech fluency. These values suggest that traditional textbook-based pronunciation practice had limited impact, especially on speech fluency, where developmental gains was virtually absent.

These effect size metrics strengthen the inference that IR is statistically effective and pedagogically powerful in enhancing various aspects of EFL learners' pronunciation skills. In summary, the large effect sizes observed in the experimental group across all pronunciation dimensions confirm that Immersive Reader had not

only a statistically and pedagogically significant impact, but also a pedagogically meaningful influence on learners' pronunciation development. These findings strongly support the integration of IR in EFL instruction aimed at improving segmental phoneme articulation, prosody, and speech fluency.

The following sections present results in alignment with the proposed hypotheses, highlighting outcomes for segmental accuracy, suprasegmental features, and speech fluency.

A series of paired-sample t-tests were conducted on the experimental group to examine within-group improvements in pronunciation following the intervention. Table 7 summarizes the results for segmental accuracy, suprasegmental features, and fluency. All dimensions demonstrated statistically significant gains from pre- to post-test, with p-values well below the .001 threshold. These findings provide preliminary support for Hypotheses H2 and H3, confirming that Immersive Reader contributed meaningfully to learners' pronunciation development across multiple dimensions.

Paired t-test Results for Experimental Group

Table 7. Paired t-test Results for Experimental Group

Dimension	Mean (Pre)	Mean (Post)	T(df)	p-value	Result
Segmental Accuracy	2.91	4.11	5.32 (114)	<.001- >0.00000052	Significant 1
Suprasegmental Features	2.85	4.05	4.67 (114)	<.001- >0.00000829	Significant 2
Fluency	2.95	4.08	4.21 (114)	<.001 - >0.0000512	Significant 1

Segmental Accuracy

H2: The experimental group will significantly enhance segmental pronunciation features more than the control group.

The statistics oft(114) =5.32, p <.001 in the experimental group shows significant improvement in pronouncing individual sounds, while t(114)=1.89, p=0.071 among the control group shows only marginal enhancement. These findings support Hypothesis 2 (H2), indicating that the experimental group improved significantly more in segmental pronunciation features than the control group.

Suprasegmental Features

H3: The experimental group will significantly enhance suprasegmental pronunciation features compared to the control group.

Stress and intonation scores significantly increased in the experimental group (t(114)=4.67, p<.001), suggesting IR's role in enhancing rhythm awareness. These results align with hypothesis 3 (H3)

Fluency

H4: The frequency of Immersive Reader use will significantly predict the degree of enhancement in pronunciation accuracy among learners in the experimental group.

Fluency measures (e.g., pauses, speech rate) improved more substantially in the experimental group (t(114)=4.21, p<.001) than in the control group (t(114)=2.02, p=0.052).

FINDINGS AND DISCUSSION

Interpretation of Findings

The study's findings offer solid proof of Microsoft Immersive Reader's (IR) ability to improve EFL learners' pronunciation accuracy. Segmental, suprasegmental, and speech fluency were the three aspects of pronunciation which the experimental group significantly improved, whereas the control group show minimal change in improving pronunciation accuracy.

These findings support the hypotheses and show Immersive Reader significantly affects language learners' pronunciation accuracy.

Pedagogical Implications

IR equipped with multi-functions simultaneously enables language teachers to provide learners with more effective and engaging lessons. While syllable segmentation and text highlighting enable real-time phonological processing, the Read Aloud feature offers instant auditory models. IR's Read Aloud feature provides immediate auditory modeling, while syllable segmentation and text highlighting allow real-time phonological processing. Thanks to that, teachers can integrate IR into oral reading fluency task sessions, pronunciation drills, or self-access practice, especially for learners with limited exposure to native-like input. The findings also suggest that IR can help standardize pronunciation developmental gains across learners, reducing outcome variability.

CONCLUSION AND LIMITATIONS

Overall, the results supported Hypotheses confirming that the experimental group using Immersive Reader achieved significantly better outcomes in segmental, suprasegmental, and speech fluency-related pronunciation features compared to the control group.

This study investigated the impact of Microsoft Immersive Reader on the phonological accuracy of EFL learners through a quasi-experimental design. The results provide strong empirical support for using Immersive Reader to enhance segmental pronunciation, prosodic features, and speech fluency.

Integrating IR's into classroom presents a significant advancement in pronouncing English. Through interactive, visual and aural functions, Immersive Readers helps EFL learners study English more efficiently and enjoyably. Thus, these findings contribute to the growing body of CALL research and pedagogical practices in SLA.

Despite limitations, including its short duration and particular sample characteristics, the study paves the way for further investigations into the long-term efficacy and comparative impact of IR on other AI-assisted techniques. This study shows how promising Immersive Reader is as a potent supplement to the EFL teaching resource.

REFERENCES

Derwing, T. M., & Munro, M. J. (2015). Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research. John Benjamins Publishing Company.

Foote, J. A., Holtby, A. K., &Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. TESL Canada Journal, 29(1), 1–22.

Kormos, J. (2006). Speech production and second language acquisition. Routledge.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. TESOL Quarterly, 39(3), 369–377.

Mayer, R. E. (2001). Multimedia learning. Cambridge University Press.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. System, 34(4), 520–531.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), Cognition and second language instruction(pp. 3–32). Cambridge University Press.

Zhang, Y., & Graham, S. (2020). The effects of digital feedback tools on EFL pronunciation. System, 95, 102367. https://doi.org/10.1016/j.system.2020.102367

****