The Effect of Text-To-Speech Applications on Listening Comprehension with an Analysis of Behavioral Engagement:A Quasi-Experimental Study

Neneng Avrianti[1], Ikhsanudin[2], Dwi Riyanti[3]

Master's Program in English Language Education, Language and Art Education Department, Teacher Training and Education Faculty, Tanjungpura University, Pontianak, Indonesia

DOI: https://doi.org/10.47772/IJRISS.2026.100601334

Received: 02 July 2026; Accepted: 07 July 2026; Published: 18 July 2026

ABSTRACT

This study examines the effect of a Text-to-Speech (TTS) application on senior high school students' listening comprehension skills and analyzes their behavioral engagement during instruction. Using a quasi-experimental design with a non-equivalent control group, the study compared an experimental group using TTSMaker with a control group receiving traditional listening instruction. Data were collected through pre-test and post-test listening comprehension tests and Time-on-Task (ToT) observations across four instructional sessions. ANCOVA results revealed a significant difference between groups after controlling for pre-test scores ($F(1,57) = 15.857$, $p < .001$, $d = 1.03$). Descriptive analysis of behavioral engagement indicated a moderate-to-high engagement level ($M = 78.89\%$) with considerable variability across students; however, the correlation between engagement and post-test scores was non-significant ($r = -.228$, $p = .363$). These findings suggest that TTS applications effectively enhance listening comprehension, yet observable behavioral engagement does not necessarily reflect deeper cognitive processing. The study provides practical implications for integrating TTS technology in English language teaching.

Keywords: Text-To-Speech, Listening Comprehension, Behavioral Engagement, Time-On-Task, Quasi-Experimental Design, EFL

INTRODUCTION

Listening comprehension constitutes a foundational skill in foreign language acquisition, yet EFL classrooms, particularly in resource-limited contexts, face persistent challenges in providing adequate, level-appropriate auditory input (Field, 2009; Richards, 2008). Conventional instructional materials, such as CD-based recordings or teacher-led reading, often lack flexibility in speech rate, pronunciation consistency, and repetition opportunities, potentially increasing extraneous cognitive load and impeding comprehension (Sweller, 2011; Vandergrift & Goh, 2012).

Recent advances in Artificial Intelligence (AI) have introduced Text-to-Speech (TTS) technology as a promising alternative for generating customizable listening materials. TTS applications convert written text into synthesized speech, offering features such as adjustable speech rate, voice selection, and repeated playback (Dida et al., 2023; Tan et al., 2021). These affordances align with Cognitive Load Theory (CLT), which posits that controllable input reduces extraneous cognitive demands and facilitates deeper processing (Sweller, 1994, 2011). Empirical studies have reported positive effects of TTS on listening comprehension (Al-Jarf, 2022; Chiang, 2019; Matsuda, 2017; Nurhamidah, 2024) and learner engagement (Amin, 2022; Huang & Liao, 2024).

However, several gaps persist in the literature. First, classroom-based quasi-experimental studies examining TTS effects on listening comprehension remain limited, particularly in Indonesian EFL contexts. Second, while engagement is widely recognized as a key factor in language learning, existing TTS research relies predominantly on self-reported measures, leaving a paucity of objective, observational data on students' actual classroom behavior (Raffoul & Jaber, 2023; Wei, 2023). Third, the assumed positive relationship between behavioral engagement and learning outcomes has rarely been critically examined within TTS-supported instruction. Fredricks et al. (2004) distinguish between behavioral, emotional, and cognitive engagement, suggesting that observable participation may not necessarily reflect deeper cognitive processing.

This study addresses these gaps by investigating: (1) the effect of TTS-supported instruction on listening comprehension compared to traditional instruction; (2) the extent and nature of students' behavioral engagement during TTS-supported listening activities; and (3) the relationship between behavioral engagement and listening comprehension achievement. By integrating objective behavioral observation with standardized comprehension assessment, this study offers a nuanced understanding of how TTS technology functions in authentic classroom settings.

METHOD

This study employed a quasi-experimental design with a non-equivalent control group (Creswell, 2012). Two intact tenth-grade classes were assigned to either the experimental group (n = 30) or the control group (n = 30) based on school availability. Both groups completed a pre-test listening comprehension assessment, participated in four instructional sessions, and completed an identical post-test.

Participants

Participants were 60 tenth-grade students from SMAN 05 Pontianak, Indonesia. The experimental group (30 students) and control group (30 students) were drawn from existing classes without random assignment. All participants had studied English for approximately six years and had no prior exposure to TTS technology in formal instruction.

Instruments

Listening Comprehension Tests. Parallel pre-test and post-test instruments were developed, each comprising 20 multiple-choice items assessing four listening comprehension aspects: identifying main ideas, recognizing specific details, making inferences, and evaluating spoken texts (Brown, 2003; Richards, 2008). Both tests used equivalent formats and cognitive levels but different audio texts (pre-test: Cristiano Ronaldo, Michael Jordan; post-test: Serena Williams, Usain Bolt) to minimize practice effects. KR-20 reliability coefficients were .79 (pre-test) and .86 (post-test).

Time-on-Task Observation. Behavioral engagement was measured using a structured Time-on-Task (ToT) observation protocol adapted from Fisher et al. (2015) and Spanjers et al. (2008). Video recordings of all four experimental group sessions were analyzed using momentary time-sampling at 2-minute intervals (15 intervals per 30-minute session). At each interval, each student was coded as on-task (1) or off-task (0) based on predefined operational criteria. Absent students were coded as 0 for all intervals of the missed session. Engagement scores were calculated as: (Total On-Task Observations ÷ Total Possible Observations) × 100.

TTS-Generated Audio Materials. All experimental group audio materials were generated using TTSMaker (https://ttsmaker.com) with standardized parameters: English US language, male voice (Christ), 0.9× speech rate, normal pitch, 100% volume, 300ms pause time, and MP3 output format.

Procedure

The study was conducted over four weeks (February–March 2026). The experimental group received TTS-supported listening instruction across pre-listening, while-listening, and post-listening stages. Sessions 1–2 focused on collaborative TTS familiarization; Sessions 3–4 involved structured individual listening tasks with worksheets. The control group received equivalent content through teacher-led reading without audio technology. Both groups completed the post-test one week after the final session.

Data Analysis

Analysis of Covariance (ANCOVA) was used to examine the effect of instructional method on post-test listening comprehension, with pre-test scores as the covariate. Assumption tests included normality of residuals (Shapiro-Wilk), homogeneity of variance (Levene's test), and homogeneity of regression slopes. For RQ3, both Spearman's rank-order and Pearson's product-moment correlations were computed; the primary analysis used complete cases ($n = 18$) to eliminate attendance bias. Effect sizes were reported as Cohen's d and partial eta squared (η^2p). All analyses were conducted using SPSS with $\alpha = .05$.

RESULTS

Effect of TTS on Listening Comprehension

Preliminary assumption testing confirmed the validity of ANCOVA. The correlation between pre-test and post-test scores was significant ($r = .420$, $p = .001$, $N = 60$), satisfying the covariate prerequisite. The homogeneity of regression slopes assumption was met ($F(1, 56) = 0.229$, $p = .634$), and Levene's test indicated equal variances ($F(1, 58) = 0.560$, $p = .457$). Residuals were normally distributed for both groups (Shapiro-Wilk $p > .05$). Descriptive statistics showed comparable baseline performance (experimental: $M = 10.63$, $SD = 3.80$; control: $M = 10.37$, $SD = 4.85$). Following the intervention, the experimental group achieved a raw gain of 5.10 points (48.0% increase), while the control group gained 2.10 points (20.3% increase).

ANCOVA results revealed a significant effect of teaching method on post-test scores after controlling for pre-test differences ($F(1, 57) = 15.857$, $p < .001$, partial $\eta^2 = .218$). The adjusted mean difference was 3.172 points (95% CI [1.577, 4.767]), with a large effect size (Cohen's $d = 1.03$). The covariate (pre-test) was also significant ($F(1, 57) = 14.662$, $p < .001$, partial $\eta^2 = .205$), explaining 20.5% of variance. The overall model accounted for 35.6% of variance in post-test scores ($R^2 = .356$).

Table1.1. ANCOVA Results for Listening Comprehension

Source	Type III SS	df	MS	F	p	Partial η^2
Corrected Model	299.469	2	149.734	15.749	< .001	.356
Pre-test (covariate)	139.402	1	139.402	14.662	< .001	.205
Teaching Method	150.761	1	150.761	15.857	< .001	.218
Error	541.931	57	9.508			

Exploratory analysis of aspect-level gains in the experimental group showed the largest improvement in Specific Details (M gain = +1.37), followed by Evaluation (+1.27), Main Idea (+1.23), and Inference (+1.23), all out of a maximum score of 5 per aspect.

Behavioral Engagement During TTS-Supported Instruction

The overall mean engagement score was 78.89% ($SD = 25.72\%$, range: 21.67%–100%). However, this figure conflated attendance and actual behavior. Among students present in all four sessions ($n = 18$), mean engagement was 95.83% ($SD = 8.77\%$), with on-task rates remarkably stable across sessions: 92.00% (S1), 98.61% (S2), 97.53% (S3), and 98.48% (S4). The apparent session-level fluctuation in aggregate engagement (76.67%–87.78%) was driven by attendance variation (83.33% in S1 to 73.33% in S4) rather than behavioral change.

Table1.2. Distribution of Engagement Levels

Category	Range	n	Percentage
High	≥ 80%	16	53.3%
Moderate	60–79%	7	23.3%
Low	< 60%	7	23.3%

Notably, all students in the "low engagement" category had multiple absences (1–3 sessions), indicating that low scores primarily reflected non-attendance rather than behavioral disengagement. Among complete attendees, the

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

lowest engagement score was 73.33%. Qualitative analysis of on-task behaviors revealed a shift across sessions: task-completion behaviors emerged only in Sessions 3–4 (66.7% and 50.0% of intervals), reflecting the transition from TTS familiarization to structured worksheet tasks. Peer discussion, encouraged during collaborative Sessions 1–2, disappeared entirely in Sessions 3–4 (0%). Note-taking was absent across all sessions (0%).

Relationship Between Engagement and Listening Comprehension

Correlation analysis revealed no significant relationship between behavioral engagement and post-test listening comprehension scores. For the full sample (N = 30), Spearman's rs = –.095 (p = .619) and Pearson's r = –.047 (p = .804) were both non-significant. For complete cases (n = 18), Pearson's r = –.228 (p = .363, 95% CI [–.671, +.215]), with only 5.2% shared variance (r^2 = .052).

Table1.3. Correlation Between Engagement and Post-Test Score

Analysis	N	r	p (2-tailed)	95% CI
Spearman (all students)	30	–.095	.619	—
Pearson (all students)	30	–.047	.804	—
Pearson (complete cases)	18	–.228	.363	[–.671, +.215]

The negative direction is negligible and should not be interpreted as a meaningful inverse relationship. Post-hoc power analysis (G*Power 3.1) indicated limited power (0.25–0.30) to detect medium-sized correlations with n = 18, suggesting the null finding may reflect insufficient statistical power rather than the genuine absence of a relationship.

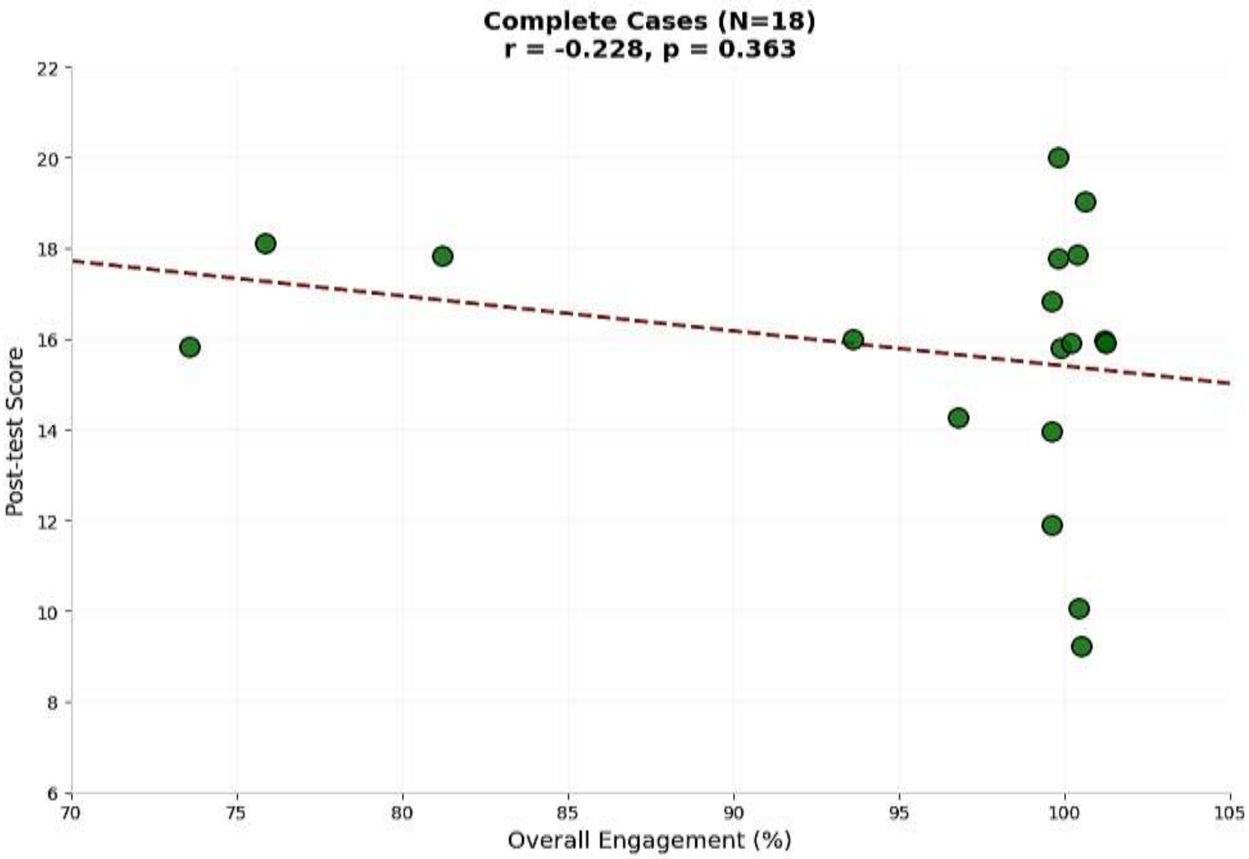

Figure 1.1: Scatterplot of Engagement and Listening Comprehension (Complete Cases, N = 18).

The dashed regression line indicates a negligible negative slope (r = –.228, p = .363). Notable cases: S07 (100% engagement, score 9) and S25 (50% engagement, score 20) exemplify the disconnect between observable participation and cognitive performance. The clustering of multiple students at 100% engagement despite spanning the full range of post-test scores (9–20) illustrates the ceiling effect that restricted correlational power.

DISCUSSION

Relationship Between Engagement and Listening Comprehension

The significant ANCOVA result supports the hypothesis that TTS-supported instruction enhances EFL listening comprehension compared to traditional teacher-led instruction. The large effect size (d = 1.03) suggests practical significance, potentially shifting a student from the 50th to approximately the 85th percentile in performance relative to traditional instruction. This finding aligns with previous studies documenting TTS benefits for listening comprehension (Al-Jarf, 2022; Chiang, 2019; Matsuda, 2017; Nurhamidah, 2024)

From a Cognitive Load Theory perspective (Sweller, 1994, 2011), TTS technology functions as a cognitive support tool by enabling learner control over speech rate, facilitating repeated exposure, and providing consistent pronunciation, features that reduce extraneous cognitive load and free working memory resources for essential comprehension processes. The exploratory finding that Specific Details showed the largest gain (+1.37) is consistent with this interpretation: controllable input appears particularly beneficial for bottom-up processing of discrete linguistic elements.

However, the quasi-experimental design with intact classes limits causal inference. The observed effect may reflect the combined influence of TTS technology, instructional structure, novelty effects, and unmeasured classroom variables. The significant covariate effect (pre-test scores) underscores that prior proficiency remains a strong predictor of outcomes, accounting for 20.5% of variance independent of instructional method.

The Technology-Engagement-Learning

The central and unexpected finding of this study is the disconnect between behavioral engagement and learning outcomes. Despite high observable engagement among present students (92–98% on-task rates), behavioral engagement did not correlate with listening comprehension performance. This reveals what we term a technology-engagement-learning: TTS improved input accessibility and produced measurable comprehension gains, yet visible participation did not predict individual achievement.

Several factors explain this paradox. First, a severe ceiling effect in engagement scores, in which 43.3% of students achieved 100% engagement, created tied ranks that reduced statistical power to detect correlations. Second, post-test scores were also concentrated at the upper end (M = 15.73/20, with 66.7% scoring $\geq$ 16), restricting range. Third, and most substantively, behavioral engagement does not equal cognitive engagement (Fredricks et al., 2004). Students may have been physically present and on-task without engaging in the deeper mental elaboration, inference-making, and schema activation required for successful listening comprehension (Vandergrift & Goh, 2012). The case of S07 (100% engagement, post-test score 9) and S25 (50% engagement, post-test score 20) exemplifies this disconnect: observable participation was neither necessary nor sufficient for high performance.

This finding challenges the common assumption that Time-on-Task serves as a reliable proxy for learning (Godwin et al., 2016, 2021). As Reschly et al. (2020) argue, behavioral engagement is one of four interrelated dimensions; effective intervention requires attention to cognitive and affective dimensions as well. The disappearance of peer discussion in later sessions and the complete absence of note-taking further suggest that TTS-supported instruction may have facilitated behavioral compliance without promoting deeper cognitive strategies.

Implications for Pedagogy

These findings carry important implications for TTS integration in EFL listening instruction. First, TTS should be positioned as instructional support rather than a standalone solution. Its effectiveness depends on purposeful pedagogical integration that channels technology use toward specific learning objectives. Second, teachers should not rely solely on observable engagement as an indicator of learning quality; structured comprehension checks, reflection tasks, and strategy instruction are necessary to ensure cognitive processing. Third, initial familiarization sessions are essential to reduce cognitive load when introducing new technology. Finally, the declining attendance across sessions (83.33% to 73.33%) suggests that sustained implementation requires attention to motivational and logistical factors beyond technology itself.

LIMITATIONS AND FUTURE DIRECTIONS

This study has several limitations. The quasi-experimental design with intact classes precludes causal attribution. The short treatment duration (four sessions) may not capture long-term effects or distinguish novelty effects from sustained learning gains. The binary ToT coding (on-task/off-task) oversimplifies engagement complexity, and the 2-minute observation interval may miss rapid attentional fluctuations. The correlational analysis for RQ3 was underpowered due to the reduced complete-cases sample ($n = 18$). Finally, the use of TTSMaker's free version may limit generalizability to more advanced TTS systems. Future research should employ randomized controlled designs with longer interventions, incorporate multidimensional engagement measures (cognitive, emotional, behavioral), use finer-grained observation protocols, and investigate individual differences (digital literacy, language proficiency) that may moderate TTS effectiveness.

CONCLUSION

This study examined the effect of Text-to-Speech (TTS) technology on EFL listening comprehension and the relationship between behavioral engagement and learning outcomes. Three key findings emerge. First, TTS-supported instruction produced significantly higher listening comprehension gains than traditional instruction ($F(1, 57) = 15.857$, $p < .001$, $d = 1.03$), supporting the pedagogical value of controllable, consistent auditory input in reducing cognitive load and facilitating bottom-up processing. Second, behavioral engagement during TTS-supported instruction was moderate to high overall ($M = 78.89\%$), but this aggregate figure masked attendance-driven variation; among present students, on-task rates were consistently high (92–98%) across sessions. Third, and most critically, behavioral engagement showed no significant correlation with listening comprehension achievement ($r = –.228$, $p = .363$), revealing a technology-engagement-learning: TTS improved comprehension without observable engagement predicting individual outcomes.

This paradox challenges the common assumption that Time-on-Task serves as a reliable proxy for learning. From a theoretical standpoint, it supports Fredricks et al. (2004) multidimensional engagement framework by demonstrating that behavioral participation alone does not guarantee cognitive processing. Students may remain physically on-task, attending to audio, completing worksheets, without engaging in the deeper mental elaboration required for comprehension. The cases of S07 (100% engagement, lowest score) and S25 (50% engagement, perfect score) exemplify this disconnect and caution against equating visible compliance with learning quality.

Practically, these findings suggest that TTS should be integrated as instructional support rather than a standalone solution. Teachers should pair TTS with structured tasks, comprehension checks, and instruction in metacognitive strategies to ensure cognitive engagement. Institutions should invest in teacher professional development and evaluate technology implementation through classroom-based research, while students should develop active listening strategies beyond reliance on repeated playback. The absence of note-taking and the disappearance of peer discussion in later sessions further indicate that TTS implementation requires intentional design to sustain collaborative, cognitively demanding activities.

Methodologically, this study highlights the importance of disaggregating attendance from behavioral engagement in Time-on-Task research. The finding that aggregate engagement trends were driven by attendance variation (83.33% to 73.33%) rather than behavioral change underscores a confounding factor often overlooked in classroom studies. The complete-cases analysis provides a model for controlling this bias, though reduced sample size limited statistical power.

Several limitations temper these conclusions. The quasi-experimental design with intact classes precludes causal inference; the short treatment duration (four sessions) may not capture long-term effects; and the binary, fixed-interval observation protocol may have missed rapid engagement fluctuations. Future research should employ randomized designs, finer-grained observation coding, and multidimensional engagement measures to clarify the conditions under which technology-enhanced instruction translates into measurable learning gains.

In sum, this study contributes empirical evidence for TTS integration in Indonesian EFL classrooms while cautioning that technology improves input accessibility more reliably than it ensures learning. The value of TTS lies not in keeping students visibly on-task, but in supporting the cognitive processes, controlled attention, repeated exposure, and reduced extraneous load that underlie comprehension. For resource-limited contexts where native-speaker audio is scarce, freely available TTS tools offer a practical alternative, but only purposeful pedagogical integration can transform accessible input into meaningful learning.

REFERENCES

1. Al-Jarf, R. (2022). Text-to-speech software for promoting EFL freshman students' decoding skills and pronunciation accuracy. Journal of Computer Science and Technology Studies, 4(2), 19–30. https://doi.org/10.32996/jcsts
2. Amin, E. A. R. (2022). Using repeated-reading and listening-while-reading via text-to-speech APPs. in developing fluency and comprehension. World Journal of English Language, 12(1), 211–220. https://doi.org/10.5430/wjel.v12n1p211
3. Brown, D. H. (2003). Language assessment: Principles and classroom practices. Longman.
4. Chiang, H.-H. (2019). A comparison between teacher-led and online text-to-speech dictation for students' vocabulary performance. English Language Teaching, 12(3), 77. https://doi.org/10.5539/elt.v12n3p77
5. Creswell, J. W. (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.). Pearson Education.
6. Dida, H. A., Chakravarthy, D., & Rabbi, F. (2023). ChatGPT and big data: Enhancing text-to-speech conversion. Mesopotamian Journal of Big Data, 2023, 31–35. https://doi.org/10.58496/mjbd/2023/005
7. Field, J. (2009). Listening in the language classroom. Cambridge University Press. https://api.pageplace.de/preview/DT0400.9780511474491_A25045246/preview-9780511474491_A25045246.pdf
8. Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (2015). Teaching behaviors, academic learning time, and student achievement: An overview. Journal of Classroom Interaction, 50(1), 6–24. https://files.eric.ed.gov/fulltext/EJ1100414.pdf
9. Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement : Potential of the Concept , State of the Evidence. Review of Educational Research, 74(1), 59–109. https://doi.org/doi.org/10.3102/00346543074001059
10. Godwin, K. E., Almeda, M. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S., & Fisher, A. V. (2016). Off-task behavior in Elementary School children. Learning and Instruction, 44, 128–143. https://doi.org/10.1016/j.learninstruc.2016.04.003
11. Godwin, K. E., Seltman, H., Almeda, M., Davis Skerbetz, M., Kai, S., Baker, R. S., & Fisher, A. V. (2021). The elusive relationship between time-on-task and learning: Not simply an issue of measurement. Educational Psychology, 41(4), 502–519. https://doi.org/10.1080/01443410.2021.1894324
12. Huang, Y.-C., & Liao, L.-C. (2024). A study of text-to-speech (TTS) in children's English learning. Teaching English with Technology, 15(1), 14–30. https://www.ceeol.com/search/article-detail?id=166912http://www.tewtjournal.org
13. Matsuda, N. (2017). Evidence of effects of text-to-speech synthetic speech to improve second language learning. JACET Journal, 61, 149–164.
14. Nurhamidah, N. (2024). The effect of text-to-speech software on improving students' listening skill [UIN Suska Riau]. https://repository.uin-suska.ac.id/79100/2/THESIS NOVIA NURHAMIDAH.pdf
15. Raffoul, S., & Jaber, L. (2023). Text-to-speech software and reading comprehension : The impact for students with learning disabilities. Canadian Journal of Learning and Technology, 49(2).
16. Reschly, A. L., Pohl, A. J., & Christenson, S. L. (2020). Student engagement: Effective academic , behavioral , cognitive , and affective interventions at school. Springer. https://doi.org/10.1007/978-3-030-37285-9 ©
17. Richards, J. C. (2008). Teaching listening and speaking from theory to practice. Cambridge University Press.
18. Spanjers, D. M., Burns, M. K., & Wagner, A. R. (2008). Systematic direct observation of time on task as a measure of student engagement. Assessment for Effective Intervention, 33(2), 120–126.

https://doi.org/10.1177/1534508407311407

19. Sweller, J. (1994). COGNITIVE LOAD THEORY , LEARNING DIFFICULTY , AND INSTRUCTIONAL DESIGN. 4, 295–312.
20. Sweller, J. (2011). Cognitive load theory. In The Psychology of Learning and Motivation (Vol. 55, pp. 37–76). Elsevier Inc. https://doi.org/10.1016/B978-0-12-387691-1.00002-8
21. Tan, X., Chen, M., Li, B., Liu, Y., Qin, T., Zhao, S., & Liu, T.-Y. (2021). AdaSpeech: Adaptive text to speech for custom voice. ICLR, 1–10. http://arxiv.org/abs/2103.00993
22. Vandergrift, L., & Goh, C. C. M. (2012). Teaching and learning second language listening: Metacognitive in action. Routledge.
23. Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. Frontiers in Psychology, 14, 1–14. https://doi.org/10.3389/fpsyg.2023.1261955