MACCA: JOURNAL OF LINGUISTIC APPLIED RESEARCH
Vol.3, No.2, 2026: Page 735-747

# The Influence Of Voice Interaction With The Virtual Assistant Chatgpt On Students' Motivation In Learning English At MTS Assalam Sidrap

**Tony Putra Hamid, Andi Sadapotto, Syahrir L**

[123]Universitas Muhammadiyah Sidenreng Rappang

Email: tonyputrahamidhamid@gmail.com

A R T I C L E I N F O

*Keywords:. ChatGPT; voice interaction; English learning motivation; EFL learners; junior secondary education*

©2026 Tony Putra Hamid

: This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License.

A B S T R A C T

*This study examined changes in students' motivation to learn English following supervised voice interaction with ChatGPT at MTs Assalam Sidrap, Indonesia. A quantitative pre-experimental one-group pretest–posttest design involved all 22 students in the selected eighth-grade class. Four treatment meetings progressed from demonstration to simple questions, guided conversations, and freer oral interaction using a researcher-supervised laptop. Motivation was measured with a 25-item, five-point questionnaire covering interest, participation, confidence, enjoyment, and effort and engagement. The source thesis reported expert review and internal-consistency coefficients of .826 at pretest and .882 at posttest. Motivation scores increased descriptively from 85.86 (SD = 10.93) to 89.55 (SD = 11.90), with a mean paired increase of 3.68 points. Thirteen students improved, seven declined, and two remained unchanged. Confidence and enjoyment showed the largest descriptive increases. However, the overall difference was not statistically significant, t(21) = 1.511, p = .146, 95% CI [−1.39, 8.75], Cohen's dz = 0.32. The null hypothesis was therefore not rejected. These findings suggest a modest positive descriptive tendency, with considerable individual variation, rather than a demonstrated overall motivational influence. Interpretation is limited by the small single-class sample, four-meeting intervention, absence of a comparison group, and reliance on self-report. Longer controlled studies should examine sustained motivation, individual exposure, and the distinction between enjoyment of AI interaction and continuing effort to learn English.*

## INTRODUCTION

Motivation influences whether learners begin an English task, continue after difficulty, and seek opportunities to communicate beyond required classroom activities. Vocabulary and pronunciation uncertainty can discourage participation. Junior secondary learners may know an answer but hesitate to express it publicly. Understanding how learning environments relate to motivation requires attention to this distinction. A medium that attracts curiosity may change the immediate experience of a lesson without necessarily changing sustained effort, persistence, or broader commitment to learning.At MTs Assalam Sidrap, the source thesis identifies limited confidence, participation, and motivation as reasons for investigating voice interaction with ChatGPT (Hamid, 2026). The school operates in a boarding-school context where students are not allowed to use personal smartphones. Consequently, the investigation used a laptop under researcher supervision. This arrangement is central to interpreting the intervention: students experienced guided access to a conversational system within school rules. They did not receive unrestricted individual access to an application. Evidence from this setting can inform practical decisions about supervised instruction, while its transferability depends on how similar other schools are in access, routines, and learner needs.

Voice interaction is relevant because it provides an additional occasion to produce and respond to spoken English. In the reported activities, students asked questions, answered prompts, repeated expressions, and engaged in short conversations. Such tasks may make language practice feel more approachable when learners can attempt an utterance without an immediate public evaluation by classmates. Nevertheless, the availability of a responsive conversational system does not determine how students experience the interaction. Task difficulty, comprehensibility, waiting time, teacher guidance, and opportunities to try again can all matter. The present study investigates motivation following this instructional arrangement, rather than evaluating every possible use of conversational AI.

Theoretical perspectives help clarify what a motivational response might involve. Deci and Ryan (2000) connect motivation with the conditions supporting autonomy, competence, and relatedness. Dörnyei (2001) emphasizes teachers' practical role in creating and sustaining motivation within language classrooms. These perspectives suggest that choice, manageable challenges, and supportive interaction deserve consideration when designing speaking activities. They do not imply that conversation with a system automatically satisfies learners' needs. A student might enjoy selecting a topic but still require linguistic support to complete the conversation. Another might perform a task successfully while remaining uncomfortable with the technology or uninterested in independent practice afterward.

Earlier chatbot scholarship provides a reason to examine these possibilities empirically. Fryer and Carpenter (2006) discussed bots as supplementary language-learning tools, while Fryer et al. (2017) investigated the development of interest with chatbot and human task partners. More recent work has examined voice-based interactions and conversational task design, including Tai and Chen (2023) and Yang et al. (2022). These studies concern related learning opportunities, but their systems and outcomes differ. Willingness to communicate, task interest, perceived enjoyment, and overall motivation should not be treated as interchangeable. A useful investigation therefore defines what its questionnaire measures and distinguishes those scores from demonstrated improvement in speaking proficiency.

Regional research also provides relevant context. Wahyuddin et al. (2025) explored AI use in EFL learning through a digital-literacy perspective, and Handayani et al. (2026) examined opportunities and challenges in AI-supported English teaching. Fatmawati et al. (2026) investigated TalkPal AI and speaking skills in a different Sidrap school. These publications demonstrate local interest in conversational technology and classroom experience. However, findings about university perceptions, primary-school teaching, or senior secondary speaking performance cannot establish the motivational effect of supervised ChatGPT voice activities for eighth-grade students. The present school context and measured construct therefore warrant a separate, explicitly bounded analysis.

Technology-supported motivation has also been investigated through applications that do not involve conversational AI. Junandar et al. (2026) examined mobile learning and English motivation at an Islamic junior secondary school, while Azis et al. (2026) addressed Wordwall-supported motivation. Abdul Jalil Jum'uatullaila et al. (2025) studied Quizizz-supported English learning. These reports provide comparisons concerning instructional engagement and digital participation. Their different tasks, intervention periods, instruments, and research designs limit direct comparisons of numerical gains. Rather than assuming that all digital applications share one motivational effect, teachers need evidence about the specific learning activity, available access, and outcomes relevant to their own students.

Other local scholarship reminds educators that motivation is embedded in instructional relationships.

Amalia et al. (2024) investigated project-based public speaking, and Saleh et al. (2026) examined supported mobile vocabulary learning. Hermansyah et al. (2026) addressed digital public-speaking materials, while Hermansyah (2026) reviewed opportunities and challenges in secondary-school digital speaking instruction. Such work broadens the instructional context beyond a single platform. ChatGPT voice interaction can be considered alongside peer communication, teacher explanation, and meaningful speaking tasks. Investigating it as a supplementary environment avoids assuming that technological novelty is itself the educational objective or that an engaging activity necessarily produces a durable change in motivation. The available thesis offers an opportunity to report both promising descriptive patterns and an overall result that did not reach statistical significance. This distinction is important because favorable averages can obscure divergent individual responses. A rise in the class mean may coexist with decreases for several learners, and increased confidence may coexist with little change in effort. Conversely, a nonsignificant test does not demonstrate that every learner was unaffected or that the approach has no possible educational value. Transparent reporting should retain these different levels of evidence, allowing readers to see what changed, how uncertain the estimate remains, and which explanations require additional research.

Accordingly, this article addresses three questions: how did total motivation scores change following the four meetings, how were individual changes and motivation categories distributed, and which questionnaire indicators showed the largest descriptive changes? The investigation uses the thesis's paired questionnaire records and preserves its inferential conclusion. It does not claim to measure speaking achievement, diagnose anxiety, or establish long-term persistence. Its contribution is a detailed account of a supervised classroom intervention in a setting with restricted personal-device access. By integrating descriptive tables with a confidence interval and clearly stated design limitations, the article supports informed instructional judgment and identifies priorities for stronger subsequent studies.

## Literature Review

Motivation in language learning involves more than enjoying a lesson. Riyanti (2019) discusses its importance in EFL settings, while Dörnyei (2001) connects motivational practice with classroom conditions and sustained participation. In the present study, motivation is operationalized through interest, participation, confidence, enjoyment, and effort and engagement. These indicators describe related emotional and behavioral orientations, but they need not change together. A student might become more confident during a conversation without increasing study effort afterward. This possibility makes indicator-level reporting useful, provided that descriptive differences are not presented as independently established treatment effects or as validated evidence for constructs.

Self-determination theory provides an interpretive framework. Deci and Ryan (2000) distinguish forms of motivation and discuss psychological needs, while Annamalai et al. (2023) examine chatbot-supported English learning through that perspective. Guided conversational activities may offer choices, achievable challenges, and responsive exchanges. However, those features are potential supports rather than direct measures of need satisfaction. The thesis did not administer a dedicated autonomy, competence, or relatedness scale. Consequently, self-determination theory informs possible explanations and future measurement decisions; it does not establish a mechanism through which the reported score changes occurred.

### Conversational systems and spoken participation

Fryer and Carpenter (2006) positioned bots as supplementary resources for language practice. Their early discussion helps distinguish opportunities to converse from the instructional work required to make those conversations useful. Chen et al. (2020) review wider educational applications of AI, while Kohnke et al. (2023) discuss ChatGPT in language teaching and learning. These publications support examining educational affordances alongside limitations. The present article makes no assumption that text-based chatbots, voice assistants, and contemporary generative conversational systems are equivalent. A study's findings remain linked to its actual tasks, interface, supervision, and learner population.

Voice-specific scholarship offers closer comparisons. Tai and Chen (2023) examined adolescent EFL learners' willingness to communicate with Google Assistant, whereas Yang et al. (2022) investigated a task-based voice chatbot as an English conversation partner. These studies show why the design of spoken tasks and learners' perceptions deserve attention. Their outcomes do not directly predict a composite motivation score in another school. Fryer et al. (2017) further distinguish stimulating interest from sustaining it. This distinction is particularly relevant to a four-meeting intervention, where an enjoyable experience may represent initial curiosity rather than an enduring change in learning commitment.

### Regional evidence on AI and digital instruction

Wahyuddin et al. (2025) provide local evidence about EFL learners' experiences with AI and digital literacy, while Handayani et al. (2026) address classroom opportunities and practical challenges. Fatmawati et al. (2026) examine AI-supported speaking with TalkPal AI, a different system and educational setting. Their relevance lies in contextualizing potential participation, feedback, and confidence benefits. Their findings cannot establish what occurred in the present intervention. Hermansyah's (2026) review similarly draws attention to task design and access conditions in digital speaking instruction, supporting documentation of the supervised laptop arrangement rather than treating all learners as independently connected users.

Junandar et al. (2026) and Azis et al. (2026) connect digital learning activities with motivation in regional secondary settings. Abdul Jalil Jum'uatullaila et al. (2025) address Quizizz-supported learning, and Wahyuddin et al. (2024) investigate digital literacy within blended learning. These studies establish a broader context for technology-supported instruction without proving that technology alone produces motivation. Azizah et al. (2026) examine smartphone perceptions in another Sidrap school, highlighting the relevance of access and classroom use. Comparisons with the present study should therefore consider differences in school policy, available devices, learner age, and the learning activities enabled by those devices.

### Measurement and the remaining research question

Evidence from non-AI approaches reinforces the importance of instructional context. Amalia et al. (2024) examine project-based speaking, Saleh et al. (2026) investigate mobile vocabulary support, and Hermansyah et al. (2026) develop digital speaking materials. Such studies suggest possible complementary activities, rather than competitors that can be ranked using unrelated outcome measures. Within the present investigation, questionnaire reliability also requires careful interpretation. Cronbach (1951) provides the foundation for internal-consistency estimation, but a satisfactory alpha does not by itself establish content coverage, dimensionality, sensitivity to change, or equivalence between pretest and posttest interpretations. The research gap is a bounded empirical question about supervised voice interaction and motivation in the selected eighth-grade cohort. The thesis provides paired total scores, indicator summaries, and an inferential analysis, making it possible to distinguish mean change from variation between learners. Lakens (2013) supports transparent reporting of standardized differences alongside their calculation basis. The study's contribution is therefore not the assumption of a positive AI effect, but a documented examination of whether motivation changed and how uncertain that estimate remains. This framework allows a nonsignificant overall result to be interpreted alongside descriptive patterns without converting either into an unsupported causal conclusion.

The absence of a statistically significant average change also leaves room for different individual experiences. Reporting those experiences separately from aggregate estimates can prevent a class mean from concealing students whose motivation decreases during the same activities.

## Method

This article adapts the quantitative pre-experimental investigation reported by Hamid (2026). The study used a one-group pretest–posttest design at MTs Assalam Sidrap, Talawe Village, Watang Sidenreng Subdistrict, South Sulawesi, Indonesia. The school population comprised 70 students: 20 in Grade VII, 22 in Grade VIII, and 28 in Grade IX. All 22 students in the selected eighth-grade class participated and completed both measurements. Class selection was not reported as randomized. Thus, the sample represented complete participation within the selected class, rather than random sampling across the school population. No control or comparison group was included.

### Intervention and procedure

Motivation was measured before and after four treatment meetings. The researcher introduced and demonstrated ChatGPT voice interaction in Meeting 1. Meeting 2 involved asking and answering simple questions; Meeting 3 used short conversations on assigned topics; Meeting 4 allowed freer interaction, including questions and opinions. Students used a researcher-supervised laptop because personal smartphones were prohibited in the boarding-school setting. Reported activities included oral responses, repetition, and immediate system replies. Table 1 summarizes this sequence. The source does not establish exact session lengths, individual speaking time, model version, complete prompts, or exposure outside the sessions; these details are not invented.

**Table 1. Design and four-meeting intervention**

| Stage | Reported activity |
|---|---|

| | |
|---|---|
| Design | One-group pretest–posttest; 22 eighth-grade students; no comparison group |
| Pretest | 25-item motivation questionnaire before the four meetings |
| Meeting 1 | Introduction and demonstration of ChatGPT voice interaction |
| Meeting 2 | Simple spoken questions and answers |
| Meeting 3 | Short guided conversations on assigned topics |
| Meeting 4 | Freer oral interaction, opinions and questions |
| Posttest | Repeat administration of the same motivation questionnaire |
| Access arrangement | Researcher-supervised laptop; personal smartphones prohibited |

*Note. Adapted from the source thesis. Session duration, individual speaking time, exact prompts and system version are not established.*

## Instrument and scoring

A 25-item questionnaire measured five motivation indicators, with five statements per indicator: interest, participation, confidence, enjoyment, and effort and engagement. Responses ranged from strongly disagree to strongly agree on a five-point scale. Positively worded statements were scored from one to five. Negatively worded items 3, 5, 8, 10, 13, 15, 18, 20, 23, and 25 were reverse-scored. Total scores could range from 25 to 125, and each indicator from five to 25. The same questionnaire was administered twice. Higher totals indicated greater reported motivation according to this instrument, not higher English proficiency or independently observed participation.

## Instrument quality

The source reports expert review of relevance, clarity, and alignment with the intended indicator. It states that all items received the highest rating and were accepted without revision. This is reported content-review evidence, rather than a new validation conducted for this article. Cronbach's alpha, estimated separately from the main-study administrations after reverse scoring, was .826 at pretest and .882 at posttest. These coefficients indicate acceptable internal consistency under the source's .70 criterion. Table 2 summarizes the instrument. A separate pilot study, factor analysis, measurement-invariance assessment, and established minimally important difference were not documented in the available account.

**Table 2. Questionnaire structure and reported internal consistency**

| Indicator or property | Items or value | Score range |
|---|---|---|
| Interest in learning English | 1–5 | 5–25 |
| Participation in learning activities | 6–10 | 5–25 |
| Confidence in English interaction | 11–15 | 5–25 |
| Enjoyment through voice interaction | 16–20 | 5–25 |
| Effort and engagement | 21–25 | 5–25 |
| Overall questionnaire | 25 items; five response options | 25–125 |
| Reverse-scored items | 3, 5, 8, 10, 13, 15, 18, 20, 23, 25 | 1↔5; 2↔4; 3 unchanged |
| Pretest Cronbach's alpha | .826; n = 22 | Not applicable |
| Posttest Cronbach's alpha | .882; n = 22 | Not applicable |

*Note. Blueprint, reverse-scoring rules and alpha coefficients are reported in the thesis. Alpha was estimated from the main-study administrations; it does not establish all aspects of validity.*

## Analysis and verification

Descriptive analyses included means, standard deviations, medians, ranges, category frequencies, and individual score changes. Categories followed the thesis: very low, 25–45; low, 46–65; moderate, 66–85; high, 86–105; and very high, 106–125. A Shapiro–Wilk test examined paired difference scores, followed by a two-sided paired-samples t-test at alpha .05. The mean difference, 95% confidence interval, and Cohen's dz were reported, with dz calculated using the standard deviation of paired differences (Lakens, 2013). The 22 displayed score pairs were independently recalculated for this article and reproduced the thesis's total-score inferential results. Reliability and indicator summaries remain identified as source-reported results.

## Interpretation and research safeguards

Indicator changes were examined descriptively; no unreported indicator-level significance tests were added. No interviews, observed anxiety scores, speaking assessments, or delayed outcomes were generated. The thesis acknowledges school permission, but detailed consent, assent, account arrangements, and formal ethics documentation require author confirmation before submission. Anonymous student codes were retained in analytical tables. Because the design lacked a comparison condition, results were interpreted as changes following the intervention rather than causal effects attributable exclusively to ChatGPT. Missing implementation details and possible changes in how students interpreted technology-specific questionnaire statements were retained as limitations.All recalculations used the printed anonymous pairings, and reported percentages retained the full denominator of twenty-two participating students.

## Results

The available analytical record contained pretest and posttest motivation totals for all 22 participants. There were no missing totals in the displayed paired table. Exposure remains uncertain. The pretest scores summed to 1,889, while posttest scores summed to 1,970. The resulting aggregate increase was 81 points across the cohort. Independent recalculation reproduced the reported means and paired-test statistics. The results below therefore retain the thesis's principal conclusion and distinguish verified total-score calculations from indicator and reliability summaries reported in the source.

The four-meeting sequence moved from demonstration toward freer conversation, as shown in Table 1. This describes the instructional opportunity provided, rather than a measured progression in motivation. Similarly, the questionnaire blueprint and reported reliability coefficients in Table 2 describe the measurement framework. The higher posttest alpha does not establish that motivation improved; it concerns consistency among questionnaire responses at that administration. No individual interaction-time records were available to connect score changes with different amounts of practice. Therefore, participant coverage should not be interpreted as evidence that all students received identical conversational exposure or experienced the activities in the same way.

## Overall descriptive change

The mean total motivation score increased from 85.86 to 89.55. The average paired difference, calculated before rounding, was 3.68 points. Subtracting the two displayed rounded means produces 3.69, a rounding difference rather than a discrepancy in the paired records. Pretest standard deviation was 10.93 and posttest standard deviation was 11.90. Thus, the posttest scores were somewhat more dispersed, despite the higher average. The median increased from 86.50 to 90.00. Table 3 presents these descriptive statistics together, allowing comparison of central tendency and dispersion without assuming that the average increase represented a uniform response across the class.

**Table 3. Descriptive statistics of total motivation scores**

| Statistic | Pretest | Posttest |
|---|---|---|
| n | 22 | 22 |
| Total | 1,889 | 1,970 |
| Mean | 85.86 | 89.55 |
| Standard deviation | 10.93 | 11.90 |
| Median | 86.50 | 90.00 |
| Minimum | 62 | 65 |
| Maximum | 105 | 110 |
| Mean per item | 3.43 | 3.58 |

*Note. Recalculated from the 22 paired totals. The mean paired change is 81/22 = 3.6818; subtraction of rounded displayed means may differ by .01.*

The minimum score increased from 62 to 65, while the maximum increased from 105 to 110. These values remained within the questionnaire's possible range of 25–125. Neither the higher minimum nor the higher maximum establishes that the same students occupied those positions at both administrations. They describe the distribution at each occasion. Expressed per questionnaire item, the overall mean moved from approximately 3.43 to 3.58 on the five-point response scale. This added calculation makes the magnitude easier to interpret, but it does not provide a validated threshold of educational importance. Such a threshold

was not specified for the instrument.

## Distribution of motivation categories

At pretest, 12 students were categorized as high, nine as moderate, and one as low. No student was categorized as very high or very low. At posttest, two students were very high, 15 high, four moderate, and one low. Consequently, the proportion in the combined high and very high categories increased from 12/22, or 54.55%, to 17/22, or 77.27%. The net difference was five students, equivalent to 22.73 percentage points. Table 4 presents the category frequencies and percentages. These categories are the thesis's score bands and should not be treated as externally validated clinical or diagnostic classifications.

**Table 4. Distribution across the thesis motivation categories**

| Category and score range | Pretest n | Pretest % | Posttest n | Posttest % |
|---|---|---|---|---|
| Very high: 106–125 | 0 | 0.00 | 2 | 9.09 |
| High: 86–105 | 12 | 54.55 | 15 | 68.18 |
| Moderate: 66–85 | 9 | 40.91 | 4 | 18.18 |
| Low: 46–65 | 1 | 4.55 | 1 | 4.55 |
| Very low: 25–45 | 0 | 0.00 | 0 | 0.00 |
| Total | 22 | 100.00 | 22 | 100.00 |

*Note. Categories reproduce thesis-defined score bands. Percentages are recalculated and may total 100.01% because of rounding. High plus very high increased from 54.55% to 77.27%.*

Category changes were not entirely upward. Comparing the paired records showed that eight students moved to a higher category, two moved to a lower category, and 12 remained within the same category. The number classified as low remained one, but this count concealed a change in which student occupied that band. Student S04 moved from low to high, whereas S01 moved from moderate to low. Student S21 moved from high to moderate. These examples demonstrate why unchanged category frequencies do not establish individual stability. Conversely, remaining in one category does not imply an unchanged total score, because each category covers multiple possible totals.

## Individual changes and indicator patterns

Thirteen students, or 59.09%, had higher posttest scores. Seven, or 31.82%, had lower scores, and two, or 9.09%, had identical scores. Individual differences ranged from −21 to +36, with a median difference of four points. The standard deviation of differences was 11.43, substantially larger than the mean increase of 3.68. This variation indicates that the modest positive average summarized a mixture of upward, downward, and unchanged responses. Table 5 preserves all 22 anonymous score pairs, making this heterogeneity visible. No causal explanation for an individual increase or decrease can be established from the score totals alone.

**Table 5. Individual motivation totals and paired changes**

| Student code | Pretest | Posttest | Change |
|---|---|---|---|
| **S01** | 76 | 65 | -11 |
| **S02** | 96 | 90 | -6 |
| **S03** | 78 | 94 | +16 |
| **S04** | 62 | 98 | +36 |
| **S05** | 93 | 86 | -7 |
| **S06** | 94 | 92 | -2 |
| **S07** | 94 | 99 | +5 |
| **S08** | 69 | 73 | +4 |
| **S09** | 81 | 90 | +9 |
| **S10** | 90 | 90 | 0 |
| **S11** | 86 | 90 | +4 |
| **S12** | 96 | 96 | 0 |
| **S13** | 105 | 110 | +5 |

| Student code | Pretest | Posttest | Change |
|---|---|---|---|
| **S14** | 81 | 87 | +6 |
| **S15** | 76 | 87 | +11 |
| **S16** | 79 | 77 | -2 |
| **S17** | 77 | 95 | +18 |
| **S18** | 97 | 104 | +7 |
| **S19** | 94 | 89 | -5 |
| **S20** | 100 | 110 | +10 |
| **S21** | 87 | 66 | -21 |
| **S22** | 78 | 82 | +4 |

*Note. Change = posttest − pretest. Totals: 1,889 and 1,970; sum of changes = 81. Student codes are anonymous. Eight moved up a category, two moved down, and 12 remained in the same category.*

All five reported indicator means increased descriptively. Confidence rose from 15.59 to 16.82, and enjoyment from 17.45 to 18.64. Their reported mean increases were 1.23 and 1.18, respectively, calculated from the underlying indicator values before rounding. Interest increased by 0.59, participation by 0.45, and effort and engagement by 0.23. Table 6 reproduces these source summaries. Small differences between subtraction of rounded displayed means and the reported changes reflect rounding. The ranking describes the observed pattern, not a statistically established difference between indicators. No indicator-level confidence intervals or hypothesis tests were reported, and none are implied by this comparison.

**Table 6. Source-reported motivation indicator means**

| Indicator | Pretest M | Posttest M | Change |
|---|---|---|---|
| **Interest** | 18.27 | 18.86 | +0.59 |
| **Participation** | 16.50 | 16.95 | +0.45 |
| **Confidence** | 15.59 | 16.82 | +1.23 |
| **Enjoyment** | 17.45 | 18.64 | +1.18 |
| **Effort and engagement** | 18.05 | 18.27 | +0.23 |

*Note. Values reproduce the thesis's indicator summaries. Differences were calculated from unrounded values; displayed subtraction may differ by .01. No indicator-level inferential tests are claimed.*

## Inferential analysis and uncertainty

The Shapiro–Wilk test of paired differences yielded W = .949 and p = .298. This result did not indicate a statistically detectable departure from normality in the available sample, although it does not prove normality. The paired-samples t-test yielded t(21) = 1.511 and p = .146. Because the p-value exceeded .05, the null hypothesis of zero mean change was not rejected. The 95% confidence interval for mean change extended from −1.39 to 8.75 points. This interval includes zero and values on both sides of it, showing uncertainty about the direction and magnitude of the underlying mean change under the statistical model.

Cohen's dz was 0.32, calculated by dividing the mean paired difference by its standard deviation. This standardized estimate describes the observed within-participant contrast and is not an independent-groups treatment effect. Table 7 summarizes the inferential results. Taken together, the analysis supports a modest positive descriptive tendency with appreciable uncertainty and heterogeneous individual responses. It does not support a statistically significant overall increase in motivation following the four meetings. Nor does it establish equivalence, prove an absence of educational value, or demonstrate improved speaking proficiency. Those questions require different designs, measures, or explicitly specified equivalence criteria.

**Table 7. Paired analysis of overall motivation**

| Statistic | Value |
|---|---|
| Mean paired change | 3.68 points |
| SD of paired changes | 11.43 |
| SE of mean change | 2.44 |
| 95% CI for mean change | −1.39 to 8.75 |
| Shapiro–Wilk test of differences | W = .949; p = .298 |
| Paired-samples t-test | t(21) = 1.511; two-sided p = .146 |
| Paired Cohen's dz | 0.32 |
| Decision at alpha .05 | Do not reject H0; no statistically significant overall change |
| Individual directions | 13 increased; 7 decreased; 2 unchanged |

*Note. Recalculation of paired total scores reproduced the source results. CI = confidence interval; SD = standard deviation; SE = standard error. A nonsignificant result is not an equivalence finding or proof of no effect.*

## Discussion

The study found a higher average motivation score after four meetings of supervised ChatGPT voice interaction, but the overall paired difference was not statistically significant. This combination is the central finding. The gain of 3.68 points describes what occurred in the selected class, while p = .146 and the confidence interval indicate uncertainty about a consistent mean change. Presenting only the improved average would exaggerate the evidence, whereas treating the nonsignificant test as proof that the activities were worthless would also overstate the result. The findings support a cautious account of classroom experience and identify questions that a stronger evaluation should address.

Individual variation helps explain why the mean alone is insufficient. More students increased than decreased, yet seven learners reported lower posttest motivation and two showed no change. The difference-score standard deviation exceeded the mean increase, demonstrating substantial variability around the average. Different responses might reflect task difficulty, confidence, familiarity, or access to speaking turns, but these explanations were not measured. The results therefore justify examining heterogeneity without assigning unverified reasons to particular students. Future work should document individual exposure and ask learners about their experiences, allowing researchers to distinguish implementation barriers from differences in preferences or starting conditions.

Confidence and enjoyment showed the largest descriptive indicator changes. A responsive conversational activity might help some learners attempt English without feeling the same pressure as a public classroom performance. Tai and Chen (2023) and Yang et al. (2022) provide relevant voice-interaction comparisons, although their systems and outcomes differ. The present study did not directly measure anxiety or willingness to communicate, so it cannot claim that either changed. The interpretation should remain close to the questionnaire: learners' reported confidence and enjoyment increased descriptively. Establishing why they increased would require dedicated measures, conversational records, or interviews linked to the intervention.

The smaller change in effort and engagement is equally informative. Enjoyment of an activity and sustained commitment to language learning are related possibilities, but they represent different educational questions. Dörnyei (2001), Riyanti (2019), and Fryer et al. (2017) provide reasons to attend to persistence and the distinction between initial interest and continuing engagement. Four meetings may capture an early response without documenting longer-term habits. However, treatment duration was not experimentally varied, so the study cannot conclude that additional meetings would necessarily improve motivation. A longitudinal design should examine whether interest persists, diminishes, or becomes connected with measurable changes in independent learning behavior.

Self-determination theory offers an interpretive vocabulary for future research. Choice of topic could support autonomy, manageable conversation could support perceived competence, and responsive exchanges could affect the experience of interaction. Deci and Ryan (2000) and Annamalai et al. (2023) inform these possibilities. They were not tested as mediating processes here. Supervised access through one laptop also differs from autonomous personal use, and a system's responsiveness cannot be assumed to replace human relationships. Future studies should measure these processes directly and compare activity arrangements, rather than labeling every positive response as evidence that psychological needs were

satisfied through AI interaction.

Regional research provides useful context without overriding the present result. Wahyuddin et al. (2025), Handayani et al. (2026), and Fatmawati et al. (2026) examine AI-supported language learning in other populations and settings. Their work identifies relevant issues concerning learner experience, instructional use, and speaking practice. The current investigation contributes a different outcome: an uncertain overall motivational change after a short supervised intervention. A positive report about speaking performance or perceptions does not require a statistically significant change in a broad motivation scale, particularly when the system, age group, and instructional arrangements differ.

Comparisons with mobile learning and gamification also require attention to design. Junandar et al. (2026) investigated an extended mobile-learning intervention with a comparison group, whereas the present study used four meetings in one class. Azis et al. (2026) and Abdul Jalil Jum'uatullaila et al. (2025) address different digital activities. Their findings cannot establish that ChatGPT should produce equivalent changes. Wahyuddin et al. (2024) and Azizah et al. (2026) further contextualize digital literacy and school technology use. Together, these studies support describing access and pedagogy precisely. They do not justify ranking applications by unrelated mean gains or assuming one platform is universally motivating.

The boarding-school setting has practical implications. The supervised laptop arrangement respected the reported restriction on personal smartphones and provided access within existing school routines. Yet it does not reveal how much speaking time each student received or whether the experience felt private. Teachers considering similar activities should plan clear conversational goals, manageable linguistic demands, and opportunities for every learner to participate. Complementary peer tasks are also plausible, as contextualized by Amalia et al. (2024), Saleh et al. (2026), and Hermansyah et al. (2026). These are instructional options informed by related literature, not combinations that this study tested or demonstrated to be effective.

Measurement places further limits on interpretation. Reported alpha coefficients support internal consistency, but Cronbach's (1951) coefficient does not establish all aspects of validity. The instrument combines broad English-learning orientations with enjoyment of a specific voice-interaction experience. If students had limited prior experience with that technology, some baseline responses may represent expectations, while later responses represent direct experience. This possibility requires clarification of pretest instructions and prior exposure. Repeated self-report can also be affected by expectations surrounding a researcher-led activity. Independent observations and delayed measures would help assess whether reported enthusiasm corresponds with participation, persistence, or language use beyond the intervention.

The inferential conclusion should remain proportionate to the design. As Lakens (2013) explains, effect-size reporting requires a clear calculation basis; the paired dz here describes within-person change and cannot be interpreted as a controlled treatment advantage. The confidence interval includes zero, and no equivalence margin was tested. The next study should use multiple classes, a meaningful comparison condition, documented exposure, and sufficient follow-up to examine sustained motivation. Researchers should also specify meaningful changes before analysis and preserve transparent reporting of unfavorable or mixed responses. This approach would make future findings more useful for determining which students benefit, under what conditions, and with what instructional support.

## Conclusion

Four meetings of supervised voice interaction with ChatGPT were followed by a modest descriptive increase in English-learning motivation among 22 eighth-grade students at MTs Assalam Sidrap. The mean increased from 85.86 to 89.55, with confidence and enjoyment showing the largest descriptive indicator gains. However, the paired difference was not statistically significant, $t(21) = 1.511$, $p = .146$, 95% CI [−1.39, 8.75], and individual responses varied. The findings therefore do not establish a significant overall motivational influence under the conditions examined.

Voice interaction may be considered a supplementary classroom activity when it fits learners' needs, instructional goals, and school access arrangements. This study does not demonstrate that it outperforms other approaches or produces sustained motivation. Larger controlled investigations should document individual conversational exposure, clarify baseline questionnaire interpretation, incorporate independent behavioral evidence, and include delayed follow-up. These improvements would help distinguish immediate enjoyment and confidence from continuing effort to learn English and provide stronger evidence for instructional decisions.

## References


Abdul Jalil Jum'uatullaila, Buhari, Syahrir L, & Hermansyah, S. (2025). The influences of using Quizizz in student's English learning outcomes. INTERACTION: Jurnal Pendidikan Bahasa, 12(1), 1030–1040. https://doi.org/10.36232/interactionjournal.v12i1.3526

Amalia, A. P., Hermansyah, S., Sadapotto, A., & Syahrir L. (2024). Applying project-based learning to enhance students' English public speaking skills at MTs YMPI Rappang. INTERACTION: Jurnal Pendidikan Bahasa, 11(2). https://doi.org/10.36232/interactionjournal.v11i2.1199

Annamalai, N., Eltahir, M. E., Zyoud, S. H., Soundrarajan, D., Zakarneh, B., & Al Salhi, N. R. (2023). Exploring English language learning via Chabot: A case study from a self determination theory perspective. Computers and Education: Artificial Intelligence, 5, Article 100148. https://doi.org/10.1016/j.caeai.2023.100148

Azis, N., Hermansyah, S., Hikmah, N., & Faradillah, N. (2026). Enhancing students' motivation in learning English through Wordwall at SMPN 2 Pitu Riase. INTERACTION: Jurnal Pendidikan Bahasa, 13(1). https://garuda.kemdiktisaintek.go.id/documents/detail/6273544

Azizah, N. A., Hermansyah, S., Syahrir L, & Isumarni. (2026). Teachers and students' perceptions of smartphone use in EFL classes at SMAN 2 Sidrap. Macca: Journal of Linguistic Applied Research. https://journal.adityarifqisam.org/index.php/macca/article/view/317

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. IEEE Access, 8, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297–334. https://doi.org/10.1007/BF02310555

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

Dörnyei, Z. (2001). Motivational strategies in the language classroom. Cambridge University Press. https://doi.org/10.1017/CBO9780511667343

Fatmawati, Syamsu T, Faradillah, N., Sadapotto, A., & Hermansyah, S. (2026). The effectiveness of AI chatbot-based English learning methods in improving students' speaking skills. INTERACTION: Jurnal Pendidikan Bahasa, 13(1). https://doi.org/10.36232/interactionjournal.v13i1.5675

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. Language Learning & Technology, 10(3), 8–14. https://doi.org/10.64152/10125/44068

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. Computers in Human Behavior, 75, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Hamid, T. P. (2026). The influence of voice interaction with the virtual assistant ChatGPT on students' motivation in learning English at MTs Assalam Sidrap [Unpublished undergraduate thesis]. Universitas Muhammadiyah Sidenreng Rappang.

Handayani, F., Hermansyah, S., Rahim, T. R., & Sain, Y. (2026). The integration of artificial intelligence in English language teaching: Opportunities and challenges. Macca: Journal of Linguistic Applied Research, 3(1). https://journal.adityarifqisam.org/index.php/macca/article/view/258

Hermansyah, S. (2026). Pemanfaatan teknologi digital dalam pembelajaran speaking bahasa Inggris di sekolah menengah: Tinjauan literatur tentang manfaat dan tantangannya. Jurnal Teknologi Pendidikan, 4(2). https://jurnal.umsrappang.ac.id/jtp/article/view/3170

Hermansyah, S., Usman M., & Syamsu T. (2026). Developing digital learning media for teaching public speaking to improve English skills of junior high school students. International Journal of Special Education, 41(16s), 502–509. https://www.internationalsped.com/index.php/ijse/article/view/5027

Junandar, A., Hermansyah, S., Maulina, Amin, S., & Syarif, A. R. (2026). The impact of mobile learning applications on students' motivation in learning English in MTS Neg. 3 Konawe Selatan. INTERACTION: Jurnal Pendidikan Bahasa, 13(1), 757–767. https://doi.org/10.36232/interactionjournal.v13i1.5646

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537–550. https://doi.org/10.1177/00336882231162868

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Riyanti, D. (2019). The role of motivation in learning English as a foreign language. Journal of English

Language Teaching Innovations and Materials, 1(1), 29–35. https://doi.org/10.26418/jeltim.v1i1.27788
Saleh, S. M., Maulina, Amin, S., Pratiwi, W. R., & Hermansyah, S. (2026). Penguatan peran Mobile Assisted Language Learning (MALL) dalam pembelajaran kosakata bahasa Inggris pada pembelajar ESL. Jurnal Pengabdian Masyarakat Bangsa, 4(3), 806–815. https://doi.org/10.59837/jpmba.v4i3.4360
Tai, T.-Y., & Chen, H. H.-J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. Interactive Learning Environments, 31(3), 1485–1502. https://doi.org/10.1080/10494820.2020.1841801
Wahyuddin, N. R., Yanti, N. E., Arnas, R., & Hermansyah, S. (2024). Digital literacy integrated with blended learning in improving EFL students' English language skills: A lesson learned from the Independent Campus Program. Journal of Language and Literature Studies, 4(4), 744–757. https://doi.org/10.36312/jolls.v4i4.2351
Wahyuddin, N. R., Yanti, N. E., Arnas, R., & Hermansyah, S. (2025). Utilization of artificial intelligence in EFL learning from a digital literacy perspective. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 13(2). https://doi.org/10.24256/ideas.v13i2.8331
Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. ReCALL, 34(3), 327–343. https://doi.org/10.1017/S0958344022000039

## Editorial notes for author verification

This manuscript adapts the supplied thesis by Tony Putra Hamid. Only the thesis author is named because the article's complete author list and contact details have not been supplied. Confirm final authorship, affiliations, corresponding author, funding, conflicts and contributions before submission.

The 22 paired totals were independently recalculated and match the reported mean difference, standard deviations, Shapiro–Wilk test, paired t-test, confidence interval and paired effect size. The nonsignificant overall result is preserved. Indicator means and internal-consistency estimates are source-reported; item-level reliability was not independently recalculated from original questionnaire files.

The article adds transparent descriptive summaries: 13 totals increased, seven declined and two were unchanged; eight students moved up a motivation category, two moved down and 12 remained in the same category. Overall score change and category change are different quantities.

The supplied Google Scholar profile could not be accessed directly. The 25-entry bibliography includes 12 relevant publications by Sam Hermansyah and collaborators identified through publisher or indexed records, 12 methodological or theoretical sources, and the thesis. Inclusion of every entry in that profile is not established. It must not be described as 25 verified publications drawn exclusively from the profile.

Bibliographic details not established from accessible records are omitted rather than invented. The Wordwall record is linked through its Garuda index because its publisher link returned an error. Some journal metadata changed during retrieval; export final citations from the version of record before submission.

Clarify exact study dates, duration per meeting, individual time on the device, conversation settings, system version, prior AI exposure and any use outside the sessions. The source supports four meetings and researcher-supervised laptop access; it does not establish completely private individual practice.

The source reports expert judgment and school permission. Provide the actual ethics, parental permission and student assent documentation as applicable, along with an accurate description of voice-data handling and account arrangements. These details have not been invented.

Clarify how the enjoyment-through-voice-interaction items were introduced at baseline if students had no prior experience. The instrument may capture expectations at one time and experience at another. Its alpha coefficients alone do not resolve this measurement question.

Approximate word counts excluding section headings, tables, references and these editorial notes: Abstract: 236; Introduction: 1000; Literature Review: 800; Method: 600; Results: 1000; Discussion: 1000; Conclusion: 160.