J-Shelves of Indragiri (JSI)
https://ejournal-fkip.unisi.ac.id/shelves
Email: jsipbifkipunisi@gmail.com
DOI: https://doi.org/10.61672/jsi.v8i2.3831

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

ENGLISH SPEAKING DEVELOPMENT THROUGH PROFICIENCY BASED PAIRING AND ROLE PLAY AMONG PHYSICAL EDUCATION STUDENTS

Mariska Febrianti[1*], Azizatul Banat[2]
[1,2]Universitas Dehasen Bengkulu, Indonesia
Author (s) Email: [1]mariska.febby@unived.ac.id, [2]azizatulbanat@unived.ac.id

*Corresponding Author Email: mariska.febby@unived.ac.id

Article Info

Article History:
Received: 30-09-2026
Revised: 01-11-2026
Accepted: 04-11-2026
Published:12-11-2026

Keywords:
English speaking;
Physical education;
Proficiency-based pairing;
Role play;
Classroom action research

Abstract

English speaking competence is increasingly important for Physical Education students because their academic and future professional activities may require them to explain instructions, interact with others, and communicate sport-related information in English. However, existing studies in Physical Education contexts have largely emphasized vocabulary learning or students' perceptions, leaving limited evidence on how classroom strategies can improve directly assessed speaking performance among non-English-major students. This study therefore examines changes in English speaking rubric scores across two instructional cycles in a Physical Education cohort whose classroom design combined proficiency-based pairing and role play. Forty matched student records were analyzed for fluency, vocabulary, pronunciation, and confidence. Descriptive statistics, individual change scores, and histograms were used to distinguish overall improvement from dimension-specific patterns. Mean total scores increased from 6.45 (SD = 3.12) in Cycle I to 9.03 (SD = 3.65) in Cycle II, with a mean paired increase of 2.58 raw points. Thirty-seven students (92.5%) improved, three remained unchanged, and none declined. Confidence showed the largest mean increase (+1.33), followed by vocabulary (+0.70) and fluency (+0.55), while pronunciation remained unchanged for all students. The findings extend existing role-play research by providing matched performance ratings in a non-English-major Physical Education cohort and demonstrate that overall improvement may conceal uneven development across speaking dimensions. However, no verified baseline assessment, completed participation observations, or pair identifiers were available; consequently, improvement in classroom activeness and the specific contribution of proficiency-based pairing could not be established. The results support further classroom evaluation of the combined approach, with separate monitoring of participation, clearer documentation of pair interaction, and targeted pronunciation support.

INTRODUCTION

Physical Education students need opportunities to use English for purposes that resemble their emerging professional responsibilities. Explaining a drill,

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

clarifying a game rule, interviewing a player, and giving constructive feedback require more than recognition of sport terminology. Speakers must organize a message, respond to an interlocutor, and repair misunderstanding within limited time. These demands make PE a useful context for connecting English instruction with purposeful oral interaction. The present adaptation focuses on fourth-semester undergraduates rather than the Grade 7 learners described in the source proposal (Setiawan, 2025).

Relevant local research provides a starting point. Febrianti et al. (2019) developed instructional materials using Goal.com for sports students at Dehasen University, linking English learning with sport-related content. Putra et al. (2024) examined Wordwall.net in vocabulary teaching for PE students through descriptive qualitative research and reported perceived benefits for vocabulary learning and classroom enthusiasm. These studies support the relevance of contextualized materials, but vocabulary engagement does not establish the ability to sustain an unscripted conversation. An instructional sequence that connects available vocabulary to explanation, questioning, and response therefore deserves investigation.

The supplied school proposal identifies limited vocabulary, hesitation, and low participation as classroom problems. These observations cannot be transferred to a new undergraduate cohort as established baseline findings. For PE students, diagnostic speaking tasks and classroom observation are needed to determine whether similar difficulties exist, how widely they occur, and whether students participate differently across familiar and unfamiliar sport topics. Academic semester, sporting experience, prior English exposure, and confidence may vary within the same class. The intervention must respond to this variation rather than assume uniform needs.

Proficient pairing is used here as an operational term for arranging students with different current levels of English-speaking performance to work together. A more proficient partner can model a phrase or request clarification, while the other partner contributes sport knowledge and practices independent responses. Pair composition alone cannot guarantee productive learning. Storch (2002) shows that

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

dyadic interaction varies in equality and mutuality, making the organization of participation central to the rationale for pairing. Structured turn allocation and role reversal are consequently built into the proposed intervention.

Role play adds a communicative purpose to paired work. Recent research with university EFL learners illustrates its application to professional speaking activities (Lisa et al., 2025). A systematic review by Kaygısız and Akar (2025) identifies its potential across cognitive, affective, and self-regulative outcomes while emphasizing implementation conditions. For PE, the relevant adaptation is to create scenarios in which partners must exchange information to complete a sporting or instructional task. Memorizing a dialogue would provide weaker evidence of communication than handling a new question during the same scenario.

This study framework addresses a specific instructional problem: how to move from supported vocabulary use toward active and individually assessable oral communication in a PE cohort. It does not claim that pairing or role play is a new technique. Its intended contribution lies in combining proficiency-informed pairs, PE-contextualized tasks involving introductions, campus interactions, and descriptions, progressive withdrawal of support, and separate measures of participation and speaking performance. The research questions are: How does observed activeness change across two action research cycles? How does individual English communication performance change?

LITERATURE REVIEW

Speaking performance and classroom activeness

English communication is defined as the ability to convey an understandable message and respond appropriately during an oral exchange. Following the dimensions specified in the source proposal, the assessment framework includes fluency, vocabulary, pronunciation, confidence, and clarity. Fluency concerns continuity of expression; vocabulary concerns appropriate lexical choice; pronunciation concerns intelligibility; confidence concerns observable independence in delivery; and clarity concerns whether the intended message can be followed. Grammar is considered when it affects meaning but is not introduced

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

as an additional weighted dimension, because doing so would change the original construct without justification.

Activeness concerns observable participation in learning opportunities. Fredricks et al. (2004) distinguish behavioral, emotional, and cognitive engagement. The present observation sheet addresses the behavioral component through initiation, relevant response, sustained engagement, and contribution to collaborative work. It cannot by itself measure a student's internal motivation or depth of thinking. A student who speaks frequently may still communicate unclearly, while a quieter student may perform accurately when given a fair opportunity. Reporting both constructs separately makes those differences visible.

Peer assistance and the conditions for scaffolding

The sociocultural rationale retained from the source proposal treats assisted interaction as a potential resource for development. However, assistance becomes educationally useful when it matches the learner's difficulty and enables greater independence. Van de Pol et al. (2010) identify contingency, fading, and transfer of responsibility as defining features of scaffolding. Their review concerns teacher–student interaction; applying these principles to peer pairs is a design inference that must be examined in classroom records. In this protocol, a partner first offers a limited prompt, then allows time for a response, and gradually withdraws assistance.

Storch's (2002) work provides a complementary caution: a proficiency difference does not determine whether a pair collaborates. A fluent student who writes and performs the whole dialogue may leave the partner with very little learning opportunity. Accordingly, proficient pairing should not become a permanent tutor–dependent hierarchy. Both students must ask questions, explain an idea, and respond without a complete script. Pair changes should be documented because they can alter the interactional conditions under which progress occurs.

Role play and willingness to communicate

Role play creates a setting in which language performs a recognizable function. Kaygısız and Akar (2025), reviewing 32 studies published between 2000 and 2023, discuss the importance of learner engagement and teacher roles in

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

implementation. This supports a carefully staged design rather than an assumption that acting automatically improves speaking. Role cards should specify goals and information differences, leaving students to formulate their own messages. A debriefing phase then connects performance difficulties with the next rehearsal.

The willingness-to-communicate framework of MacIntyre et al. (1998) helps explain why linguistic knowledge and participation may diverge. Readiness to speak depends partly on the immediate communicative situation. A familiar sport topic and supportive partner may make an exchange more approachable, but public evaluation or a dominant partner may inhibit it. This perspective complements the affective emphasis in the source proposal. Changes in observed confidence are treated as behavioral evidence; reduced anxiety would require its own measure or carefully interpreted interview evidence.

Physical Education context and earlier research

PE offers distinctive communicative tasks because explanation often accompanies demonstrations, spatial arrangements, rules, and feedback. A diagram of a playing area can supply a shared referent while students practice sequencing language or clarification requests. Such resources are justified by the content of the task, not by classifying all PE students as kinesthetic learners. Familiarity with football, badminton, athletics, or another sport must be checked rather than presumed. Accessible alternatives allow students with different sporting backgrounds to engage with an equivalent language demand.

Febrianti et al. (2019) provide an antecedent for using sport-related materials, whereas Putra et al. (2024) provide a local connection between vocabulary work and participation in PE English learning. The proposed study extends the outcome focus to live interaction. The two publications are treated as empirical antecedents rather than as independent named theories. Their educational settings are relevant, but their designs and vocabulary-focused outcomes cannot supply evidence for the present intervention’s effectiveness.

Evidence from adjacent settings further clarifies the limits of transfer. Lisa et al. (2025) explored professional role play with 58 English language education students through a qualitative case study; that population differs from non-English-

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

major PE undergraduates. Budiman et al. (2023) investigated role play and social skills among 30 sixth-grade pupils in PE, using a one-group pretest–posttest design. Their work supports examining the social dimension of role-based activity, but social skills among schoolchildren are not equivalent to English speaking achievement among university students. These distinctions justify collecting direct evidence in the intended cohort.

Krebt (2017) examined role-play instruction among 40 Iraqi EFL college students divided into experimental and control groups. The study reported a speaking advantage for the role-play group following 20 lessons. It provides a useful comparison for the instructional approach, although its language-student population differs from PE undergraduates. It also does not establish the separate contribution of proficiency-based pairing.

Studies of non-English majors offer more direct contextual comparisons. Hoang et al. (2023) surveyed 100 students and found favorable perceptions of the contribution of role play to communication, confidence, and participation. Zulfa et al. (2024) similarly reported positive perceptions among 31 Indonesian Information Systems and Accounting Information Systems students. These accounts establish the perceived usefulness of role play, but self-reported benefits do not independently demonstrate changes in assessed performance.

Pham and Do (2025) move beyond perceptions in a two-cycle action research study involving 41 non-English-major students. Using observation, speaking tests, and questionnaires, they reported improvements in speaking performance and active participation. Their study is an especially close precedent: assessing speaking and participation together cannot itself be claimed as a new contribution. Kaygısız and Akar's (2025) review likewise emphasizes that role-play outcomes depend on instructional design and implementation, including preparation and reflective debriefing.

Pairing research complicates the assumption that a higher-proficiency partner necessarily improves learning. Zabihi and Ghahramanzadeh (2022) found that high–high pairs generated and resolved more cognitive conflict episodes during joint composition than the other pairing configurations examined; mixed

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

proficiency could weaken engagement. Li and Liu (2022), studying 62 non-English-major students, found stronger differences by collaborative interaction pattern than by proficiency grouping in the occurrence and resolution of language-related episodes. Both studies concerned writing-related tasks, so their findings inform the rationale for oral pair work without directly demonstrating its effects.

The gap emerging from these studies concerns the connection between pairing arrangements, individual participation, and dimension-specific oral development in non-English-major classrooms. The present dataset addresses only part of that gap: it permits matched comparisons across two cycles and separates linguistic ratings from confidence. It does not permit tests of interaction quality or participation equity. Its contribution is therefore a transparent account of uneven speaking development within a PE-oriented intervention, together with a clear identification of the process evidence still needed to explain that development.

RESEARCH METHOD

The adapted protocol uses classroom action research with two cycles of planning, action, observation, and reflection, drawing on Kemmis et al. (2014). Descriptive quantitative assessment is embedded within classroom inquiry supported by observation, interviews, and recordings. This extends the predominantly qualitative analysis described in the source so that numerical speaking improvement can be evaluated transparently. The purpose is to improve practice in one class; the design does not isolate the causal contribution of pairing from that of role play.

The intended cohort comprises 40 fourth-semester PE students, as specified in the study. The setting is a PE English class at Universitas Dehasen Bengkulu,. The study must identify the actual course, procedure, attendance, and analyzed sample. An intact class would constitute classroom-based inclusion rather than random population sampling. Ages, gender distribution, previous English attainment, and study dates remain unreported because supporting records were not supplied. April-June is school timeline in the source of data taken by the researcher in MKU Bahasa Inggris Subject.

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

A baseline oral task provides the basis for arranging 20 pairs. As an initial rule, the 20 higher-scoring and 20 lower-scoring students can be matched by corresponding ranks within each half, subject to compatibility and manageable proficiency differences. The lecturer should avoid publicly labeling students as strong or weak. Pairing relies on English performance, not athletic competence. Students keep both speaker and listener responsibilities, and the lecturer records any re-pairing and its reason. If diagnosis shows little proficiency variation, the pairing rationale must be revised and reported honestly.

The three topics specified by the researcher are Introducing Yourself to New People, Campus Life, and Describing People. The adapted protocol organizes these topics into two cycles, with one topic-focused meeting per cycle and a separate baseline assessment. Meeting duration, actual dates, and instructional exposure require confirmation from teaching records. Cycle I uses model expressions and role cards; Cycle II revisits the same language functions with new partners or information and reduced prompts. Assessment tasks use comparable demands and preparation time, including an unprepared follow-up question for each learner.

Table 1. Classroom Action Research Sequence

No.	Stage	PE task and action	Evidence
1	Baseline	Introduce oneself, respond to a campus question, and give a short description using parallel prompts.	Individual rubric scores; observation opportunities; topic familiarity.
2	Cycle I	Practice all three topics with expression; switch roles and ask follow-up questions.	Recordings, observation scores, individual assessment, field notes.
3	Reflection I	Review initiation, script dependence, unequal speaking turns, and unsuccessful clarification.	Documented evidence supporting each revision.
4	Cycle II	Repeat the three functions with new information and an unprepared question.	Equivalent final tasks; repeated observation; interviews.

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

No.	Stage	PE task and action	Evidence
5	Final reflection	Compare individual trajectories across topics and examine limited or negative change.	Quantitative and qualitative evidence, including contradictory cases.

Analytic Speaking Performance Rubric according to Cambridge English in *Assessing speaking performance: Level B1* stated that:

a. Fluency

Score 1. Produces isolated words and cannot sustain a short message.

Score 2. Produces short phrases with frequent interruptions.

Score 3. Completes short connected responses despite noticeable hesitation.

Score 4. Sustains the exchange with occasional disruptive pauses.

Score 5. Develops responses continuously, with pauses that rarely interrupt the message.

b. Vocabulary

Score 1. Lacks essential words for the assigned task.

Score 2. Uses a narrow repertoire that conveys only part of the message.

Score 3. Uses sufficient everyday words to complete the main task.

Score 4. Selects appropriate wording and sometimes explains unfamiliar terms differently.

Score 5. Uses varied, precise wording and reformulates effectively when needed.

c. Pronunciation

Score 1. Most attempted messages cannot be understood.

Score 2. The listener frequently needs repetition.

Score 3. The main message is understandable despite recurring difficulties.

Score 4. Speech is clear, with occasional problems affecting understanding.

Score 5. Speech remains readily understandable throughout the task.

d. Confidence related speaking behavior

Score 1. Requires continuous individual prompting to attempt a speaking turn.

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

Score 2. Delivers responses with frequent prompting or extensive script dependence.

Score 3. Delivers prepared responses independently but needs support for follow-up questions.

Score 4. Handles most follow-up questions independently and resumes after difficulty.

Score 5. Sustains independent delivery and handles unexpected questions without a partner supplying the response.

The teaching framework described in the source manuscript combines proficiency-informed pairing with three role-play topics: Introducing Yourself to New People, Campus Life, and Describing People. Partners are expected to alternate speaking and listening responsibilities. The planned sequence uses model expressions and role cards before reducing prompts in later practice. These features describe the instructional framework; the score workbook alone cannot verify the pairing algorithm, the duration of instruction, or the extent to which role rotation and reduced support were implemented.

The topics provide a plausible link between general English and PE student life. An introduction task can involve meeting a new classmate or sports-club member; a campus-life task can involve asking for directions to a teaching room or sports facility; a description task can involve identifying a fictional participant. These examples clarify the intended task functions rather than report observed classroom episodes. Claims about specific assistance, negotiation, or independent responses require supporting recordings or field notes.

The workbook records four dimensions: fluency, vocabulary, pronunciation, and confidence. Each total is the sum of these four entries. The five-dimension rubric in the earlier manuscript, which also included clarity, was not applied in the present analysis because no clarity scores were supplied. Ratings are reported in their original units. On the 1–5 interpretation described in the source manuscript, the four-dimension total would range from 4 to 20; however, the workbook contains

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

no rating anchors, so raw totals remain the primary outcome.

All 80 recorded totals were checked against their component sums, and all matched. Student codes were matched across cycles, with no missing component ratings among the 40 paired records. Confidence was retained in the original composite but also examined separately because it includes an affective or behavioral aspect of performance. As a descriptive sensitivity check, fluency, vocabulary, and pronunciation were summed without confidence. This secondary subtotal is not presented as a validated alternative scale.

Several regularities require caution: fluency and vocabulary scores are identical for all students in Cycle I; pronunciation and confidence are also identical in that cycle. In Cycle II, vocabulary and confidence are identical for every student, and pronunciation is unchanged for every student across cycles. These entries were preserved exactly as supplied. Their recurrence warrants comparison with original rating sheets before submission and limits interpretation of the dimensions as independent indicators.

The researcher prepared three tasks for this study, namely Task 1 Introducing Yourself to New People, Task 2 Campus Life and Task 3 Describing People. Then, teacher will meet a new student at a university sports club. Introduce yourself, ask about your partner's interests, and ask at least one follow-up question. Then exchange roles.

Role A: You are an existing club member. Welcome the new student and ask about their interests.

Role B: You are a new member. Introduce yourself and ask about the club. In the second round, exchange roles and use a different club context.

Your partner is visiting the campus for the first time. Explain how to reach a sports facility and answer questions about its location or activities. Then exchange roles using a different destination.

FINDINGS AND DISCUSSION

Overall change and score distributions

The mean total score increased from 6.45 (SD = 3.12) in Cycle I to 9.03 (SD

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

= 3.65) in Cycle II. The corresponding medians were 6 and 7, and observed ranges were 4–14 and 4–18. The mean within-student increase was 2.58 points (unrounded value = 2.575; SD of change = 1.55), with a median increase of 3 and a range of 0–6. These values are calculated from the four recorded dimensions and supersede the illustrative baseline and cycle means in the earlier draft.

Figure 1 shows that improvement was accompanied by substantial concentration at lower totals. In Cycle I, 16 students scored 4 and 15 scored 6, together accounting for 77.5% of the class. In Cycle II, the most frequent score was 7, obtained by 19 students (47.5%); seven scored 10, and three scored 18. The higher mean therefore should not be interpreted as uniform high performance. The increase in standard deviation also shows that between-student variation did not disappear.

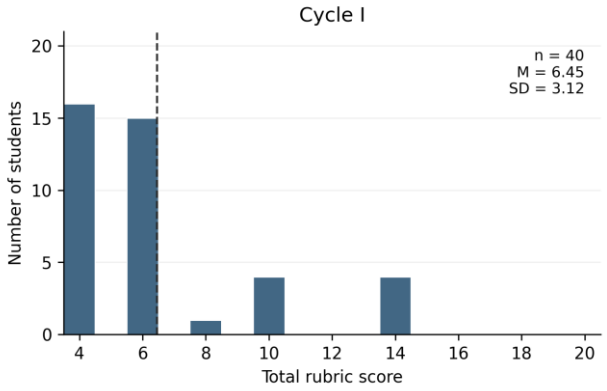


Figure 1. Distribution of total speaking rubric scores in Cycle I

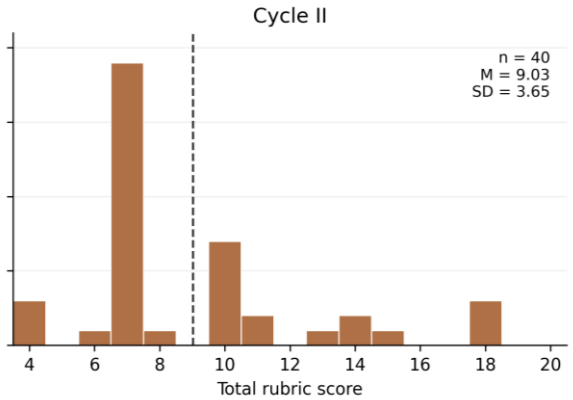


Figure 2. Distribution of total speaking rubric scores in Cycle II

Thirty-seven students (92.5%) increased their total score, three (7.5%)

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

remained unchanged, and none declined. The unchanged records were W, HH, and II, each with a total of 4 in both cycles. Gains of 1 and 4 points were equally frequent, occurring in 12 students each; ten gained 3 points, while gains of 2, 5, and 6 points each occurred once. Figure 2 shows why the class mean alone is insufficient to describe individual development.

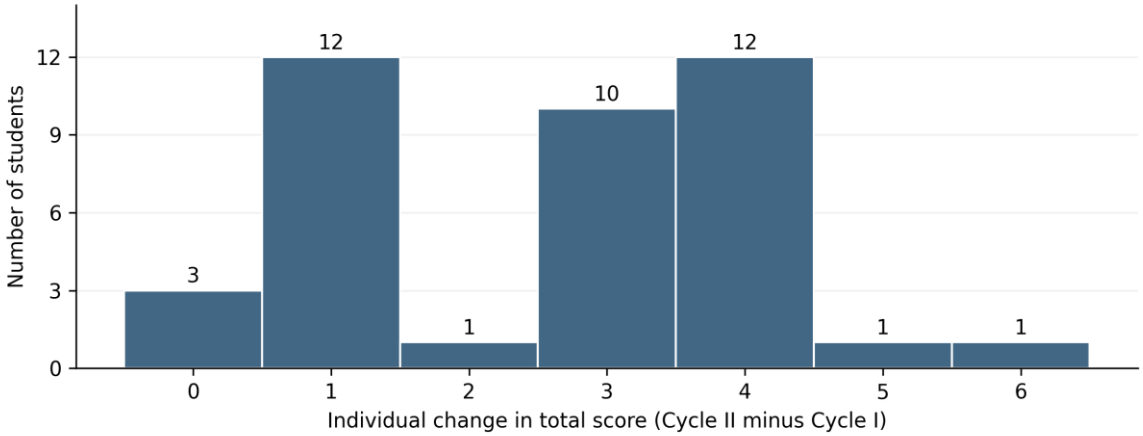


Figure 3. Distribution of individual changes in total speaking rubric scores.

Fluency increased from M = 1.925 (SD = 0.971) to M = 2.475 (SD = 1.109), a mean change of 0.550. Twenty-two students improved and 18 remained unchanged. Vocabulary increased from M = 1.925 (SD = 0.971) to M = 2.625 (SD = 1.055), a mean change of 0.700; 25 students improved and 15 remained unchanged. Neither dimension declined for any student.

Pronunciation remained at M = 1.300 (SD = 0.648) in both cycles. Every student retained the same pronunciation rating: 32 received 1, four received 2, and four received 3. Confidence increased from M = 1.300 (SD = 0.648) to M = 2.625 (SD = 1.055), a mean gain of 1.325. Thirty-seven students improved in confidence and three remained unchanged. Thus, the largest numerical contribution to composite change came from confidence, while pronunciation contributed none.

Confidence accounted for 51.5% of the summed composite increase, vocabulary for 27.2%, and fluency for 21.4%, with minor rounding differences. This arithmetic decomposition describes where the recorded change occurred; it is not a causal explanation. Excluding confidence, the mean three-dimension subtotal

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

increased from 5.15 to 6.40, a gain of 1.25 points. Consequently, the direction of improvement was not solely dependent on confidence, although confidence strongly influenced its magnitude.

Evidence concerning classroom activeness

Confidence ratings were examined as an indirect indicator of readiness to participate in spoken classroom tasks. This interpretation concerns perceived or rated assurance in speaking, rather than observed behavioral participation. The dataset does not record how often students asked questions, initiated exchanges, or took speaking turns. Accordingly, the following results describe confidence-related progress and do not establish a measured increase in classroom activeness.

In Cycle I, mean confidence was 1.300 (SD = 0.648), and the median was 1. Thirty-two students (80.0%) received a rating of 1, four (10.0%) received 2, and four (10.0%) received 3. No student received 4 or 5. The distribution was therefore concentrated at the lowest recorded rating. This pattern indicates limited rated confidence at the first available cycle assessment, although it does not establish how much each student actually participated.

In Cycle II, mean confidence increased to 2.625 (SD = 1.055), and the median rose to 2. Three students (7.5%) received a rating of 1, 20 (50.0%) received 2, nine (22.5%) received 3, five (12.5%) received 4, and three (7.5%) received 5. The number at the lowest rating fell from 32 to three. Ratings of 3 or above became more frequent, increasing from four students (10.0%) to 17 (42.5%); this is a descriptive distributional comparison, not a validated threshold for active participation.

J-Shelves of Indragiri (JSI)
https://ejournal-fkip.unisi.ac.id/shelves
Email: jsipbifkipunisi@gmail.com
DOI: https://doi.org/10.61672/jsi.v8i2.3831
November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

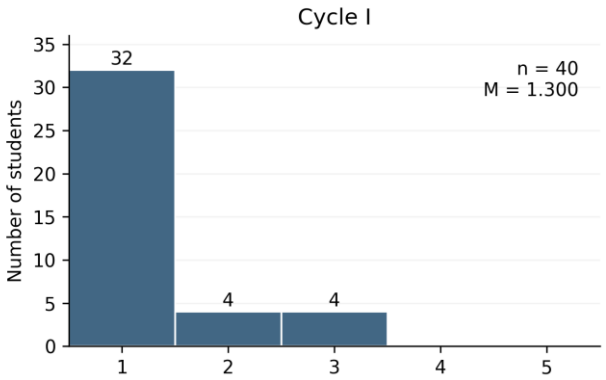


Figure 3. Histograms of activeness in Cycle I

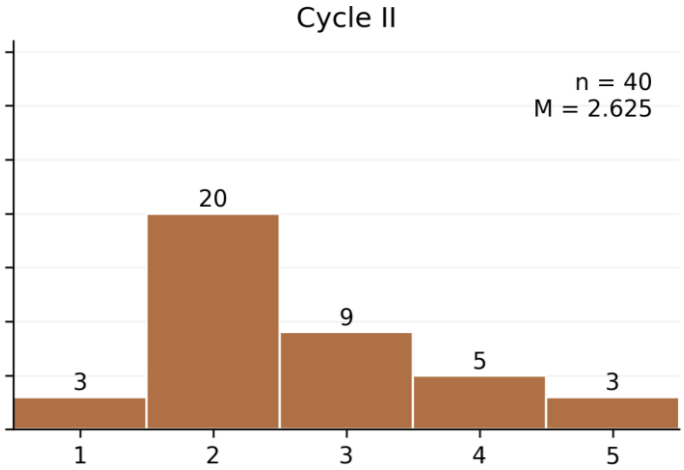


Figure 4. Histograms of activeness in Cycle II

Identical one-point and frequency are used for 40 matched students. Confidence is an indirect indicator of readiness to participate, not a direct measure of classroom activeness. The paired mean increase was 1.325 rating points. Thirty-seven students (92.5%) improved in confidence, three (7.5%) remained unchanged, and none declined. Figure 3 shows a shift away from the lowest rating toward a wider distribution across ratings 2–5. Nevertheless, half the class remained at rating 2, indicating that higher confidence was not uniformly attained. The percentage improving refers to individual confidence gains and must not be interpreted as the percentage of active students.

For presentation on a common 100-point scale, dividing the mean confidence rating by the assumed maximum of 5 and multiplying by 100 yields 26.0 in Cycle I and 52.5 in Cycle II, an increase of 26.5 index points. This is a percentage-of-maximum confidence index, conditional on the source rubric's 1–5 convention. It

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

is not an activeness percentage or a zero-based proficiency score; the minimum possible value on this conversion is 20. Raw ratings remain the primary results.

Discussion

The increase in total speaking ratings is consistent in direction with the role-play findings of Krebt (2017) and Pham and Do (2025). The comparison is strongest at the level of a shared pattern of classroom improvement, rather than equivalent effect sizes: populations, measures, and designs differ. The present analysis adds a useful qualification by showing that improvement was uneven across dimensions and individuals. A single composite would obscure both unchanged pronunciation and the three students whose totals did not increase.

The largest change occurred in confidence. Supported rehearsal and familiar role-play situations could make speaking more manageable, which is compatible with the favorable perceptions reported by Hoang et al. (2023) and Zulfa et al. (2024). However, the present confidence ratings do not demonstrate reduced anxiety or increased willingness to initiate interaction. Neither psychological construct was measured separately. The exact agreement between Cycle II vocabulary and confidence scores also makes it difficult to interpret their development as independent processes without reviewing the original scoring evidence.

The confidence histograms help qualify the participation-related interpretation. The decline in rating-1 cases and emergence of ratings 4 and 5 are compatible with greater readiness to engage in spoken tasks. This interpretation is consistent with favorable confidence-related perceptions in Hoang et al. (2023) and Zulfa et al. (2024), but confidence gains cannot verify equal speaking opportunities or more frequent initiation. Pham and Do (2025) included participation observations, whereas the present analysis cannot make that behavioral comparison. Thus, confidence offers a relevant proxy for one aspect of participation readiness while leaving the activeness component of the research gap unresolved.

Vocabulary and fluency gains offer a more narrowly linguistic indication of development. Their combined increase remained visible after confidence was removed from the subtotal. This pattern connects with local PE research

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

emphasizing lexical resources (Febrianti et al., 2019; Putra et al., 2024), while extending the assessment focus to rated spoken performance. It does not show that vocabulary learning caused fluency gains, nor that the students could transfer their performance to unfamiliar professional situations. Such transfer would require new-task assessment beyond the practiced classroom context.

Pronunciation is the clearest boundary to a general improvement claim. Its stability contrasts with accounts of broader linguistic gains summarized by Kaygısız and Akar (2025). Repeated interaction may create opportunities to speak without necessarily providing sufficiently focused feedback on intelligibility. Alternatively, unchanged ratings could reflect limited scale sensitivity or a recording issue. The current records cannot distinguish these explanations. A subsequent cycle could combine listening to short models, focused rehearsal, and feedback on comprehensibility, accompanied by independent reassessment of recorded performances.

The pairing literature helps explain why improved scores cannot be attributed to pair composition alone. Zabihi and Ghahramanzadeh (2022) caution that mixed proficiency can coincide with weaker engagement, while Li and Liu (2022) emphasize collaborative interaction patterns. The present workbook lacks pair identifiers and interaction records, so neither partner dominance nor effective peer scaffolding can be established. The finding that most individuals improved does not resolve whether benefits were balanced within each pair. Future classroom records should link pair membership to speaking opportunities and individual assessments.

This distinction also limits the claim concerning activeness. Pham and Do (2025) combine speaking assessment with observation, whereas the present empirical analysis contains speaking ratings only. A stronger participation claim would require repeated observations of who initiates, responds, requests clarification, and makes room for a partner. Confidence and activeness may be related, but they remain different constructs. Keeping them separate preserves the value of the available results while identifying the evidence needed to address the remaining research gap.

The three teaching topics provide practical opportunities to collect that

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

evidence. In introductions, both partners can initiate an exchange and ask a follow-up question. In campus-life tasks, partners can exchange different information about locations or schedules, making clarification necessary. In describing people, one learner can describe a fictional participant while the other identifies the person and checks uncertain details. These are instructional implications derived from the framework and score pattern, not observed explanations for the gains. Their value would need to be assessed through documented implementation and comparable tasks.

For PE students, contextualization should make language tasks relevant without assuming that all learners share the same sporting interests or preferred learning style. Campus and sports-club situations can support accessible practice while preserving similar linguistic demands across students. This approach builds on the local rationale for disciplinary relevance, but the current evidence supports only a classroom-specific description. It does not establish that the intervention is uniquely effective for PE students or superior to its use in other non-English departments.

The study's main limitations are the absence of a verified baseline, comparison group, participation observations, pairing records, and rater-agreement evidence. Repeated testing, task familiarity, differences in assessment difficulty, or rater expectations could contribute to the observed change. The repeated score patterns warrant verification against original records. In addition, dates, instructional exposure, and cycle-specific revisions are not established by the dataset. These limitations prevent strong causal conclusions and restrict the contribution to descriptive changes in the available classroom ratings.

The overall improvement is consistent with the findings of Krebt (2017) and Pham and Do (2025), who reported speaking development following role-play instruction. The increase in confidence also aligns with the positive perceptions reported by Hoang et al. (2023) and Zulfa et al. (2024). The present study complements these accounts by documenting matched individual ratings and showing how their distributions changed between cycles. However, confidence represents only an indirect indicator of readiness to participate; it does not establish

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

that students initiated more exchanges, asked more questions, or received more balanced speaking opportunities.

The findings also reinforce the importance of interpreting pairing arrangements cautiously. Zabihi and Ghahramanzadeh (2022) showed that mixed-proficiency pairing does not automatically produce stronger engagement, while Li and Liu (2022) emphasized the importance of collaborative interaction patterns. Because the present dataset contains no pair-level interaction records, the observed gains cannot be attributed specifically to proficiency-based pairing or peer scaffolding. The contribution of this study therefore lies in identifying dimension-specific speaking development within a non-English-major PE cohort, including the continuing need for pronunciation support. Further implementation should combine structured role rotation and individual speaking opportunities with direct observations of participation, documented pairing procedures, and independently assessed oral performance.

CONCLUSION

The findings indicate an improvement in students assessed English speaking performance from Cycle I to Cycle II within the proficiency-based pairing and role-play framework. Among the 40 Physical Education students, the mean total score increased from 6.45 to 9.03, with 37 students showing improvement and three remaining unchanged. The histograms demonstrate a shift toward higher scores, although progress varied across students and assessment dimensions. Confidence showed the largest mean increase, from 1.300 to 2.625, followed by vocabulary and fluency, while pronunciation remained unchanged. These results suggest that the observed development was concentrated in confidence and selected aspects of speaking rather than distributed evenly across all dimensions.

REFERENCES

Cambridge English. (n.d.). *Assessing speaking performance: Level B1*. https://www.cambridgeenglish.org/Images/563276-b1-preliminary-assessing-speaking.pdf

Council of Europe. (2020). *Common European Framework of Reference for*

November 2026. Vol.8, No.2
e-ISSN. 2721-205X
pp. 172-191

Languages: Learning, teaching, assessment—Companion volume. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2020/16809ea0d4

Febrianti, M., Banat, A., & Juniardi, A. (2019). Website www.goal.com as an instructional model development. *ANGLO-SAXON: Jurnal Ilmiah Program Studi Pendidikan Bahasa Inggris, 10*(1), 43–55. https://doi.org/10.33373/anglo.v10i1.1922

Hoang, T. T., Pham, A. L., & La, T. T. (2023). Non-English majored students' perspectives on the use of role-play technique to improve speaking skills at Thai Nguyen University of Technology. *International Journal of Scientific and Research Publications, 13*(6), 441–446. https://doi.org/10.29322/IJSRP.13.06.2023.p13853

IELTS. (n.d.). *IELTS speaking band descriptors*. https://ielts.org/cdn/ielts-guides/ielts-speaking-band-descriptors.pdf

Kaygısız, S., & Akar, H. (2025). Role-playing as an instructional technique in English as a foreign language and English as a second language settings: A systematic review. *Cambridge Journal of Education, 55*(3), 441–467. https://doi.org/10.1080/0305764X.2025.2480579

Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research.* Springer. https://doi.org/10.1007/978-981-4560-67-2

Krebt, D. M. (2017). The effectiveness of role play techniques in teaching speaking for EFL college students. *Journal of Language Teaching and Research, 8*(5), 863–870. https://doi.org/10.17507/jltr.0805.04

Li, N., & Liu, Y. (2022). Effects of dyadic patterns and proficiency pairing on Chinese EFL learners' second language learning in collaborative writing. *Frontiers in Psychology, 13*, Article 1004924. https://doi.org/10.3389/fpsyg.2022.1004924

Pham, T. T., & Do, P. H. (2025). Using role-playing activities to enhance speaking skills of second-year non-English major students: Action research at a university in Hung Yen. *Journal of Applied Science and Technology, 48*, 68–74. https://jst.utehy.edu.vn/index.php/jst/article/view/836

Putra, A. R., Febrianti, M., Pradana, A., & Banat, A. (2024). The use of Wordwall.net as learning media to enhance vocabulary mastery of physical education students. *Scope: Journal of English Language Teaching, 9*(1), 74–79. https://doi.org/10.30998/scope.v9i1.24846

Zabihi, R., & Ghahramanzadeh, B. (2022). Proficiency pairing and engagement with language in L2 peer interaction. *System, 105*, Article 102725. https://doi.org/10.1016/j.system.2022.102725

Zulfa, P. F., Suryanti, E., & Tazali, I. (2024). ESP students' perception on role playing for boosting their speaking skill. *English Education and Literature Journal, 4*(2), 187–199. https://doi.org/10.53863/ejou.v4i02.1171