



THE IMPACT OF ESA TEACHING SEQUENCE ON EFL UNIVERSITY STUDENTS' SPEAKING FLUENCY AND CLASSROOM INTERACTION

Abdul Ayiz^{1*}, Kurniawan Yudhi Nugroho², Benni Ichsanda Rahman Hz³, Nurussyfa Luthfia Kholifah⁴, and Diego Javier Romero⁵

^{1,4}Universitas Duta Bangsa, Indonesia

²University Islam Sultan Agung, Indonesia

³Universitas Islam Negeri Sumatera Utara, Indonesia

⁵University of Osaka, Japan

*Corresponding Author: abdul_ayiz@udb.ac.id

DOI: <https://doi.org/10.65431/jrell.v2i1.34>

Abstract

The ESA (Engage, Study, Activate) teaching sequence introduced by Jeremy Harmer offers a flexible, recursive alternative to linear instructional models such as Presentation-Practice-Production (PPP). However, empirical evidence regarding its effectiveness for speaking fluency and classroom interaction, particularly in university EFL contexts, remains relatively limited. This quasi-experimental study investigated the impact of the ESA teaching sequence on speaking fluency and classroom interaction among university EFL learners, comparing it to conventional PPP instruction. Sixty first-year undergraduate students from two intact classes participated in an eight-week intervention. The control group (n = 30) received PPP instruction, while the experimental group (n = 30) received ESA instruction following the boomerang variation (Engage → Activate → Study → Activate). Speaking fluency was measured through speech rate (syllables per minute) and pause frequency (pauses per 100 syllables) using picture-cued speaking tasks administered as pretest and posttest. Classroom interaction was measured through systematic observations at Weeks 2, 4, 6, and 8, capturing student talk time, student-initiated turns, and negotiation episodes. The ESA group demonstrated statistically significant improvements in speech rate and pause frequency with large effect sizes, whereas the PPP group showed non-significant gains. Regarding classroom interaction, the ESA group exhibited substantially higher student talk time, more student-initiated turns, and more numerous negotiation episodes compared to the PPP group at post-intervention. These findings provide strong evidence within this context that the ESA teaching sequence enhances speaking fluency and promotes more balanced, learner-centered classroom interaction in university EFL contexts. Pedagogical implications include reallocating class time to communicative activities, reducing teacher talk time, delaying correction during fluency tasks, and explicitly teaching interactional strategies during the Study phase. Future research should examine long-term retention, accuracy outcomes, and replication across diverse educational settings.

Keywords: ESA Teaching Sequence, Classroom Interaction, PPP Instruction, Speaking Fluency, University EFL

INTRODUCTION

Speaking proficiency remains one of the most challenging competencies for university-level English as a Foreign Language (EFL) learners worldwide. Despite years of formal instruction, many first-year undergraduate EFL students in Indonesian universities (Setyawan et al., 2026). This persistent difficulty has been attributed to several factors, including limited opportunities for meaningful oral practice, high levels of speaking anxiety, and instructional models that prioritize grammatical accuracy over communicative fluency (Wang et al., 2024); the ESA (Engage-Study-Activate) sequence is proposed as a flexible alternative that directly addresses these limitations by emphasizing meaningful communication, lowering the affective filter, and providing recursive opportunities for authentic language use (Harmer, 1998).



Research has consistently demonstrated that traditional language teaching approaches, particularly the Presentation-Practice-Production (PPP) model, may inadvertently hinder the development of speaking fluency. The PPP sequence, characterized by teacher-led presentation of grammatical structures, controlled practice through drills, and delayed production, has been criticized for its artificial separation of skill development and its tendency to produce learners who can recite grammatical rules but cannot engage in real-time conversation. As Harmer (1998) observed, the linear nature of PPP fails to accommodate the recursive, unpredictable nature of authentic communication, where learners must simultaneously attend to meaning, form, and interactional dynamics.

The consequences of such instructional limitations are particularly evident in university EFL classrooms. Recent discourse analyses have revealed that classroom interaction remains heavily dominated by teacher speech, with students producing only minimal contributions, typically limited to short responses to teacher initiations (Rahmi et al., 2026). This pattern, often described as the Initiation-Response-Feedback (IRF) cycle, leaves learners with insufficient practice in initiating turns, negotiating meaning, or sustaining extended discourse (Pramudanti & Wasiatiningsih, 2025), whereas the ESA framework, particularly its boomerang variation, is designed to replace this teacher-dominated sequence with opportunities for student-initiated interaction and negotiated meaning (Harmer, 1998). Moreover, studies examining speech act distribution in EFL classrooms have found that lecturers predominantly employ representative and directive speech acts (explaining content, organizing activities), while students produce mostly expressive responses, with commissive and declarative acts being virtually absent (Rahmi et al., 2026). This interactional asymmetry reflects broader pedagogical and psychological dimensions. These include student anxiety, lack of confidence, and limited learner agency (Lu & Ramanair, 2025).

In response to the limitations of linear instructional models, Jeremy Harmer proposed the Engage-Study-Activate (ESA) framework as a more flexible and psychologically informed approach to language teaching (Harmer, 1998). The ESA sequence consists of three essential elements, each serving a distinct pedagogical purpose. The Engagement phase aims to emotionally and cognitively hook learners through provocative materials, personal anecdotes, or thought-provoking questions. This phase is theoretically grounded in Krashen's (1982) affective filter hypothesis, which posits that low-anxiety, high-motivation states facilitate language acquisition. The Study phase involves focused attention on language form, including grammar, vocabulary, discourse markers, or pronunciation, typically through guided discovery or explicit instruction. Unlike PPP's deductive presentation, Study in the ESA framework can be inductive and recursive, allowing teachers to respond to emergent learner needs. The Activate phase requires learners to use all available language resources in communicative tasks such as role-plays, debates, problem-solving activities, and information-gap exercises. During this phase, correction is minimal and delayed, prioritizing fluency over accuracy.

A distinctive feature of the ESA framework is its flexibility. Harmer (1998) describes three variations as the *straight arrow* sequence ($E \rightarrow S \rightarrow A$), suitable for predictable lessons; the *boomerang* sequence ($E \rightarrow A \rightarrow S \rightarrow A$), which begins with activation to diagnose learner needs before focused study; and the *patchwork* sequence (multiple recursive cycles of E, S, and A), designed for

complex lessons where learners require repeated exposure and practice (Harmer, 1998). This flexibility represents a significant advantage over PPP, as it allows teachers to adapt instruction based on ongoing assessment of learner performance.

Speaking fluency has emerged as a critical construct in second language acquisition research, particularly as scholars have moved beyond simplistic definitions based solely on speech rate. Contemporary frameworks distinguish between three dimensions of fluency: cognitive fluency (the speaker's ability to plan and retrieve language efficiently), utterance fluency (temporal features of speech such as speed and pausing), and perceived fluency (listener judgments of ease of communication) (Setyawan et al., 2026). Empirical studies have demonstrated that these dimensions are interrelated yet distinct, each contributing uniquely to overall speaking proficiency.

Recent meta-analytic research has identified speech rate (syllables per minute) and pause frequency (silent and filled pauses per 100 syllables) as the most robust behavioral indicators of utterance fluency (Suzuki et al., 2021). These measures have been shown to correlate strongly with holistic assessments of speaking proficiency and with listener perceptions of comprehensibility (Ginther et al., 2010; Suzuki et al., 2021). Longitudinal studies have, furthermore, documented that fluency development follows predictable patterns, with initial gains in speed followed by reductions in pausing behavior as automatization progresses. The relationship between fluency and accuracy has also received considerable attention. While some pedagogical approaches have treated fluency and accuracy as competing goals, recent evidence suggests they are complementary dimensions of proficiency that can develop simultaneously under appropriate instructional conditions (Wang et al., 2024). Specifically, tasks that provide opportunities for repeated, meaningful communication, such as those characteristics of the Activate phase in ESA, have been shown to promote both fluency and accuracy gains, particularly when learners receive opportunities for focused practice on problematic structures (Abdi Tabari et al., 2025; Buhot & Xing, 2025).

The quality of classroom interaction is widely recognized as a determinant of language learning outcomes. Drawing on Long's (1996) interaction hypothesis, which posits that negotiated interaction facilitates acquisition by providing comprehensible input, feedback, and opportunities for output modification, researchers have examined the patterns of talk that characterize effective language classrooms (Pramudanti & Wasiatiningsih, 2025). Of particular interest is the distribution of turn-taking between teachers and students, as well as the extent to which students initiate rather than merely respond to teacher prompts.

Observational studies of university EFL classrooms have documented persistent interactional asymmetries. The dominant pattern remains the IRF sequence, in which the teacher initiates a question, a student responds, and the teacher provides feedback (Rehman et al., 2025). While efficient for managing large classes and checking comprehension, this pattern has been criticized for limiting student talk time and reducing opportunities for extended discourse (Dalia & Putra, 2024). Moreover, students in IRF-dominated classrooms rarely initiate topics, ask clarification questions, or engage in peer-to-peer negotiation, behaviors that are essential for developing communicative competence (Alanazi & Widin, 2018; Rehman et al., 2025).

Affective factors also play a significant role in shaping classroom interaction. Speaking anxiety, defined as the fear or apprehension associated with real or anticipated oral communication in a second language, has been shown to negatively predict both willingness to communicate and actual speaking performance (Lu & Ramanair, 2025). University EFL learners are particularly susceptible to speaking anxiety due to performance pressures, fear of negative evaluation, and perceived inadequacy relative to peers. Recent questionnaire validation studies have confirmed that speaking anxiety, learning motivation, and learning attitudes are distinct but interrelated constructs that collectively influence speaking outcomes.

Although the ESA framework is widely cited in teacher training literature, empirical research on its effectiveness remains relatively limited. Existing studies have primarily focused on secondary school contexts and on general speaking outcomes rather than on specific fluency or interaction measures. Despite these limitations, several recent empirical studies have provided support for the effectiveness of ESA and CLT-based approaches in enhancing EFL speaking skills. In Myanmar, Seintita (2025) and Jotika (2025) reported substantial speaking gains among secondary school students following ESA-integrated and CLT-based interventions, respectively, with qualitative feedback indicating increased learner confidence and engagement. In the Indonesian context, Yogiswara et al. (2025) and Romadhona et al. (2023) similarly documented significant improvements in speaking performance among junior high school students, with Romadhona et al. reporting a statistically significant increase from pre-test ($M = 52.78$) to post-test ($M = 80.28$, $p < 0.001$). Agganyarni & Tran (2025) further extended these findings to Grade 6 students in Myanmar, reporting improved fluency, pronunciation, vocabulary, and interaction alongside reduced anxiety.

Collectively, these findings underscore the potential of ESA and CLT-based instructional frameworks to foster measurable gains in speaking proficiency across grade levels and national contexts. However, several gaps remain in the existing literature. First, no studies have specifically examined the impact of ESA on speaking fluency using objective temporal measures such as speech rate and pause frequency. Second, the effect of ESA on classroom interaction patterns, including turn-taking distribution and student-initiated talk, has not been systematically quantified. Third, the majority of ESA research has been conducted at the primary and secondary levels, with limited attention to university EFL contexts. Fourth, few studies have compared ESA directly to PPP, the dominant instructional model in many university programs.

However, several gaps remain in the existing literature. First, no studies have specifically examined the impact of ESA on speaking fluency using objective temporal measures such as speech rate and pause frequency. Second, the effect of ESA on classroom interaction patterns, including turn-taking distribution and student-initiated talk, has not been systematically quantified. Third, the majority of ESA research has been conducted at the primary and secondary levels, with limited attention to university EFL contexts. Fourth, few studies have compared ESA directly to PPP, the dominant instructional model in many university programs.

The present study is informed by three complementary theoretical perspectives. First, Krashen's (1982) affective filter hypothesis suggests that low-anxiety, high-motivation learning environments facilitate language acquisition

(Krashen, 1982). The Engagement phase of ESA is specifically designed to lower the affective filter by creating interest and personal relevance before demanding production. Second, Segalowitz's (2010) cognitive fluency framework conceptualizes fluency as the product of automatized lexical access and proceduralized grammatical knowledge (Segalowitz, 2010). The recursive Activate phases in the boomerang and patchwork variations of ESA provide repeated opportunities for controlled-to-automatic processing under real-time constraints, a condition essential for fluency development. Third, Long's (1996) interaction hypothesis, which has been supported and extended by more recent researches (Lambert & Philip, 2023; Sato & Ballinger, 2016), emphasizes the role of negotiated interaction in driving acquisition. The Activate phase of ESA, which requires learners to engage in unscripted communication with peers, is designed to maximize opportunities for meaning negotiation, clarification requests, and modified output, interactional features that have been empirically linked to language development.

Addressing the gaps identified above, this quasi-experimental study investigates the impact of the ESA teaching sequence on two key outcomes among university EFL learners: speaking fluency, operationalized through speech rate and pause frequency, and classroom interaction, operationalized through student talk time, student-initiated turns, and negotiation episodes. The study directly compares the ESA sequence to conventional PPP instruction over an eight-week intervention period. Two research questions guide the investigation. The first question examines whether the ESA teaching sequence significantly improves EFL university students' speaking fluency compared to conventional PPP instruction. The second question explores whether the ESA teaching sequence leads to more balanced classroom interaction than PPP instruction, as reflected in increased student talk time, a higher frequency of student-initiated turns, and more numerous negotiation episodes.

LITERATURE REVIEW

Theoretical Foundations of the ESA Framework

The Engage-Study-Activate (ESA) teaching sequence, introduced by Harmer (1998, 2007), differs fundamentally from linear instructional models such as Presentation-Practice-Production (PPP) in its theoretical underpinnings. Unlike PPP, which proceeds through fixed stages of teacher-led presentation, controlled practice, and final production, ESA offers a flexible, recursive framework that accommodates the unpredictable nature of real-time communication (Harmer, 2007). The three phases serve distinct pedagogical purposes: Engagement emotionally and cognitively hooks learners, lowering the affective filter and creating a need for communication (Krashen, 1982). Study focuses learners' attention on specific language forms through guided discovery or explicit instruction; and Activate requires learners to deploy all available linguistic resources in communicative tasks with minimal correction.

A distinctive feature of the ESA framework is its adaptability. Harmer describes three variations: the *straight arrow* sequence ($E \rightarrow S \rightarrow A$), suitable for predictable lessons where learners have minimal prior knowledge; the *boomerang* sequence ($E \rightarrow A \rightarrow S \rightarrow A$), which begins with activation to diagnose learner needs before focused study; and the *patchwork* sequence, which

incorporates multiple recursive cycles of E, S, and A for complex lessons (Harmer, 2007). This flexibility represents a significant pedagogical advantage over PPP, as it allows teachers to respond dynamically to emergent learner needs and to revisit problematic structures as they arise during communicative activities.

Speaking Fluency: Construct and Measurement

Speaking fluency has emerged as a critical construct in second language acquisition research, particularly as scholars have moved beyond simplistic definitions based solely on speech rate. Segalowitz (2010) proposed a tripartite framework distinguishing among cognitive fluency (the speaker's ability to plan and retrieve language efficiently), utterance fluency (temporal features of speech such as speed and pausing), and perceived fluency (listener judgments of ease of communication). This multidimensional conceptualization has informed subsequent empirical investigations, with researchers increasingly recognizing that fluency is not a unitary construct but rather a constellation of related but distinct phenomena (Suzuki et al., 2021; Szyzka et al., 2026). Of particular relevance to the present study, speech rate (syllables per minute) and pause frequency (pauses per 100 syllables) are established behavioral measures of utterance fluency, capturing the temporal features of speech that reflect automatization and cognitive processing efficiency (Suzuki et al., 2021).

The relationship between fluency and accuracy has also received considerable attention. While some pedagogical approaches have treated fluency and accuracy as competing goals, recent evidence suggests they are complementary dimensions of proficiency that can develop simultaneously under appropriate instructional conditions (Hasnain & Halder, 2024; Setyawan et al., 2026).

Task Repetition and Fluency Development

A substantial body of research has demonstrated that task repetition, the opportunity to perform the same or similar communicative tasks multiple times, promotes fluency development. Abdi Tabari et al. (2025) conducted a comprehensive meta-analysis synthesizing findings from multiple primary studies on task repetition and its effects on Complexity, Accuracy, Lexical Complexity, and Fluency (CALF) in L2 oral production. Their findings revealed that task repetition has a substantial impact on syntactic complexity and accuracy, with learners producing more complex and less erroneous utterances, while fluency, although positively influenced, exhibits smaller gains. This finding is particularly relevant to the ESA framework, as the boomerang variation's dual Activate phases (Activate 1 and Activate 2) constitute a form of task repetition: learners perform a communicative task, receive focused instruction during the Study phase, and then repeat a similar task, providing the repeated practice conditions that prior research has shown to promote complexity and accuracy development (Abdi Tabari et al., 2025; Harmer, 2007). Wang and Chen (2018) similarly found that immediate exact task repetition enhanced learners' discourse performance across multiple dimensions, including increased quantity of discourse, increased cohesion, and increased quality of lexis.

Critically, Buhot & Xing (2025) demonstrated that task repetition promotes fluency development, but the addition of focused practice on target structures is necessary to achieve simultaneous accuracy gains. In their experimental study,

learners who received accuracy enhancement with reduced cognitive load outperformed both control groups and groups without structured focus in terms of fluent use of target structures, with a large effect size. This finding aligns directly with the recursive structure of the ESA boomerang sequence, in which an initial Activate phase diagnoses learner needs, followed by a focused Study phase addressing problematic structures, and a second Activate phase allowing learners to apply newly acquired forms under communicative conditions.

Classroom Interaction and Turn-Taking Patterns

Observational studies of university EFL classrooms have documented persistent interactional asymmetries. The dominant pattern remains the Initiation-Response-Feedback (IRF) sequence, in which the teacher initiates a question, a student responds, and the teacher provides feedback (Rehman et al., 2025). While efficient for managing large classes and checking comprehension, this pattern has been criticized for limiting student talk time and reducing opportunities for extended discourse. Moreover, students in IRF-dominated classrooms rarely initiate topics, ask clarification questions, or engage in peer-to-peer negotiation, behaviors that are essential for developing communicative competence. The ESA framework directly addresses this limitation through its Activate phase, which deliberately shifts interactional control from teacher to students by requiring learners to participate in unscripted communicative tasks such as role-plays, debates, and problem-solving activities; tasks that naturally elicit student-initiated talk, clarification requests, and peer negotiation (Harmer, 2007). The distribution of turn-taking between teachers and students has been shown to correlate with learning outcomes. Teacher-dominated classrooms, characterized by extended teacher monologues and minimal student contributions, have been associated with lower levels of student engagement and reduced opportunities for output practice (Riad & Trimasse, 2023).

Conversely, classrooms with balanced turn allocation, in which students initiate topics and respond to one another, have been linked to higher levels of willingness to communicate and improved speaking performance (Nematizadeh & Wood, 2019). Dewi et al., (2025) found that task-based instruction encouraged collaborative speech acts and increased student-initiated turns compared to teacher-fronted instruction. Affective factors also play a significant role in shaping classroom interaction. Speaking anxiety, defined as the fear or apprehension associated with real or anticipated oral communication in a second language, has been shown to negatively predict both willingness to communicate and actual speaking performance (Sosa-López & Mora, 2022; Szyszka et al., 2026).

University EFL learners are particularly susceptible to speaking anxiety due to performance pressures, fear of negative evaluation, and perceived inadequacy relative to peers. Lu & Ramanair (2025) recently validated instruments measuring EFL speaking strategy use, speaking anxiety, learning motivation, and learning attitudes, confirming that these constructs are distinct yet interrelated and collectively influence speaking outcomes. Kitjaroonchai & Maywald (2024) similarly found that technology-mediated speaking practice reduced anxiety and improved fluency and accuracy among EFL university students.

METHOD

This study employed a quasi-experimental, non-equivalent control group, pretest-posttest design to investigate the impact of the ESA teaching sequence on speaking fluency and classroom interaction among university EFL learners. Sixty first-year undergraduate students from two intact classes at a mid-sized university in Indonesia participated in the study. The use of intact classes was necessitated by institutional constraints that prohibited random assignment of individual students to experimental conditions, a common limitation in educational research that maintains ecological validity by preserving authentic classroom structures (Creswell & Creswell, 2017). The control group ($n = 30$, PPP instruction) and experimental group ($n = 30$, ESA instruction) were comparable at baseline, with no significant differences in age, gender distribution, proficiency level, or pretest speaking fluency scores ($p > .05$). All participants provided written informed consent, and the study received institutional ethics approval.

Speaking fluency was assessed using a picture-cued speaking task administered as a pretest (Week 1) and posttest (Week 8), with two parallel forms counterbalanced across participants. Each participant had 30 seconds to prepare and then spoke for two minutes. Audio recordings were transcribed and analyzed using Praat software (Boersma & Weenink, 2021). Speech rate was calculated as syllables per minute, and pause frequency was calculated as the total number of silent pauses (≥ 0.25 seconds) and filled pauses per 100 syllables (Suzuki et al., 2021). Inter-rater reliability was excellent ($ICC = .92$ for speech rate; $.88$ for pause frequency). Prior to formal coding, two independent raters underwent a two-hour training session in which they reviewed the operational definitions of speech rate (syllables per minute, excluding hesitations and repetitions) and pause frequency (silent pauses ≥ 0.25 seconds and filled pauses such as 'um' or 'uh' per 100 syllables), practiced coding five pilot recordings not included in the main dataset, and discussed discrepancies until consensus was reached.

Classroom interaction was measured using a validated observation checklist. The checklist captured four variables: student talk time (percentage), teacher talk time (percentage), student-initiated turns (frequency per 30 minutes), and negotiation episodes (frequency per 30 minutes). Observations were conducted during 30-minute video-recorded sessions at four time points: Weeks 2, 4, 6, and 8.

The eight-week intervention consisted of 90-minute weekly sessions covering identical speaking topics across both groups. The control group received conventional PPP instruction: Presentation (25 min of explicit teacher-led input), Practice (35 min of controlled drills and gap-fills with immediate correction), and Production (25 min of final communicative tasks with ongoing correction). The experimental group received Harmer's (2007) boomerang ESA sequence: Engagement (15 min of affective/cognitive hook), Activate 1 (20 min of communicative task with teacher observation and no correction), Study (25 min of focused instruction based on observed gaps), Activate 2 (25 min of a second communicative task with minimal correction), and delayed feedback (5 min). Fidelity of implementation was confirmed via a checklist for 25% of sessions (ESA: 94%; PPP: 91%). This proportion aligns with methodological recommendations in educational research, which indicate that sampling 20–25% of sessions is sufficient for establishing fidelity when inter-rater reliability is high and observations are randomly selected (Strehlenert et al., 2024). Data analysis employed paired-

samples and independent-samples t-tests for fluency measures (speech rate and pause frequency) with effect sizes reported as Cohen's *d*. Mann-Whitney U tests were used for interaction variables (student talk time, student-initiated turns, negotiation episodes) due to violated normality assumptions. Chi-square tests compared teacher versus student talk time proportions. The alpha level was set at $p < .05$ for all analyses. Because the study employed only two primary fluency measures (speech rate and pause frequency) and conducted a limited number of planned comparisons (three per research question), no adjustment for multiple comparisons (e.g., Bonferroni correction) was applied, consistent with recommendations that corrections are most appropriate for large sets of post-hoc or exploratory tests (Armstrong, 2014). SPSS Version 28 (IBM Corp., 2021) was used for all statistical calculations.

RESULTS

Preliminary Analyses

Prior to addressing the research questions, preliminary analyses were conducted to ensure the comparability of the control and experimental groups at baseline. Independent-samples t-tests revealed no statistically significant differences between groups in age ($t(58) = 0.67, p = .51$), pretest speech rate ($t(58) = 0.81, p = .42$), or pretest pause frequency ($t(58) = 0.44, p = .66$). Chi-square tests indicated no significant differences in gender distribution ($\chi^2(1) = 0.07, p = .79$) or proficiency level distribution ($\chi^2(1) = 0.07, p = .79$). These findings confirm that the two groups were equivalent at the outset of the intervention.

Speech Rate in Speaking Fluency

Prior to addressing the research questions, preliminary analyses were conducted to ensure the comparability of the control and experimental groups at baseline. Independent-samples t-tests revealed no statistically significant differences between groups in age ($t(58) = 0.67, p = .51$), pretest speech rate ($t(58) = 0.81, p = .42$), or pretest pause frequency ($t(58) = 0.44, p = .66$). Chi-square tests indicated no significant differences in gender distribution ($\chi^2(1) = 0.07, p = .79$) or proficiency level distribution ($\chi^2(1) = 0.07, p = .79$). These findings confirm that the two groups were equivalent at the outset of the intervention.

Research Question 1 examined whether the ESA teaching sequence improves EFL university students' speaking fluency compared to conventional PPP instruction. Fluency was operationalized using two temporal measures: speech rate (syllables per minute) and pause frequency (pauses per 100 syllables). The results for each measure are presented below.

Descriptive statistics for speech rate are presented in Table 1. The experimental (ESA) group demonstrated a substantial increase in speech rate from pretest ($M = 98.4, SD = 13.9$) to posttest ($M = 128.7, SD = 14.6$), representing a mean gain of 30.3 syllables per minute (30.8% improvement). A paired-samples t-test confirmed that this within-group gain was statistically significant ($t(29) = 4.91, p < .01$), with a large effect size (Cohen's $d = 1.27$). In contrast, the control (PPP) group showed a more modest increase from pretest ($M = 101.2, SD = 14.3$) to posttest ($M = 113.4, SD = 15.1$), representing a mean gain of 12.2 syllables per minute (12.1% improvement). This within-group gain did not reach statistical

significance ($t(29) = 1.82$, $p = .08$), with a small-to-medium effect size (Cohen's $d = 0.47$).

Table 1. Descriptive and inferential statistics for speech rate (syllables per minute)

Group	Pretest (SD)	M	Posttest (SD)	M	Mean change	t(29)	p	Cohen's d
PPP (n=30)	101.2 (14.3)		113.4 (15.1)		+12.2	1.82	.08	0.47
ESA (n=30)	98.4 (13.9)		128.7 (14.6)		+30.3	4.91	<.01	1.27

A between-groups comparison of posttest speech rates using an independent-samples t-test revealed that the ESA group significantly outperformed the PPP group ($t(58) = 3.87$, $p < .01$), with a large effect size (Cohen's $d = 1.02$). These findings indicate that the ESA teaching sequence produced statistically significant and practically meaningful gains in speaking rate compared to conventional PPP instruction, as evidenced by the large effect size (Cohen's $d = 1.02$).

Pause Frequency

Descriptive statistics for pause frequency are presented in Table 2. The ESA group demonstrated a marked reduction in pause frequency from pretest ($M = 15.2$, $SD = 3.4$) to posttest ($M = 8.1$, $SD = 2.5$), representing a mean reduction of 7.1 pauses per 100 syllables (46.7% improvement). A paired-samples t-test confirmed that this within-group reduction was statistically significant ($t(29) = 5.33$, $p < .05$), with a large effect size (Cohen's $d = 1.41$). In contrast, the PPP group showed a minimal reduction from pretest ($M = 14.8$, $SD = 3.1$) to posttest ($M = 14.2$, $SD = 2.9$), representing a mean reduction of only 0.6 pauses per 100 syllables (4.1% improvement). This within-group reduction was not statistically significant ($t(29) = 0.41$, $p = .68$), with a negligible effect size (Cohen's $d = 0.11$).

Table 2. Descriptive and inferential statistics for pause frequency (pauses per 100 syllables)

Group	Pretest M (SD)	Posttest M (SD)	Mean change	t(29)	p	Cohen's d
PPP (n=30)	14.8 (3.1)	14.2 (2.9)	-0.6	0.41	.68	0.11
ESA (n=30)	15.2 (3.4)	8.1 (2.5)	-7.1	5.33	<.05	1.41

A between-groups comparison of posttest pause frequencies using an independent-samples t-test revealed that the ESA group produced significantly fewer pauses than the PPP group ($t(58) = -4.62$, $p < .01$), with a large effect size (Cohen's $d = 1.20$). These findings indicate that the ESA teaching sequence produced substantially greater reductions in pausing behavior compared to conventional PPP instruction, reflecting more automatized speech production.

Taken together, the fluency results demonstrate that the ESA teaching sequence significantly improved both dimensions of utterance fluency examined in this study, with large effect sizes for speech rate (Cohen's $d = 1.27$ within-group; $d = 1.02$ between-group) and for pause frequency (Cohen's $d = 1.41$ within-group; $d = 1.20$ between-group). These findings provide strong evidence that the ESA

sequence, particularly the boomerang variation with its recursive Activate phases, promotes greater automatization of speech production than conventional PPP instruction. The ESA group increased speech rate by 30.8% (large effect) and reduced pause frequency by 46.7% (large effect), whereas the PPP group showed non-significant improvements of 12.1% and 4.1%, respectively. These findings provide strong evidence that the ESA sequence, particularly the boomerang variation with its recursive Activate phases, promotes greater automatization of speech production than conventional PPP instruction.

Research Question 2 asked whether the ESA teaching sequence leads to more balanced classroom interaction than PPP instruction, as measured by student talk time (percentage), student-initiated turns, and negotiation episodes. Because interaction data violated normality assumptions, non-parametric Mann-Whitney U tests were employed for between-group comparisons. Descriptive statistics and inferential results for Week 8 (post-intervention) are presented in Table 3; trends across all observation weeks are summarized subsequently.

Table 3. Descriptive and inferential statistics for classroom interaction measures at week 8 (post-intervention)

Variable	PPP group (n=30) Median (Range)	ESA group (n=30) Median (Range)	Mann-Whitney U	p
Student talk time (%)	31% (24–38)	54% (48–62)	112.5	<.01
Student-initiated turns (per 30 min)	6 (3–9)	19 (14–25)	89.0	<.001
Negotiation episodes (per 30 min)	2 (0–4)	11 (7–16)	75.5	<.001

Student Talk Time

At Week 8, the ESA group demonstrated substantially higher student talk time (median = 54%, range = 48–62%) compared to the PPP group (median = 31%, range = 24–38%). A Mann-Whitney U test confirmed that this difference was statistically significant ($U = 112.5$, $p < .01$). Examination of trends across the eight-week intervention revealed that student talk time in the ESA group increased progressively from Week 2 (median = 29%) to Week 4 (median = 38%), Week 6 (median = 46%), and Week 8 (median = 54%). In contrast, student talk time in the PPP group remained relatively stable across the same period, increasing only marginally from Week 2 (median = 28%) to Week 4 (median = 29%), Week 6 (median = 30%), and Week 8 (median = 31%). A chi-square test comparing the proportion of teacher versus student talk time at Week 8 confirmed that the ESA group achieved a more balanced distribution (54% student, 46% teacher) compared to the PPP group (31% student, 69% teacher; $\chi^2(1) = 8.94$, $p < .01$).

Student-Initiated Turns

At Week 8, the ESA group produced substantially more student-initiated turns per 30-minute observation (median = 19, range = 14–25) compared to the PPP group (median = 6, range = 3–9). A Mann-Whitney U test confirmed that this difference was statistically significant ($U = 89.0$, $p < .001$). Longitudinal trends showed that student-initiated turns in the ESA group increased markedly from Week 2 (median = 5) to Week 4 (median = 10), Week 6 (median = 15), and Week

8 (median = 19). In the PPP group, student-initiated turns remained low and relatively stable across the intervention: Week 2 (median = 4), Week 4 (median = 5), Week 6 (median = 5), and Week 8 (median = 6). These findings indicate that the ESA sequence encouraged students to take greater communicative initiative without waiting for teacher nomination.

Negotiation Episodes

At Week 8, the ESA group produced substantially more negotiation episodes per 30-minute observation (median = 11, range = 7–16) compared to the PPP group (median = 2, range = 0–4). A Mann-Whitney U test confirmed that this difference was statistically significant ($U = 75.5$, $p < .001$). Longitudinal trends showed that negotiation episodes in the ESA group increased steadily from Week 2 (median = 2) to Week 4 (median = 5), Week 6 (median = 8), and Week 8 (median = 11). In the PPP group, negotiation episodes remained minimal throughout the intervention: Week 2 (median = 1), Week 4 (median = 1), Week 6 (median = 2), and Week 8 (median = 2). Qualitative observation notes indicated that negotiation episodes in the ESA group typically involved clarification requests (e.g., "What do you mean by...?"), comprehension checks (e.g., "So you're saying that...?"), and repair sequences following communication breakdowns. Such episodes were virtually absent in the PPP classroom.

Qualitative Observations of Interaction Patterns

Beyond the quantitative measures, observers noted distinct qualitative differences in interaction patterns between the two groups. In the ESA classroom, students frequently interrupted each other politely using expressions such as "Sorry, can I add something?" and employed backchannels (e.g., "Right," "Uh-huh," "I see") to signal active listening. Teacher talk time decreased progressively from approximately 60% in Week 2 to 40% in Week 8. In contrast, the PPP classroom remained teacher-fronted throughout the intervention. Even during the Production phase, teachers continued to nominate speakers, and student turns were typically limited to single sentences rather than extended exchanges. Students rarely spoke without explicit teacher nomination, and peer-to-peer negotiation was minimal.

DISCUSSION

The findings of the present study both corroborate and extend the existing empirical literature on the ESA teaching sequence. As summarized in table 2 and 3 previously, the present study's results align with previous research in demonstrating positive effects of ESA on speaking outcomes (Seintita, 2025; Yogiswara et al., 2025; Romadhona et al., 2023), while extending the literature by (a) focusing on university EFL learners, (b) employing objective temporal fluency measures (speech rate and pause frequency), and (c) systematically quantifying classroom interaction patterns. The significant improvements in speaking performance observed in the ESA group are consistent with multiple earlier investigations. Hairunnisa et al. (2022) demonstrated that ESA instruction produced statistically significant improvements in tenth-grade students' speaking ability. Similarly, Khoshshima & Shokri (2016) found that university EFL learners receiving ESA instruction significantly outperformed a control group receiving audio-lingual instruction on a posttest of speaking ability. In the Myanmar context, Seintita (2025)

reported that an eight-week CLT-ESA integrated intervention produced substantial improvements across pronunciation, grammar accuracy, vocabulary use, fluency, and comprehension, with qualitative feedback indicating increased learner confidence and active engagement. The present study aligns with these findings while extending them by providing objective, fine-grained temporal measures of fluency, speech rate and pause frequency, rather than relying solely on holistic speaking scores.

While the majority of studies have reported positive effects of ESA, one study has produced divergent findings that warrant discussion. Hulwana (2024) compared ESA and PPP methods and found that ESA did not produce significantly different speaking outcomes from conventional instruction, whereas PPP did show a positive effect. Several factors may explain this divergence from the present study's findings: (1) ESA variation, Hulwana did not specify which ESA variation was used, whereas the present study employed the boomerang variation with dual Activate phases; (2) outcome measures, Hulwana used holistic speaking tests, while the present study employed fine-grained temporal fluency measures more sensitive to ESA effects; and (3) intervention duration, the present study's eight-week intervention may have provided sufficient exposure for ESA effects to manifest. This divergence highlights that ESA effectiveness depends on implementation quality, variation selection, and alignment between outcome measures and instructional goals.

The present study extends the existing ESA literature in four key respects. Unlike previous research that relied on holistic speaking tests, this study employed objective temporal fluency measures (speech rate and pause frequency) grounded in Segalowitz's (2010) cognitive fluency framework. It also systematically quantified classroom interaction patterns—student talk time, student-initiated turns, and negotiation episodes—using validated observation instruments, variables linked to language acquisition through Long's (1996) interaction hypothesis. Additionally, this study directly compared ESA to PPP under controlled conditions and focused on university-level EFL learners, an under-researched population, demonstrating ESA's effectiveness for adult learners facing academic speaking demands and performance anxiety.

The convergence of findings across multiple studies, despite differences in educational levels (secondary vs. university), geographical contexts (Indonesia, Myanmar, Iran), and outcome measures, provides preliminary evidence that the ESA framework may possess theoretical robustness and cross-contextual applicability within similar Asian EFL settings. The present study's findings, interpreted alongside previous research, support three theoretical propositions consistent with established SLA frameworks. First, the recursive structure of ESA (particularly the boomerang variation) promotes automatization, consistent with Segalowitz's (2010) cognitive fluency framework. Second, the Engagement phase lowers the affective filter, consistent with Krashen's (1982) affective filter hypothesis. Third, the Activate phases generate negotiation episodes, consistent with Long's (1996) interaction hypothesis. Collectively, the accumulating evidence from multiple studies across diverse contexts supports the conclusion that Harmer's ESA teaching sequence is an effective and theoretically sound approach for developing speaking skills in EFL learners.

CONCLUSION

This quasi-experimental study investigated the impact of Jeremy Harmer's ESA teaching sequence on speaking fluency and classroom interaction among university EFL learners, comparing it to conventional PPP instruction over an eight-week intervention. The findings provide encouraging evidence that the ESA sequence, implemented in the boomerang variation, outperforms PPP on both dimensions within the context of this study. The ESA group demonstrated substantial improvements in speech rate and pause frequency, key behavioral indicators of utterance fluency, whereas the PPP group showed only minimal, non-significant gains. Furthermore, the ESA group exhibited markedly more balanced classroom interaction patterns, characterized by greater student talk time, a higher frequency of student-initiated turns, and more numerous negotiation episodes compared to the PPP group. These findings align with the theoretical predictions of Krashen's affective filter hypothesis, Segalowitz's cognitive fluency framework, and Long's interaction hypothesis, while extending the limited empirical literature on ESA by providing evidence of its effectiveness for objective fluency development and systematic interactional outcomes in a university EFL context.

The pedagogical implications of these findings are substantial. Based on the ESA group's superior fluency gains (speech rate: +30.8%; pause frequency: -46.7%), instructors should replace PPP with ESA's boomerang variation, allocate sufficient time for communicative activities, and delay correction during fluency tasks. In light of the ESA group's more balanced interaction patterns (student talk time: 54% vs. 31%; student-initiated turns: 19 vs. 6; negotiation episodes: 11 vs. 2), interactional strategies should be explicitly taught during the Study phase. Finally, curriculum designers should adopt the ESA framework as a principled alternative to linear instructional models.

Several limitations of this study warrant acknowledgment, each with potential implications for interpreting the results. The quasi-experimental design may have introduced selection bias, though baseline equivalence was confirmed. The eight-week intervention may capture short-term gains rather than durable automatization. The absence of a delayed posttest means retention of effects is unknown. The lack of accuracy measures leaves potential fluency-accuracy trade-offs unexplored. The single-context sample limits generalizability to other populations and settings. Readers should therefore interpret the findings as encouraging but preliminary, pending replication and extension.

References

- Abdi Tabari, M., Zhuang, J., & Farahanynia, M. (2025). Task repetition and L2 oral performance: A meta-analysis. *System*, 135, 103868. <https://doi.org/10.1016/j.system.2025.103868>
- Alanazi, M. J., & Widin, J. (2018). *Exploring the Role of Teacher Talk in Saudi EFL Classroom: Importance of F-Move in developing Students' Spoken Skill*. 9(1), 307–320. <https://doi.org/10.24093/awej/vol9no1.22>
- Boersma, P., & Weenink, D. (2021). *Praat: Doing Phonetics by Computer* [Graphic]. <https://praat.org/>
- Buhot, N., & Xing, Q. (2025). The addition of a target structure to task repetition as an accuracy enhancement: The necessity of reducing cognitive load. *Languages*, 10(6), 128. <https://doi.org/10.3390/languages10060128>

- Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*.
- Dalia, A. C., & Putra, F. R. (2024). The initiation, response, and feedback (IRF) in English language teaching: a literature review. *Jurnal Ilmiah Spectral*, 10(1), 043–053. <https://doi.org/10.47255/ya5nfp19>
- Dewi, I. S., Fatimah, F., Fauziningrum, E., Kritandani, W., & Sani, A. (2025). Anchoring task-based learning to foster cadets' sustainable English-speaking skills. *EduLite: Journal of English Education, Literature and Culture*, 10(2), 254–270. <https://doi.org/10.30659/e.10.2.254-270>
- Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, 27(3), 379–399. <https://doi.org/10.1177/0265532210364407>
- Hairunnisa, A., Rahman, M. A., & Baa, S. (2022). The Effects of Engage, Study, Activate (ESA) method on 10 year students' English speaking ability. *Journal of Excellence in English Language Education*, 1(2). <https://eprints.unm.ac.id/28971/>
- Harmer, J. (1998). *How to teach English: An introduction to the practice of English language teaching*. Longman.
- Harmer, J. (2007). *The practice of English language teaching*. Harlow, England: Pearson Longman. http://archive.org/details/practiceofenglis0000harm_d2l2
- Hasnain, S., & Halder, S. (2024). Intricacies of the multifaceted triad-complexity, accuracy, and fluency: A review of studies on measures of oral production. *Journal of Education*, 204(1), 145–158. <https://doi.org/10.1177/00220574221101377>
- Hulwana, H. (2024). The effect of ESA and PPP method on students' speaking skill | *Jurnal Pendidikan Bahasa Inggris Undiksha*. <https://ejournal.undiksha.ac.id/index.php/JPBI/article/view/65547>
- Khoshsima, H., & Shokri, H. (2016). The effects of ESA Elements on speaking ability of intermediate EFL learners: A task-based approach. *Theory and Practice in Language Studies*, 6(5), 1085. <https://doi.org/10.17507/tpls.0605.24>
- Kitjaroonchai, N., & Maywald, S. (2024). The effects of reading assistant software on the speech fluency and accuracy of EFL university students. *JET (Journal of English Teaching)*, 10(2), 183–197. <https://doi.org/10.33541/jet.v10i2.5763>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition* (1st ed). Pergamon.
- Lambert, C., Aubrey, S., & Bui, G. (2023). *The Role of the learner in task-based language teaching: Theory and research methods* (1st ed.). Routledge. <https://doi.org/10.4324/9781003227267>
- Long, M. H. (1996). *The role of the linguistic environment in second language acquisition*. Academic Press. <https://www.scirp.org/reference/referencespapers?referenceid=476624>
- Lu, W., & Ramanair, J. (2025). Development and validation of EFL speaking strategy, speaking anxiety, learning motivation, and learning attitude

- questionnaires. *Journal of Global Research in Education and Social Science*, 19(3), 33–49. <https://doi.org/10.56557/jogress/2025/v19i39354>
- Nematizadeh, S., & Wood, D. (2019). Willingness to communicate and second language speech fluency: An investigation of affective and cognitive dynamics. *The Canadian Modern Language Review*, 75(3), 197–215. <https://doi.org/10.3138/cmlr.2017-0146>
- Pramudanti, Y., & Wasiatiningsih, E. (2025). Turn-taking patterns in online English lessons for young learners: An IRF-based discourse analysis. *The Proceedings of English Language Teaching, Literature, and Translation (ELTLT)*, 14, 675–689.
- Rahmi, R., Uswatunnisa, & Marlina, R. (2026). *Patterns of lecturer and student speech acts in EFL classroom interaction: Qualitative study insights | new dimensions*. <https://journal.unesa.ac.id/index.php/nld/article/view/48062?articlesBySimilarityPage=2#articlesBySimilarity>
- Rehman, A., Noor, F., Chaudhary, H. N., & Yasmeen, I. (2025). Analyzing initiation-response-feedback patterns: An experimental investigation of classroom discourse in Pakistani educational settings. *Journal of Applied Linguistics and TESOL (JALT)*, 8(3), 2571–2587. <https://doi.org/10.63878/jalt1285>
- Sato, M., & Ballinger, S. (Eds.). (2016). *Peer Interaction and Second Language Learning: Pedagogical potential and research agenda* (Vol. 45). John Benjamins Publishing Company. <https://doi.org/10.1075/llt.45>
- Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. Routledge. https://www.academia.edu/28808259/Cognitive_Bases_of_Second_Language_Fluency
- Seintita, V. (2025). Development of communicative language teaching (CLT) Integrated with the ESA Instructional model to enhance English-speaking skills in basic education high school grade-ten students in Myo Chaung, Myanmar. *Journal of Asian Language Teaching and Learning*, 6(2), 194–200.
- Setyawan, S., Dewangga, V., & Pudyastuti, Z. E. (2026). Speech fluency and accuracy as indicators of academic speaking proficiency. *Journal of English in Academic and Professional Communication*, 12(1), 79–91. <https://doi.org/10.25047/jeapco.v12i1.6019>
- Sosa-López, G., & Mora, J. C. (2022). The Role of speaking anxiety on L2 English speaking fluency, accuracy and complexity. *Pronunciation in Second Language Learning and Teaching Proceedings*, 12(1). <https://doi.org/10.31274/psllt.13362>
- Strehlenert, H., Hedberg Rundgren, E., Sjunnestrand, M., & Hasson, H. (2024). Fidelity to and adaptation of evidence-based interventions in the social work literature: A scoping review. *The British Journal of Social Work*, 54(3), 1356–1376. <https://doi.org/10.1093/bjsw/bcad170>
- Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal*, 105(2), 435–463. <https://doi.org/10.1111/modl.12706>
- Szyszk, M., Peltonen, P., & Lintunen, P. (2026). Unravelling the relationship between language anxiety and foreign language speech fluency in a

monologue production. *Journal of Multilingual and Multicultural Development*, 47(1), 323–337.
<https://doi.org/10.1080/01434632.2024.2387149>

Wang, W., Rezaei, Y. M., & Izadpanah, S. (2024). Speaking accuracy and fluency among EFL learners: The role of creative thinking, emotional intelligence, and academic enthusiasm. *Heliyon*, 10(18), e37620.
<https://doi.org/10.1016/j.heliyon.2024.e37620>