

The Impact of Online Task Planning Conditions on L2 Utterance Fluency and Self-Perceived Fluency: A Study of Chinese EFL Learners

Xu Han¹

Department of English, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Kuala Lumpur, Malaysia

Yasir Yahya²

Department of English, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Kuala Lumpur, Malaysia

Ngee Thai Yap

Department of English, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Kuala Lumpur, Malaysia

Abstract—This study examines the effects of online task planning conditions on L2 learners' utterance fluency and self-perceived fluency among Chinese EFL learners. Ninety participants were randomly assigned to Pressured Online Planning (POP), Unpressured Online Planning (UOP), and Hybrid Online Planning (HOP) groups. Participants completed a narrative task based on a Mr. Bean video clip under their respective planning conditions. Utterance fluency was measured using temporal and linguistic indicators, while self-perceived fluency was assessed through CEFR self-assessment and an Analytic Fluency Perception Scale. Results indicate that the hybrid condition yielded significantly higher speech and articulation rates, with fewer disfluencies compared to other conditions ($p < .001$). Notably, self-perceived fluency did not consistently align with objective measures or rater evaluations, particularly in the pressured condition. The study reveals complex relationships between task planning conditions, objective fluency measures, and learners' self-perceptions, contributing to our understanding of L2 speech production processes and informing task-based language teaching practices.

Index Terms—online task planning, utterance fluency, perceived fluency

I. INTRODUCTION

In the field of second language acquisition (SLA), achieving fluency continues to be a major challenge for English as a Foreign Language (EFL) learners, particularly in the Chinese context. Despite extensive classroom instruction, many students struggle to translate their theoretical knowledge into practical, real-time language use, revealing a crucial gap between understanding and applied linguistic skills (Ellis, 2017; Ur, 2013). Task-Based Language Teaching (TBLT) has gained prominence as an approach to bridge this gap, with task planning conditions recognized as key factors influencing learners' utterance performance, particularly in terms of complexity, accuracy, and fluency (CAF) (Robinson, 2015; Skehan, 2016).

Recent research has explored various task planning strategies, including pre-task planning (Khatib & Farahanynia, 2020; Lampropoulou, 2023; Qiu & Bui, 2022), within-task planning (Awwad & Alhamad, 2021; Gu & Van den Branden, 2024; Kim, 2020) and hybrid planning approaches (Hsu, 2017; Wang et al., 2019), offering valuable insights into how these conditions affect L2 learners' utterance performance. However, there remains a notable gap in understanding the impact of specific online task planning conditions—namely pressured, unpressured, and hybrid—on both utterance fluency and learners' self-perceived fluency, especially in dialogic tasks that mirror real-world communication more closely.

Theoretical frameworks such as Skehan's (1998) Limited Attentional Capacity Hypothesis (LACH) and Robinson's (2001) Cognition Hypothesis (CH) have been influential in task-based language research. LACH posits that learners have limited attentional resources, leading to trade-offs between different aspects of performance, while CH suggests that increasing task complexity can lead to improvements in multiple performance areas simultaneously. These models provide different perspectives on how task characteristics and conditions influence L2 performance, yet their implications for online task planning conditions remain underexplored. This study aims to fill this gap by examining the

¹ Corresponding Author¹. Email: gs59514@student.upm.edu.my

² Corresponding Author². Email: muhammadyasir@upm.edu.my

effects of different online task planning conditions on the utterance fluency and self-perceived fluency among Chinese EFL learners.

II. LITERATURE REVIEW

A. *Task-Based Language Teaching (TBLT) and Task Planning Conditions*

(a). *TBLT and Theoretical Models of Task Provision*

In the field of second language acquisition (SLA) and instruction, Task-based Language Teaching (TBLT) has garnered significant attention for over three decades and remains one of the most prevalent approaches (Ur, 2013). TBLT emphasizes communicative and collaborative activities as the foundation for planning and delivering instruction (Bygate, 2016; Long, 2016). It operates on the premise that language is a means of communication and can be effectively acquired through exposure and discussion during the performance of communicative tasks (Ur, 2013). Students engage in task-based activities by completing real-world tasks that focus on meaningful interaction and the practical use of language (Ellis, 2017). This principle suggests that students learn the target language more effectively when they are task-focused and given opportunities for authentic language usage in the classroom.

In psycholinguistic task-based language research, both the Limited Attention Capacity Hypothesis (LACH) and the Cognition Hypothesis (CH) have been influential, particularly in terms of task sequencing criteria in TBLT. Both hypotheses view tasks as the primary unit of analysis in language instruction, suggesting they should be sequenced according to specific criteria. While LACH and CH differ in defining task characteristics and conditions, both frameworks allow teachers to sequence tasks and determine their influence on second language (L2) students' task performance in terms of complexity, accuracy, and fluency (CAF).

Skehan (2009, 2014, 2015, 2016) explained LACH by distinguishing between task characteristics, which determine task difficulty based on the tasks' inherent features (e.g., cognitive complexity, code complexity), and task conditions, which determine task difficulty based on task performance (e.g., communicative stress). On the other hand, Robinson's (2001, 2005, 2007, 2015) CH offers a triadic componential framework for task sequencing, consisting of task complexity (cognitive factors related to task design), task condition (interactional factors regarding participation and participant variables), and task difficulty (learner factors regarding affective and ability variables). LACH posits that task characteristics and conditions should be included in evaluating task difficulty, with difficult tasks resulting in either an increase in complexity or accuracy at the expense of the other (Skehan, 2016). CH, however, argues that only task characteristics (i.e., task complexity) are used to evaluate task complexity, with complex tasks associated with simultaneous improvements in accuracy and complexity (Robinson, 2015).

Skehan (1998) and Robinson (2001) have proposed models representing the crucial elements of task complexity and describing how these elements contribute to task performance. Task complexity is critical to pedagogical task design and sequencing decisions. Skehan's model suggests that attentional capacity is limited, meaning that fluency, accuracy, and complexity compete for cognitive resources. Studies manipulating planning time or task requirements have found effects on both accuracy and complexity (Foster & Skehan, 1997; Skehan, 2009). Conversely, Robinson maintains that students can simultaneously draw on multiple cognitive resources, with fluency, accuracy, and complexity not necessarily in competition. Robinson categorizes task complexity dimensions into resource-directing and resource-dispersing.

(b). *Task Planning Conditions and Previous Studies*

Planning in oral task performance can be categorized into pre-task planning and within-task (online) planning (Ellis, 2005, 2009). Pre-task planning includes rehearsal and strategic planning, where students prepare for tasks beforehand. Online planning occurs during task performance and can be pressured or unpressured. Pressured planning involves time constraints, whereas unpressured planning allows adequate time for careful monitoring of language (Ahmadian & García Mayo, 2017). Combining pre-task and online planning can reduce cognitive load and enhance language output, referred to as 'hybrid online planning' (Ellis, 2009). Task difficulty, according to Robinson's CH, includes factors related to individual differences in learning and task implementation, classified into ability dimensions (e.g., working memory, aptitude) and affective dimensions (e.g., task motivation, processing anxiety). While Robinson did not explicitly distinguish between task characteristics and conditions, this study uses these terms consistently with Skehan's explanation.

Several studies have explored the effects of planning time conditions on oral performance, focusing on accuracy and complexity but giving less attention to fluency (Yuan & Ellis, 2003). Pre-task planning has been shown to improve L2 students' fluency, lexical complexity, and syntactic complexity. For example, Khatib and Farahanynia (2020) found that pre-task planning and repetition improved fluency and syntactic complexity, supporting the Trade-off Hypothesis. D'Ely et al. (2019) demonstrated that pre-task planning and repetition positively affected accuracy, complexity, and lexical density. Nielson and DeKeyser (2019) found that guided pre-task planning improved accuracy, while unguided pre-task planning improved fluency. These studies collectively highlight that pre-task planning generally benefits multiple dimensions of oral performance, suggesting that providing students with time to prepare before tasks can lead to enhanced language output across various metrics.

Research on within-task planning conditions indicates that unpressured planning allows for better monitoring and careful language use, improving fluency and accuracy (Awwad & Alhamad, 2021; Kim, 2020). However, findings on the effects of pressured planning are mixed, with some studies showing improvements in fluency and others highlighting trade-offs with accuracy and complexity. For instance, Panahzadeh and Asadi (2019) found that pressured online planning led to higher fluency and coherence but at the expense of grammatical accuracy. Similarly, Kim (2018) noted that while pressured planning improved fluency, it did not significantly enhance accuracy or complexity. These mixed results suggest that the benefits of within-task planning may depend on the specific conditions and the learners' abilities to manage cognitive load under time constraints.

Meta-analyses and comparative studies indicate that pre-task planning generally enhances fluency, complexity, and accuracy more effectively than within-task planning. Johnson and Abdi Tabari's (2022) meta-analysis, which included 52 studies, demonstrated that pre-task planning significantly improves syntactic complexity and accuracy, as well as fluency in L2 production. Similarly, Bakhtiary et al. (2021) found that strategic planning was more beneficial for oral production than unpressured online planning. These findings align with the Cognition Hypothesis (Robinson, 2011), suggesting that allowing learners to plan before tasks enables them to allocate cognitive resources more effectively, thus improving overall performance.

Hybrid online planning, which combines pre-task and within-task planning, has been shown to enhance accuracy and complexity of speech. Studies by Wang et al. (2019) and Hsu (2017) suggest that integrating multiple planning methods can optimize task performance and address trade-offs between fluency, accuracy, and complexity. Wang et al. (2019) found that hybrid planning conditions led to better accuracy and complexity compared to either pre-task or within-task planning alone. Hsu (2017) also noted that combining rehearsal with online planning improved language output significantly. These studies highlight the potential of hybrid planning to balance the demands of fluency, accuracy, and complexity, providing a more comprehensive approach to task-based language instruction.

Theoretical models such as LACH and CH provide valuable insights into how task planning conditions influence L2 speech production. While pre-task conditions generally improve accuracy, complexity, and fluency, within-task conditions remain controversial regarding their effect on fluency. This study aims to examine different task conditions (hybrid, unpressured, and pressured) and their impact on learners' utterance fluency, addressing the identified research gap.

B. Exploring Fluency in L2 Speech Performance

(a). Definitions and Theories of Fluency

The nature of L2 speech performance is multi-componential. The CAF framework (i.e., complexity, accuracy, and fluency) was developed to assist researchers in analyzing this multi-dimensional and multi-faceted phenomenon of L2 speech performance (Housen et al., 2012), and these language constructs or aspects can be used to describe, quantify, and measure L2 students' language production and language proficiency (Housen & Kuiken, 2009). Norris and Ortega (2009) established empirically that the characteristics of CAF components are interdependent and distinctive, which indicates each component of CAF has its own specific role in L2 proficiency. L2 students are, however, said to have limited pools of attention and limited proficiency in L2 to be able to pay attention to only one aspect of CAF at a time (Skehan, 1998, 2015). Based on Skehan's Limited Attentional Capacity Hypothesis (LACH), a trade-off effect is expected to emerge between the three dimensions (i.e., CAF) of speech performance. Thus, the current study focuses on one dimension (i.e., fluency) of L2 speech performance.

During the earlier stages, Fillmore (1979) defined fluency as the ability to talk for long periods without pauses, to talk in coherent and structured sentences, to communicate effectively in a variety of social and professional settings, and to use the language in a creative and innovative manner. Although Fillmore's definition of fluency was originally proposed for first language (L1), it is considered one of the seminal perspectives on fluency and has provided the foundation for L2 discussions of the concept. Lennon (1990) expanded on Fillmore's work, distinguishing between broad and narrow fluency. Broad fluency pertains to oral proficiency, while narrow fluency refers to the speed at which speech is produced. Tavakoli and Hunter (2018) further illustrated fluency as a pyramid with four layers, distinguishing between broad and narrow perspectives. For this study, fluency is defined using Lennon's narrow definition and Tavakoli and Hunter's narrow and very narrow perspectives.

(b). L1 and L2 Speech Production Models

A well-established theoretical framework for interpreting fluent and non-fluent speech performances is Levelt's (1989) first language (L1) production model, which includes three main stages: conceptualizing the intended message, formulating a preverbal message, and articulating speech. Fluent speech is defined as knowing the ideas in advance (conceptualizing), possessing the language to express these ideas (formulating), and being able to construct or translate these ideas into speech (articulating). This L1 speech production model is applied in Kormos's (2006) L2 speech production model, highlighting some significant distinctions between L1 and L2 processes. In the conceptualization stage, bilingual speakers choose which language to use when forming communication intentions. The formulation and articulation stages are intuitive for L1 speakers; however, L2 speakers require assistance with conceptualization and monitoring to perform fluently.

The three stages of L2 speech production are related to utterance fluency, which can be assessed through speed, breakdown, and repair indexes (Lambert et al., 2021; Skehan, 2016; Suzuki et al., 2021). Speed fluency can assess the overall efficiency of these stages. Filled or unfilled pauses (breakdown) are related to the conceptualization and formulation stages, indicating that the L2 speaker is planning or considering (De Jong et al., 2021; Kormos, 2006; Tavakoli et al., 2020; Van Os et al., 2020) what to say. Monitoring in speech production can be reflected in repair, reformation, or repetition. Therefore, fluency requires parallel and stable relationships between the phases of the speech model. In line with this model, pre-task planning assists the conceptualization stage, resulting in more fluency, whereas within-task planning improves the formulation of the message, leading to increased complexity and accuracy (Ellis, 2009; Skehan, 1998). The present study evaluates performance on a specific language dimension, fluency, in L2 by using hybrid online planning.

(c). Characteristics and Measurements of L2 Fluency

From a psycholinguistic perspective, Segalowitz (2010) describes three distinct characteristics of fluency based on Levelt's (1989) model: cognitive, utterance, and perceived fluency. Cognitive fluency refers to the ability to control the underlying cognitive systems associated with speech production to ensure the smooth flow of speech. Utterance fluency is an extension of cognitive fluency and can be measured through temporal features such as speed, breakdown, and repair. Perceived fluency is what listeners judge about speakers' performances, often influenced by speed and breakdown aspects.

Temporal measures related to speed and pausing were extensively used in earlier L2 fluency research (Lennon, 1990; Raupach, 1980). Skehan (2003) and Tavakoli and Skehan (2005) added repair measures, developing a comprehensive fluency framework consisting of speed, breakdown, and repair fluency. Speed fluency measures how fast a speaker delivers the message, breakdown fluency is concerned with disruptions in the flow of speech, and repair fluency describes monitoring processes and repair strategies. These dimensions have been widely accepted and used in assessing L2 fluency (De Jong, 2018; Suzuki & Kormos, 2022; Tavakoli et al., 2020; Van Os et al., 2020). Common measures of utterance fluency include articulation rate (AR), speech rate (SR), mean length of run (MLR), and phonation time ratio (PTR). Breakdown fluency measures include the frequency, duration, and location of filled and silent pauses. Repair fluency measures include false starts, repetitions, reformulations, and replacements.

Perceived fluency is typically evaluated by native speaker listeners. However, in this study, in addition to evaluations by native speaker raters, students will also participate in assessing their own speaking performance using a self-assessment scale to provide their self-perception of fluency. Specifically, this research examines: (1) the impact of pressured, unpressured, and hybrid planning conditions on objective measures of utterance fluency and subjective measures of perceived fluency and (2) the relationship between objective fluency measures and subjective self-assessments under different planning conditions.

III. METHODOLOGY

A. Participants and Study Setting

Ninety native Chinese speakers, consisting of sixty females and thirty males aged 18 to 21, participated in this study as volunteers. They were first-year Medical Science students at a university in Hainan and had not lived in an English-speaking country for more than three months. All of them had passed the National College Entrance Examination (NEMT) with band scores between 90 and 110 (maximum possible = 150), which grades between A and B in the oral component of this examination. Based on their self-reported proficiency from a pre-task questionnaire, their overall English proficiency ranged from 6 to 7 on the IELTS, with speaking scores between 5.5 and 6. These scores roughly correspond to B2 to C1 levels of the Common European Framework of Reference for Languages (CEFR), indicating intermediate to advanced English proficiency. The current study chose to maintain individuals in their original classrooms in case it was too disruptive to their learning to move them from their normal learning environment and would threaten the ecological validity of the study. The three classes were randomly assigned to two experimental groups, namely the Unpressured Online Planning (UOP) group (N=28), the Hybrid Online Planning (HOP) group (N=32), as well as a control group, the Pressured Online Planning (POP) group (N=30). For the UOP and HOP conditions, the video clips were slowed to 60% of their normal speed, while for the POP condition, the video was played at normal speed. Participants in the UOP and POP conditions received no pre-task planning time. In contrast, the HOP condition allowed participants to view the video once at normal speed without narrating, fostering content preparedness, before narrating during a second, slowed playback.

B. Research Instruments and Procedure

All participants were instructed to narrate the events of a five-minute-long Mr. Bean video clip as they unfolded. This choice was driven by the clip's minimal dialogue and straightforward comedic content, making it ideal for middle-level students. The setup minimized listening demands, allowing students to focus solely on content narration without the need to decipher key words, thereby enhancing their engagement and concentration on the task.

(a). Utterance Fluency Measures

This study operationalizes fluency as the speaker's speed of delivery, flow of speech, and absence of unnecessary hesitations or interruptions, aligning with Lennon's (1990) conceptualization. Following Tavakoli and Hunter's (2018) narrow view of utterance fluency, we further delineate fluency into three primary dimensions: speed fluency, breakdown fluency, and repair fluency.

Speed fluency was assessed using two primary measures: speech rate and articulation rate. Speech rate, calculated as the number of syllables produced per minute of total speaking time including pauses, provides a comprehensive overview of fluency by accounting for both production speed and pausing behavior (Kormos & Dénes, 2004). Articulation rate, measured as the number of syllables produced per minute of phonation time excluding pauses, offers a more focused measure of speech production speed (Tavakoli, 2016). These measures were selected based on their demonstrated strong predictive power for perceived fluency in previous studies. Breakdown fluency analysis focused on the frequency and duration of pauses, incorporating both filled pauses (non-lexical fillers such as "um," "uh," "er") and unfilled pauses (silent pauses lasting 250 milliseconds or more) (De Jong, 2018). Following Riggensbach's (1991) classification, we differentiated pauses into micro-pauses (≤ 0.2 seconds), hesitations (0.2-0.4 seconds), and unfilled pauses (≥ 0.5 seconds). To ensure a comprehensive analysis, we examined the number of filled and unfilled pauses, as well as the mean length of mid-clause and end-clause unfilled pauses. This approach was informed by methodologies from Derwing et al. (2009), Kormos and Dénes (2004) and Rossiter (2009). Repair fluency was measured using Skehan and Foster's (1999) framework, analyzing four types of disfluency episodes: false starts (abandoned utterances before completion), repetitions (verbatim reiteration of words, phrases, or clauses without modification), reformulations (repeated phrases and clauses with some alteration), and replacements (immediate substitution of lexical items).

All fluency measures were analyzed using PRAAT speech analyzer software, examining both transcriptions and waveforms. To ensure comparability across participants and tasks, all measures of utterance fluency were standardized and calculated per 60 seconds of speech. The selection of these specific measures was based on their recognized reliability in evaluating L2 oral performance, particularly in fluency assessment (Segalowitz, 2010), and their effectiveness in examining online planning processes in L2 speech production (Skehan & Foster, 2005).

(b). Perceived Fluency Measures

To complement the narrative task, perceived fluency was assessed through a comprehensive approach combining student self-assessment and rater evaluation, both based on performance in a dialogic narrative task. Two primary tools were employed for this purpose: the CEFR Self-Assessment Grid (Council of Europe, 2001) and the Analytic Fluency Perception Semantic Differential Scale (Bosker et al., 2013) (see Appendix A and B).

The CEFR Self-Assessment Grid, aligned with the Common European Framework of Reference, serves as the foundation for evaluation. This grid encompasses six levels of fluency, ranging from A1 to C2, and provides detailed descriptors for both spoken interaction and spoken production. These descriptors address various aspects of fluency, including ease of expression, frequency of pauses, complexity of language use, and the ability to handle spontaneous speech in different contexts. Following the dialogic narrative task, students engage in self-assessment by selecting the level on the grid that best reflects their perceived performance. They are instructed to consider factors such as the natural flow of their speech, the occurrence of hesitations, and their ability to express thoughts clearly within the narrative context. Concurrently, trained raters independently evaluate the students' performances by reviewing recorded speech samples from the task. These raters assess the same criteria (i.e., fluency, spontaneity, and interaction) with the aim of providing an objective measure of spoken fluency. They assign fluency levels from A1 to C2 based on the grid's descriptors, and scores from multiple raters are averaged to ensure reliability and minimize potential bias.

Meanwhile, the Analytic Fluency Perception Semantic Differential Scale is administered to students immediately following their completion of the dialogic narrative task. This scale comprises eight statements, each focusing on a specific aspect of fluency, such as smoothness of expression, rhythm, speed, effortlessness, frequency of hesitations and self-corrections, and the ability to maintain continuous speech without pausing to search for words. Students indicate their level of agreement with each statement on a 5-point Likert scale, ranging from 'Strongly Agree' (1) to 'Strongly Disagree' (5). For instance, they assess whether they could express themselves smoothly, maintain a reasonable rhythm, speak quickly, or communicate effortlessly without frequent pauses or self-corrections. Each student's responses are converted into numerical scores, with the total score obtained by summing the individual ratings. A lower total score indicates a higher perceived fluency level, suggesting that the student believes they performed more fluently during the speaking task.

C. Data Coding and Analysis

The data analysis process in this study followed a systematic approach to ensure accuracy and reliability in assessing L2 fluency. The initial step involved transcribing the spoken data, followed by segmenting the unpruned transcriptions into AS-Units (Foster, 2000) and clauses. The AS-unit was employed as a basic unit to segment the transcriptions into clauses, providing a standardized approach to analyzing spoken language. This segmentation was crucial for investigating pauses in mid and end-clause positions, offering insight into the temporal aspects of speech production. For the analysis of fluency measures, we utilized PRAAT software (Boersma & Weenink, 2013), which allowed for precise identification and measurement of speed and pausing features in the speech samples. The data annotation process in PRAAT was conducted manually by an expert researcher with extensive experience in analyzing similar data.

This process involved listening to speech extracts, inspecting spectrograms produced by PRAAT, and identifying and tagging fluency features such as pauses and repairs on a corresponding grid.

To ensure high reliability of the analysis, several measures were implemented. Despite PRAAT's difference and accuracy, the entire data set underwent PRAAT analysis for a second time to ensure a high level of intra-rater reliability. Additionally, 20% of the transcribed files were rechecked and recoded by a second expert rater to verify the reliability of the researchers' data coding. The Pearson correlation coefficient showed a high agreement of 92% between the coding of the researchers and the second-rater regarding fluency measures. This adequate inter-rater reliability permitted us to proceed with further data analysis. Following the reliability checks, the data was analyzed using the SPSS Statistical Package Software (Aldrich, 2018).

IV. FINDINGS

A. Effects of Task Conditions on L2 Speech Performance Measures

This study investigated the impact of task pressure on second language (L2) oral fluency. Through one-way analysis of variance (ANOVA) and post-hoc tests, we found that task conditions significantly affected multiple indicators of oral fluency (see Table 1).

In terms of speech rate, there was a significant main effect of task condition ($F(2, 87) = 13.831, p < .001, \eta^2 = .241$). Post-hoc Bonferroni tests revealed that the Unpressured Group ($M = 165.81, SD = 43.85$) had significantly higher speech rates than both the Pressured Group ($M = 115.61, SD = 39.53, p < .001$) and the Hybrid Group ($M = 125.80, SD = 33.12, p < .001$). This result suggests that reduced time pressure may allow learners to produce more language output. The analysis of articulation rate presented a similar pattern, with a significant main effect of task condition ($F(2, 87) = 17.474, p < .001, \eta^2 = .287$). Post-hoc tests showed significant differences between all groups: the Unpressured Group ($M = 276.83, SD = 37.28$) outperformed the Hybrid Group ($M = 252.32, SD = 30.49, p = .008$), which in turn outperformed the Pressured Group ($M = 229.97, SD = 22.58, p < .001$). This result further supports the impact of time pressure on oral fluency.

Regarding pauses, we observed interesting patterns. The number of filled pauses (such as "um", "uh") was significantly affected by task condition ($F(2, 87) = 18.066, p < .001, \eta^2 = .293$). The Unpressured Group ($M = 14.84, SD = 3.16$) had significantly more filled pauses than both the Pressured Group ($M = 10.89, SD = 2.86, p < .001$) and the Hybrid Group ($M = 12.39, SD = 1.26, p = .001$). Conversely, for unfilled pauses, the Pressured Group ($M = 22.56, SD = 8.14$) had significantly more than both the Unpressured Group ($M = 16.43, SD = 6.52, p = .001$) and the Hybrid Group ($M = 18.50, SD = 1.65, p = .035$). This contrast may reflect different cognitive strategies adopted by learners under various task conditions. Moreover, the mean length of mid-clause pauses was also significantly affected by task condition ($F(2, 87) = 7.856, p = .001, \eta^2 = .153$). The Pressured Group ($M = 0.73, SD = 0.36$) had significantly shorter mid-clause pauses than both the Unpressured Group ($M = 1.22, SD = 0.64, p = .002$) and the Hybrid Group ($M = 1.18, SD = 0.57, p = .005$). This may indicate that under time pressure, learners' language production becomes more fragmented, with reduced time for intra-sentential planning.

TABLE 1
ONE-WAY ANOVA RESULTS FOR THE EFFECTS OF TASK CONDITION ON L2 SPEECH PERFORMANCE MEASURES

Variables	Task Condition Group			F (2, 87)	p	partial η^2
	Pressured	Unpressured	Hybrid			
Speech Rate	115.60 (39.52)	165.81 (43.85)	125.80 (33.12)	13.83	<.001	.241
Articulation Rate	252.32 (30.49)	229.96 (22.57)	276.83 (37.28)	17.47	<.001	.287
No. of Filled Pauses	10.89 (2.86)	14.84 (3.16)	12.39 (1.26)	18.07	<.001	.293
No. of Unfilled Pauses	22.56 (8.14)	16.43 (6.52)	18.50 (1.65)	7.85	<.001	.153
ML of Mid-Clause UP	0.73 (0.36)	1.22 (0.64)	1.18 (0.57)	7.86	<.001	.153
ML of End-Clause UP	5.58 (1.01)	5.73 (1.59)	5.45 (1.75)	0.28	.759	.006
No. of Disfluencies	8.62 (1.91)	7.64 (0.89)	8.33 (2.14)	2.53	.085	.055
CEFR Self-Assessment	3.37 (0.61)	3.47 (0.73)	3.07 (0.70)	2.80	.066	.061
CEFR Rater Assessment	4.37 (0.61)	4.20 (0.48)	4.17 (0.46)	1.25	.291	.028
Semantic Difference	17.87 (3.03)	18.77 (2.89)	18.87 (2.96)	1.04	.357	.023

Note: ML = Mean Length; UP = Unfilled Pauses. Values in parentheses represent standard deviations.

Notably, we found no significant effects of task condition on other indicators such as the mean length of end-clause pauses, the number of disfluencies, CEFR self-assessment scores, teacher CEFR assessment scores, and semantic difference (all $p > .05$). This suggests that these aspects may be less sensitive to task pressure manipulation or may require different measurement approaches.

B. Relationship Between Utterance Fluency, Perceived Fluency, and Semantic Difference

To examine the relationships among utterance fluency, students' self-assessment, raters' assessment, and semantic difference, we conducted a series of Spearman's rank-order correlations. The resulting correlation matrix is presented as a heatmap (see Figure 1), visually illustrating the relationships between the variables. The color intensity in the heatmap represents the strength of the correlations, with positive correlations indicated by shades of red and negative correlations by shades of blue.

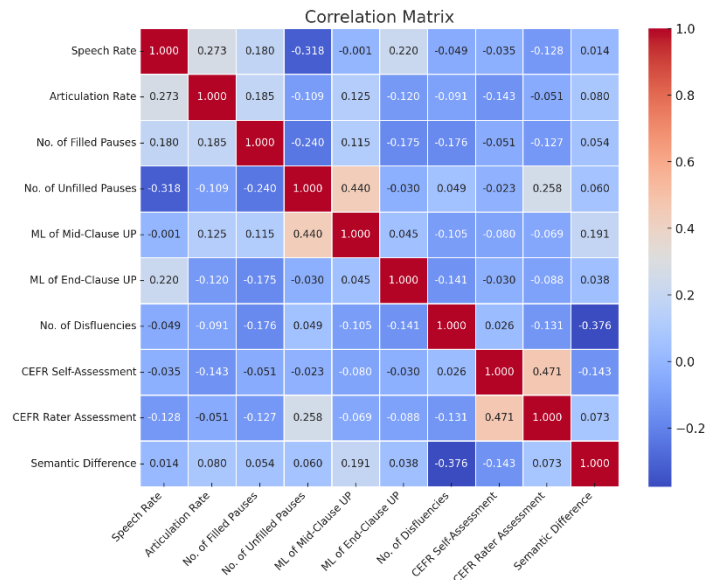


Figure 1. Correlation Matrix of Fluency Dimensions

The correlation analysis of speech fluency measures revealed several significant relationships. Speech Rate showed a moderate positive correlation with Articulation Rate ($r = 0.273$, $p = 0.009$), and a significant negative correlation with the Number of Unfilled Pauses ($r = -0.318$, $p = 0.002$). A strong positive correlation was observed between the Number of Unfilled Pauses and the Mean Length of Mid-Clause UP ($r = 0.440$, $p < 0.001$). However, the correlation between Mean Length of Mid-Clause and End-Clause UPs was weak and not statistically significant ($r = 0.045$, $p = 0.675$). The Number of Disfluencies demonstrated a notable negative correlation with Semantic Difference ($r = -0.376$, $p < 0.001$). In terms of proficiency assessments, CEFR Self-Assessment and CEFR Rater Assessment showed a strong positive correlation ($r = 0.471$, $p < 0.001$). Interestingly, Speech Rate had weak and non-significant correlations with both CEFR Self-Assessment ($r = -0.035$) and CEFR Rater Assessment ($r = -0.128$). Lastly, the analysis revealed a weak, non-significant correlation between Semantic Difference and CEFR Rater Assessment ($r = 0.073$, $p = 0.492$).

V. DISCUSSION

This study provides significant insights into the complexity of second language (L2) oral fluency, particularly regarding the interrelationships between task pressure and various fluency indicators. Our findings deepen the understanding of how task conditions affect L2 performance and elucidate the intricate dynamics of fluency components. The significant impact of task pressure on speech rate and articulation rate aligns with previous research by Ellis (2009) and Yuan and Ellis (2003) on the benefits of planning time. Our findings support Robinson's (2011) Cognition Hypothesis and Skehan's (2016) Limited Attention Capacity Hypothesis, suggesting that different planning conditions can be strategically combined to enhance L2 oral fluency. Notably, the hybrid planning condition yielded the highest speech and articulation rates, consistent with studies by Wang et al. (2019) and Hsu (2017). This aligns with recent work by Suzuki and Kormos (2020), which emphasizes the importance of task design in facilitating L2 fluency development.

Our study revealed a positive correlation between speech rate and articulation rate, challenging the common assumption that faster speech necessarily compromises clarity. This finding echoes Kormos and Dénes (2004) and is further supported by recent research from Tavakoli et al. (2020), who found that higher speech rates are often associated with improved overall fluency ratings. This relationship between speed and clarity underscores the complex nature of fluency and aligns with McDonough and Sato's (2019) call for a more nuanced understanding of fluency development in L2 learners. The analysis of pausing patterns revealed complex interactions between task pressure and different types of pauses. The pressured group exhibiting the highest number of disfluencies corroborates Skehan's (1998) Limited Attentional Capacity Hypothesis and is consistent with recent findings by Wang et al. (2019) on the effects of time pressure on L2 speech production. This result underscores the importance of considering potential trade-offs between fluency and other aspects of language production when designing task-based activities, a point emphasized in recent work by Hasnain and Halder (2023).

Contrary to earlier studies suggesting that time pressure might enhance fluency, our study found that the pressured group exhibited the highest number of disfluencies. This aligns with recent research by Khatib and Farahanynia (2020), who found that excessive time pressure can lead to cognitive overload in L2 speakers, particularly in complex narrative tasks. These findings highlight the need for careful consideration of task conditions in different learning contexts, as also noted by Ahmadian et al. (2023). The unpressured group's higher frequency of filled pauses aligns with Ellis's (2009) assertion that unpressured planning facilitates more careful language monitoring. This interpretation is supported

by recent work from Tavakoli and Wright (2020), who argue for a more comprehensive view of pausing behaviors in L2 speech.

Our findings on CEFR assessments revealed non-significant effects of task conditions on both self-assessment and rater evaluation of CEFR levels, suggesting a complex relationship between task planning conditions and perceived fluency. This aligns with recent research by Suzuki et al. (2021) on the multidimensionality of L2 fluency assessment. A notable discrepancy emerged in the participants' self-perceived fluency compared to rater evaluations, aligning with recent findings by Sadeghi et al. (2017) on the complexities of self-assessment in L2 fluency. The weak correlation between semantic difference and CEFR rater assessment is further corroborated by recent work from Saito et al. (2018) on the role of lexical sophistication in perceived fluency. These results collectively emphasize the complex and multifaceted nature of speech fluency and its assessment. They suggest that effective language teaching methodologies should address various aspects of fluency simultaneously, as proposed by Lambert and Kormos (2022) for a more integrated approach to fluency instruction.

VI. CONCLUSION

This study contributes significantly to the growing body of research on task planning conditions in L2 acquisition, offering valuable insights into the effects of pressured, unpressured, and hybrid planning on oral fluency among Chinese EFL learners. Our findings both support and challenge aspects of existing theoretical frameworks, contributing to a more nuanced understanding of L2 speech production processes.

Consistent with prior research, our study demonstrates that hybrid planning proves to be the most effective in improving fluency, as measured by speech rate, articulation rate, and the reduction of disfluencies. This finding supports the Cognition Hypothesis and underscores the practical value of incorporating diverse task planning strategies into L2 instruction. However, the discrepancies observed in self-assessed fluency, particularly under pressured conditions, highlight the complexity of fluency perception and suggest areas for future research, especially regarding the psychological factors influencing learner self-assessment. The superior performance of the hybrid planning condition in terms of speech rate and articulation rate suggests that combining different planning strategies may be particularly effective in enhancing L2 oral fluency. However, our study also reveals the need for a more comprehensive approach to fluency assessment that accounts for the multifaceted nature of L2 speech production, including aspects such as pausing patterns and semantic variability.

Our research advances the field of Task-Based Language Teaching (TBLT) and L2 acquisition by providing empirical evidence for the effectiveness of hybrid planning conditions and highlighting the complex relationship between objective fluency measures and learners' self-perceived fluency. It underscores the need for a more nuanced approach to fluency assessment and instruction in L2 education, considering both cognitive and affective factors that influence language production under various task conditions. In light of these findings, educators should strive to create balanced task conditions that promote fluency while also fostering accurate self-evaluations, contributing to more effective language learning strategies. By integrating these insights into curriculum design and pedagogical practices, language instructors can better support learners in developing both their oral fluency and their ability to accurately assess their own performance.

Future research should further explore the cognitive processes underlying the relationships between task pressure and fluency components, as well as investigate the long-term effects of different planning conditions on L2 fluency development. Additionally, studies focusing on the psychological aspects of self-assessment in relation to task pressure could provide valuable insights into learner perceptions and metacognitive processes in L2 acquisition.

In conclusion, this study not only contributes to our theoretical understanding of L2 fluency but also offers practical implications for language teaching and assessment. By adopting a more holistic approach to fluency development that considers task conditions, cognitive processes, and learner perceptions, we can enhance the overall effectiveness of L2 education and better prepare learners for real-world communication challenges.

APPENDIX A. CEFR SELF-ASSESSMENT GRID

By placing an "✓" in the below table from 1 to 6, please indicate how well you understand your spoken fluency in general based on the descriptions in the following second table.

Range	Fluency Scale					
	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
No.	Statements					
	Spoken Interaction			Spoken Production		
1	<ul style="list-style-type: none"> I can interact in a simple way provided the other person in prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics. 			<ul style="list-style-type: none"> I can use simple phrases and sentences to describe where I live and people I know. 		
2	<ul style="list-style-type: none"> I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself. 			<ul style="list-style-type: none"> I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job. 		
3	<ul style="list-style-type: none"> I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events). 			<ul style="list-style-type: none"> I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions. 		
4	<ul style="list-style-type: none"> I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views. 			<ul style="list-style-type: none"> I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options. 		
5	<ul style="list-style-type: none"> I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contributions skilfully to those of other speakers. 			<ul style="list-style-type: none"> I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion. 		
6	<ul style="list-style-type: none"> I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it. 			<ul style="list-style-type: none"> I can present a clear smoothly flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points. 		

APPENDIX B. ANALYTIC FLUENCY PERCEPTION SEMANTIC DIFFERENTIAL SCALE

Please indicate the degree to which the statements apply to your perception of fluency in the speaking task that you have just narrated under the appropriate number from 1 to 5 by marking the continuum with an "✓". We are interested in your personal opinion. Thank you very much for your help.

No.	Statements	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
		1	2	3	4	5
1.	I can express myself smoothly in English.					
2.	I can express myself with reasonable rhythm in English.					
3.	I can express myself quickly in English.					
4.	I can express myself effortlessly in English.					
5.	I can express myself with few hesitations in English.					
6.	I can express myself with few self-corrections in English.					
7.	I can express myself with reasonable speed in English.					
8.	I do not have to stop and search for words or expressions when speaking in English.					

REFERENCES

- [1] Ahmadian, M. J., & García Mayo, M. del P. (2017). *Recent Perspectives on Task-Based Language Learning and Teaching*. De Gruyter. <https://doi.org/10.1515/9781501503399>
- [2] Aldrich, J. O. (2018). *Using IBM SPSS statistics: An interactive hands-on approach*. SAGE Publications.
- [3] Awwad, A., & Alhamad, R. (2021). Online task planning and L2 oral fluency: *Does manipulating time pressure affect fluency in L2 monologic oral narratives?* *International Review of Applied Linguistics in Language Teaching*, 0(0), 000010151520200178. <https://doi.org/10.1515/iral-2020.0178>
- [4] Bakhtiary, M. R., Rezvani, E., & Namaziandost, E. (2021). Effects of strategic and unpressured within-task planning on Iranian intermediate EFL learners' oral production. *Journal of Nusantara Studies (JONUS)*, 6(2), 97–115. <https://doi.org/10.24200/jonus.vol6iss2pp97-115>
- [5] Boersma, P., & Weenink, D. (2013). Praat software. *Amsterdam: University of Amsterdam*.
- [6] Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & de Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30(2), 159–175. <https://doi.org/10/gnxxfn>
- [7] Bygate, M. (2016). Sources, developments and directions of task-based language teaching. *The Language Learning Journal*, 44(4), 381–400. <https://doi.org/10/gqmr9>
- [8] De Jong, N. H. (2018). Fluency in Second Language Testing: Insights From Different Disciplines. *Language Assessment Quarterly*, 15(3), 237–254. <https://doi.org/10.1080/15434303.2018.1477780>
- [9] de Jong, N. H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice*, 28(4), 456–476. <https://doi.org/10.1080/0969594X.2021.1951162>
- [10] Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The Relationship Between L1 Fluency and L2 Fluency Development. *Studies in Second Language Acquisition*, 31(04), 533. <https://doi.org/10/b6pdj>
- [11] Ellis, R. (Ed.). (2005). *Planning and task performance in a second language*. John Benjamins Pub. Co.
- [12] Ellis, R. (2009). The Differential Effects of Three Types of Task Planning on the Fluency, Complexity, and Accuracy in L2 Oral Production. *Applied Linguistics*, 30(4), 474–509. <https://doi.org/10/ddpgs5>
- [13] Ellis, R. (2017). Position paper: Moving task-based language teaching forward. *Language Teaching*, 50(4), 507–526. <https://doi.org/10/gjk48w>
- [14] Fillmore, C. J. (1979). On fluency. In *Individual differences in language ability and language behavior* (pp. 85–101). Elsevier.
- [15] Foster, P. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. <https://doi.org/10/dmd8tm>
- [16] Foster, P., & Skehan, P. (1997). Modifying the task: The effects of surprise, time and planning type on task based foreign language instruction. *Thames Valley Working Papers in English Language Teaching*, 4, 86–109.
- [17] Gu, Q., & Van den Branden, K. (2024). The effects of two methods of on-line planning on L2 task-based speaking performance and strategy use. *Language Teaching Research*, 13621688241239460. <https://doi.org/10.1177/13621688241239460>
- [18] Hasnain, S., & Halder, S. (2023). Task-based language teaching approach for improving speaking fluency: Case study of trainee teachers in west Bengal. *World Futures*, 79(7–8), 747–775. <https://doi.org/10.1080/02604027.2021.1996189>
- [19] Housen, A., & Kuiken, F. (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4), 461–473. <https://doi.org/10/b53crr>
- [20] Housen, A., Kuiken, F., & Vedder, I. (2012). Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA. *Language Learning & Language Teaching*. Volume 32. *Language Learning & Language Teaching (MS)*. <https://doi.org/10/gqmt6k>
- [21] Hsu, H. C. (2017). *The Effect of Task Planning on L2 Performance and L2 Development in Text-Based Synchronous Computer-Mediated Communication*. *Applied Linguistics*, 38(3), 359–385.
- [22] Johnson, M. D., & Abdi Tabari, M. (2022). Task Planning and Oral L2 Production: A Research Synthesis and Meta-analysis. *Applied Linguistics*, 43(6), 1143–1164. <https://doi.org/10.1093/applin/amac026>
- [23] Joo, M. (2022). Effects of pre-task and on-line planning on complexity, fluency, and accuracy in computer-based English speaking and writing tests. *Korean Journal of English Language and Linguistics*, 22, 938–956.
- [24] Khatib, M., & Farahanynia, M. (2020). Planning conditions (strategic planning, task repetition, and joint planning), cognitive task complexity, and task type: Effects on L2 oral performance. *System*, 93, 102297. <https://doi.org/10.1016/j.system.2020.102297>
- [25] Kim, N. (2018). The Effects of Online Planning on CAF in L2 Spoken and Written Performance. *English Teaching*, 73(3), 3–28. <https://doi.org/10.15858/engtea.73.3.201809.3>
- [26] Kim, N. (2020). Conditions and tasks: The effects of planning and task complexity on L2 speaking. *Journal of Asia TEFL*, 17(1), 34–40.
- [27] Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates.
- [28] Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164. <https://doi.org/10/cmhpnk>
- [29] Lambert, C., Aubrey, S., & Leeming, P. (2021). Task Preparation and Second Language Speech Production. *TESOL Quarterly*, 55(2), 331–365. <https://doi.org/10/gnj7zq>
- [30] Lampropoulou, L. (2023). The use and impact of pre-task planning time in the monologic task of LanguageCert speaking tests. *Language Education & Assessment*, 6(1), 1–18. <https://doi.org/10.29140/lea.v6n1.1180>
- [31] Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387–417. <https://doi.org/10/bqbw5k>
- [32] Levelt, W. J. (1989). *Speaking: From intention to articulation*. MIT press.
- [33] Long, M. H. (2016). In Defense of Tasks and TBLT: Nonissues and Real Issues. *Annual Review of Applied Linguistics*, 36, 5–33. <https://doi.org/10/gjk48x>

- [34] McDonough, K., & Sato, M. (2019). Promoting EFL students' accuracy and fluency through interactive practice activities. *Studies in Second Language Learning and Teaching*, 9(2), 379–395. <https://doi.org/10/gqmsx9>
- [35] Nielson, K. B., & DeKeyser, R. (2019). Working memory and planning time as predictors of fluency and accuracy. *Journal of Second Language Studies*, 2(2), 281–316. <https://doi.org/10.1075/jsls.19004.bro>
- [36] Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 555–578. <https://doi.org/10/ds7cjr>
- [37] Panahzadeh, V., & Asadi, B. (2019). On the Impacts of Pressured vs. Unpressured On-line Task Planning on EFL Students' Oral Production in Classroom and Testing Contexts. *Eurasian Journal of Applied Linguistics*, 5(3), 341–352. <https://doi.org/10.32601/ejal.651267>
- [38] Qiu, X., & Bui, G. (2022). Pre-task planning effects on learner engagement in face-to-face and synchronous computer-mediated communication. *Language Teaching Research*, 13621688221135280. <https://doi.org/10.1177/13621688221135280>
- [39] Riggensbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14(4), 423–441. <https://doi.org/10.1080/01638539109544795>
- [40] Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. <https://doi.org/10/d2tjx3>
- [41] Robinson, P. (2005). Cognitive Complexity and Task Sequencing: Studies in a Componential Framework for Second Language Task Design. *IRAL - International Review of Applied Linguistics in Language Teaching*, 43(1), 1–32. <https://doi.org/10/b88cmt>
- [42] Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. *Investigating Tasks in Formal Language Learning*, 7, 26.
- [43] Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), *Task-Based Language Teaching* (Vol. 8, pp. 87–122). John Benjamins Publishing Company. <https://doi.org/10.1075/tblt.8.04rob>
- [44] Robinson, P. J. (Ed.). (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. John Benjamins Pub. Co.
- [45] Rossiter, M. J. (2009). Perceptions of L2 Fluency by Native and Non-native Speakers of English. *The Canadian Modern Language Review*, 65(3), 395–412. <https://doi.org/10.3138/cmlr.65.3.395>
- [46] Sadeghi, K., Mousavi, M. A., & Javidi, S. (2017). Relationship Between EFL Learners' Self-Perceived Communication Competence and Their Task-Based and Task-Free Self-Assessment of Speaking. *Journal of Research in Applied Linguistics*, 8(2), 31–50.
- [47] Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*, 39(3), 593–617. <https://doi.org/10/gfr22w>
- [48] Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- [49] Skehan, P. (1998). *A Cognitive Approach to Language Learning*. OUP Oxford.
- [50] Skehan, P. (2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, 30. <https://doi.org/10/bq67g4>
- [51] Skehan, P. (2014). The context for researching a processing perspective on task performance. In *Processing Perspectives on Task Performance* (pp. 1–26). John Benjamins.
- [52] Skehan, P. (2015). Limited attention capacity and cognition. Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (pp. 123–156). Amsterdam/Philadelphia: John Benjamins.
- [53] Skehan, P. (2016). Tasks versus conditions: Two perspectives on task research and their implications for pedagogy. *Annual Review of Applied Linguistics*, 36, 34–49. <https://doi.org/10/gqmrhc>
- [54] Suzuki, S., & Kormos, J. (2022). The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition*, 1–27. <https://doi.org/10/gqmsxk>
- [55] 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/gjvx8h>
- [56] Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, 54(2). <https://doi.org/10/gnxxgh>
- [57] Tavakoli, P., & Hunter, A.-M. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, 22(3), 330–349. <https://doi.org/10.1177/1362168817708462>
- [58] Tavakoli, P., Nakatsuhara, F., & Hunter, A. (2020). Aspects of Fluency Across Assessed Levels of Speaking Proficiency. *The Modern Language Journal*, 104(1), 169–191. <https://doi.org/10.1111/modl.12620>
- [59] Ur, P. (2013). *Language-teaching method* (67(4), 468–474.). *ELT Journal*.
- [60] Van Os, M., De Jong, N. H., & Bosker, H. R. (2020). Fluency in Dialogue: Turn-Taking Behavior Shapes Perceived Fluency in Native and Nonnative Speech. *Language Learning*, 70(4), 1183–1217. <https://doi.org/10/gqmszb>
- [61] Wang, Z., Skehan, P., & Chen, G. (2019). Q1-The effects of hybrid online planning and L2 proficiency on video-based speaking task performance. *Instructed Second Language Acquisition*, 3(1), 53–80. <https://doi.org/10.1558/isla.37398>
- [62] Yuan, F., & Ellis, R. (2003). *The Effects of Pre-Task Planning and On-Line Planning on Fluency, Complexity and Accuracy in L2 Monologic Oral Production*. *Applied linguistics*, 24(1), 1–27.

Xu Han is a PhD candidate majoring in the English Language, at the Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Malaysia. She graduated from York University in the UK. She takes a lively interest in spoken English and language fluency assessment.

Yasir Yahya received his Ph.D. in Education from the University of Reading, UK. He is currently a Senior Lecturer at Universiti Putra Malaysia (UPM). His research interests include Applied Linguistics, Task-Based Language Teaching, and L2 Speech. He has published several papers on ESL learning strategies and task-based language teaching.

Yap Ngee Thai received her Ph.D. in Linguistics from the University of Delaware, USA. She is an Associate Professor at Universiti Putra Malaysia (UPM). Her research focuses on phonetics, second language acquisition, and bilingualism. She has published extensively on ESL vowel production and vocabulary learning.